

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة النهرين

المعلوماتية الحيوية Bioinformatics

E. shawalteri
E. elara
E. varians
E. variata
E. variata

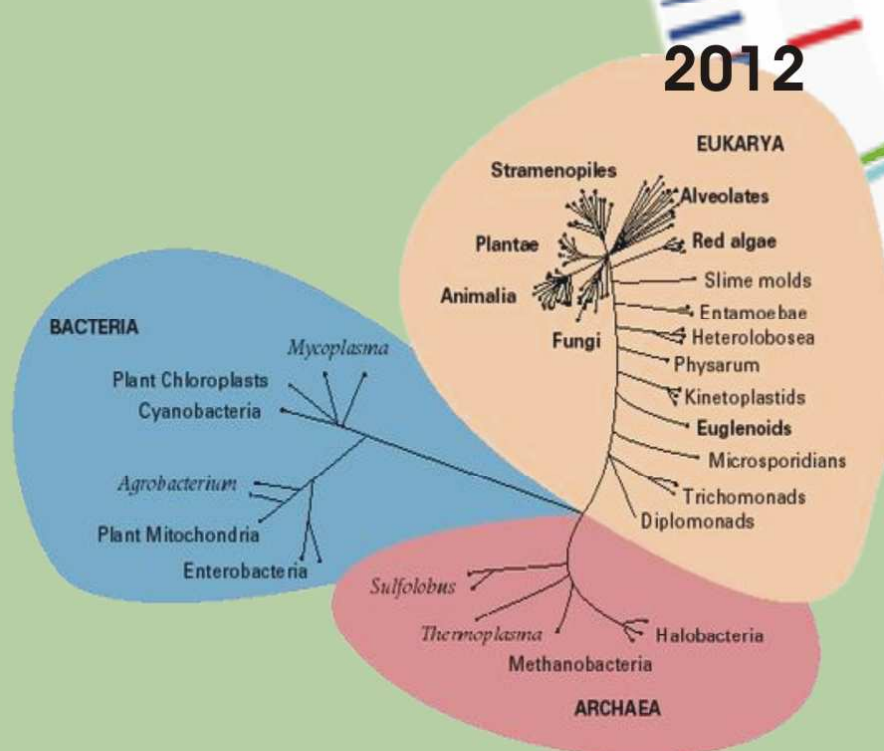
دكتور
علي عبد الحافظ ابراهيم

قسم التقنيات الحيوية / كلية العلوم
جامعة النهرين

دكتورة
زهرة محمود الخفاجي

معهد الهندسة الوراثية والتقنيات الحيوية
للدراسات العليا / جامعة بغداد

2012



المعلوماتية الحيوية

Bioinformatics

دكتور

علي عبد الحافظ إبراهيم

قسم التقنيات الحيوية \ كلية العلوم \
جامعة النهرين \ العراق

دكتورة

زهرة محمود الخفاجي

معهد الهندسة الوراثية والتقنيات الحيوية
للدراسات العليا \ جامعة بغداد

2012

أَعُوذُ بِاللَّهِ مِنَ الشَّيْطَانِ الرَّجِيمِ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿ إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ ﴿٤٩﴾ وَمَا أَمْرُنَا إِلَّا وَاحِدَةٌ كَلَمْحٍ بِالْبَصَرِ ﴿٥٠﴾ ﴾

القمر: ٤٩ — ٥٠

﴿ لَقَدْ أَحْصَاهُمْ وَعَدَّهُمْ عَدًّا ﴿٩٤﴾ وَكُلُّهُمْ عِندَ يَوْمِ الْقِيَامَةِ فَرْدًا ﴿٩٥﴾ ﴾

مريم: ٩٤ — ٩٥

﴿ لَيَعْلَمَنَّ أَنَّ قَدْ أَبْلَغُوا رَسُولَكَ رَبِّهِمْ وَأَحَاطَ بِمَا لَدَيْهِمْ وَأَحْصَى كُلَّ شَيْءٍ عَدَدًا ﴿٢٨﴾ ﴾

الجن: ٢٨

﴿ إِنَّا نَحْنُ نُحْيِي الْمَوْتَى وَنَكْتُبُ مَا قَدَّمُوا وَءَاثَرَهُمْ وَكُلَّ شَيْءٍ أَحْصَيْنَاهُ فِي

إِمَامٍ مُبِينٍ ﴿١٢﴾ ﴾

يس: ١٢

شكر وتقدير

الحمد لله حمدا طيبا مباركا فيه كما يحب ربنا ويرضى والصلاة والسلام على سيدنا محمد وعلى آله الطيبين وأصحابه الغر الميامين ومن تبعهم بإحسان الى يوم الدين.

نتقدم بالشكر الجزيل الكبير الى د. عصام فاضل الجميلي رئيس قسم التقنية الحيوية / معهد الهندسة الوراثية والتقنيات الحيوية للدراسات العليا / جامعة بغداد للاستيعاب والتفهم والمساندة. كما نود ان نشكر لجنة مناقشة المفردات المؤلفة من الدكتور اياد محمد علي رئيسا / قسم التقنيات الحيوية والدكتور جمال محمد كاظم عضوا / قسم الحاسبات والدكتورة احلام جميل خليل عضوا / قسم الرياضيات (جامعة النهرين) ، كما نود تقديم الشكر الجزيل الى المقوم العلمي الدكتور اياد محمد علي / قسم التقنيات الحيوية / جامعة النهرين ، والشكر والتقدير للدكتور مؤيد احمد يونس قسم وقاية النبات / كلية الزراعة / جامعة بغداد لمتابعته الحثيثة طيلة تأليف الكتاب . والشكر والتقدير للدكتور فالح عبد الكريم الشخيلي لما قدمه من خبرات طيلة مدة تأليف الكتاب. كما نتقدم بالشكر الوافي للأخت ندى خالد والسيدة ورقاء يحيى والمهندسة طيبة بشير محمود والأستاذ غيث لطفى للقيام بطبع المسودات الأولية للكتاب ، كما نتقدم بالشكر الكبير للأستاذ نذير بشير محمود لتعامله مع برامج الحاسوب وفهرسة الكتاب وتصميم الغلاف . والشكر موصول للأخت المدرس المساعد زينب هاتف والمدرس علي عماد العزاوي / معهد الهندسة الوراثية والتقنيات الاحيائية للرفقة الطبية والتشجيع أثناء إعداد الكتاب . والأستاذ حسن محمود أبو المعالي لرفده بالمراجع الملائمة.

كما أتقدم بالجزيل من الامتنان لأخي بشير محمود الخفاجي الذي كان الودد بعد الله عز وجل الذي استند اليه عندما تطيش مراكبي لتشجيعه ومتابعة سير تقدم الكتاب. وكلمة أخيرة أود ان اعبر عنها " ان هذا العمل ولد من رحم العذاب وترعرع في أحضان الآلام وقصم ظهر العافية " . ونسألكم الدعاء لنا ولوالدينا .

المؤلفان

كانون الأول \ 2011

المقدمة

تفتقر المكتبة العربية وتكاد تخلو من الكتب التي تناقش موضوع المعلوماتية الحيوية ، لذلك كانت المبادرة بوضع كتاب يتناول الأسس لهذا الحقل الذي هو في تطور هائل في أنحاء العالم ويقع الكتاب في ست فصول : الأول يتناول مخطط عام للمعلوماتية الحيوية والتعريف بها وبعض الجوانب التي يتعامل معها . اما الفصل الثاني فيخصص وسائل المعلوماتية الحيوية وهي البرامج وقواعد البيانات اذ تم تناول الأخيرة بشكل العام ، وكانت هناك وقفات مع قواعد البيانات في الفصول التي تليه عندما تكون الحاجة ملحة وكذلك في حالة التبيان ، اما البرامج فقد تم تناول الأسس دون ذكر المواقع على الانترنت لانها قد تخفي من شبكة الانترنت لضعفها وعدم أدائها الجيد وقد ذكر النزر القليل منها والتي تعد بالمئات او الآلاف .

وتتناول الفصل الثالث بعض الجوانب الحيوية للأحياء التي تعد المادة التي يعمل فيها مجال المعلوماتية الحيوية . ولم يتم تناول الجوانب الحيوية بشكل مفصل وانما أعطيت الأطر الخارجية وبقدر التعلق بالمعلوماتية الحيوية ، وقد تم تناول بعض الجوانب بشي من العمق عند الحاجة اليها في توضيح توجهات البرامج التي تناولها . وتتناول الفصل الرابع جوانب الانترنت باعتباره واسطة التواصل والبحث وكذلك تناول الجوانب الإحصائية التي تعد الركائز القوية التي تربط بين المعلوماتية الحيوية ووسائل المعلوماتية لتحليلها .

اما الفصل الخامس فقد اخذ حصة الأسد من الكتاب اذ انه يمثل جوهر حقل المعلوماتية الحيوية وتم تناول العديد من النواحي التي تخفي الجزيئات الحيوية وبالدرجة الرئيسية DNA والبروتينات و RNA . اما الفصل الأخير فيتناول القيل من جوانب الحياة لتي تتدخل المعلوماتية الحيوية بقوة والبعض منها جديد الاستحداث . وغطي الكتاب فضلا عن قائمة المحتويات لكل فصل بفهارس للمفردات العربية والانكليزية لتسهيل الاستعمال وزيادة الفائدة . والله موفق .

الفصل الاول

نظرة عامة

(9-1)

الفصل الثاني

وسائل المعلوماتية الحيوية

(67-10)

الفصل الثالث

الأنظمة الحيوية

(263-68)

الفصل الرابع

الانترنت والإحصاء وتطبيقاته في الحاسوب

(368-264)

الفصل الخامس

تحليل البيانات

(676-369)

الفصل السادس

المعلوماتية وبعض جوانب الحياة

(730-677)

المراجع

(743-730)

الفصل الأول

نظرة عامة Overview

المعلوماتية الحيوية Bioinformatics
أهمية المعلوماتية الحيوية
أسباب تطور المعلوماتية الحيوية
تطور ومهام المعلوماتية الحيوية
الانترنت والمعلوماتية الحيوية
الإحصاء والمعلوماتية الحيوية
الحاسوب والمعلوماتية الحيوية
المعلوماتية الحيوية مقابل المنظومة البايولوجية
مرافقات المعلوماتية الحيوية
تطبيقات المعلوماتية الحيوية
المحددات في المعلوماتية الحيوية
ما يحتاجه العامل في مجال المعلوماتية الحيوية

نظرة عامة Overview

دخلت تقنيات المعلوماتية Informatics مختلف جوانب الحياة وذلك لأسباب عدة ، فهناك المعلوماتية الحيوية Bioinformatics والمعلوماتية الكيماوية Chemoinformatics والمعلوماتية المناعية Immunoinformatics وغيرها. ومجال الكتاب يتحدد بالمعلوماتية الحيوية التي سوف يضم توضيح العديد من جوانبه وليس كل الجوانب لان ذلك سيكون شاقاً.

المعلوماتية الحيوية Bioinformatics

قد لا يوجد تعريف واضح المعالم وقياسي لهذا الحقل ولكن يمكن القول انه "الادارة والتعامل والاستغلال للمعلومات الحيوية باستعمال طرق الحاسوب"، لذا يكون الحقل أحد حقول المعرفة المتداخلة بين العلوم. ويتضح انه دمج بين علوم الحياة المختلفة وعلوم الحاسوب، ولكن المجال يخص العلوم الجزيئية على وجه الخصوص Molecular biology وبذا يمكن ان يطلق عليه اسم علم الحياة الحاسوبي Computational molecular biology والذي يضم بين جنبيه عمليات المحاكاة الحيوية Biosimulation والتصوير الحيوي Bioimaging وغيرها لتصب في النهاية الى ادارة وتحليل المعلومات الحيوية.

ومن هذا يتضح ان المعلومات الحيوية آتية من تحليل الجزيئات الحيوية العملاقة مثل DNA، RNA وما ينتج من الفعاليات الجارية فيها وهو إنتاج البروتينات. ومن الجانب الثاني يقوم الحاسوب بترتيب وتصنيف هذه المعلومات لغرض إيجاد أنماطاً لتحليلها واستغلال المعلومات التي فيها Data biomining ، لذا فهو يساعد في الكشف عن حقائق بايولوجية جديدة تساعد في توحيد علوم الحياة.

أهمية المعلوماتية الحيوية

تظهر أهمية المعلوماتية الحيوية من ازدياد الاهتمام بها على كافة الأصعدة ودخولها الى أروقة الجهات المنفذة ، فقد أدخلت في منهاج الدراسات الأولية والدراسات العليا ومجالات البحوث العلمية على مختلف مشاربها، وأنشأت الأقسام العلمية الخاصة بها وكذلك الكليات ووصل البعض الى إنشاء جامعات خاصة بها لتخريج كوادر تعمل في هذا المجال، ومن هنا جاءت الزيادات في التخصيصات المالية واستحداث درجات علمية خاصة بهذا المجال.

ففي مجال المعلوماتية الحيوية والدراسات التي تجري في الحاسوب *In silico* يتم وضع الخطط واستعمال أفضل الظروف وذلك ببناء موديلات حاسوب للخلايا الحية وشبكات أيضاً الداخلية بالاعتماد على تجارب المختبر Wet-lab-data وتطبيق المحاكاة الحاسوبية *In silico* simulation لتحسين الحاصل والإنتاجية في حالة إجراء دراسات الصناعات الحيوية عند استخدام الأحياء المجهرية على سبيل المثال، ومن هنا نشأ ما يسمى التقنية الحيوية الحاسوبية *In silico* biotechnology للوصول الى أفضل الظروف المثالية - النظرية- قبل بدء عمليات الإنتاج الفعلي للعديد من المواد مثل الأدوية وغيرها.

وفضلاً عن الجوانب الاقتصادية المذكورة آنفاً والتي تعد المحرك الرئيس لأي عملية تطوير، هناك جوانب أخرى قد تسبق الجوانب الاقتصادية، منها باستعمال الوسائل الحيوية يمكن استغلال المعلومات الفياضة التي ضمت في قواعد البيانات Databases (كما سيأتي ذكره لاحقاً) والتي تتم بطرق مختلفة. ومن هذه المعلومات التعرف على مواقع الجينات في التواليات DNA (DNA sequences) وتحديد تركيب البروتينات الناتجة من الجينات أو التواليات المشفرة Coding genes و Coding sequences (وتشير المشفرة هنا إلى قابلية ترجمتها لإعطاء البروتينات)، وبلي عملية تحديد التركيب ما يترتب عليه من الوظائف، والقيام بعمليات التجميع والعنقدة Clustering لنواتج الجينات بعد التعبير عنها. وهذا يجر بمجملة إلى زيادة المعرفة وفهم الخلايا الحية وكيفية القيام بوظائفها على المستوى الجزيئي لغرض استغلالها بالنحو الأمثل.

أسباب تطور المعلوماتية الحيوية

من الواضح للعيان زيادة المعلومات في مجال علم الحياة وخاصة على المستوى الجزيئي بشكل هائل جداً ومذهل، فمثلاً يتضاعف عدد الجينومات التي يحدد تواليها والمودعة في مراكز خزن المعلومات حوالي مرة سنوياً أو أقل من سنة، ومن هذا المنطلق أخذت بعض المؤسسات المعنية على عاتقها إصدار نشرات دورية شهرية أو تقارير شهرية تحت عنوان "Genomes of the month" وتقوم هذه المؤسسات بنشذيب المعلومات وإزالة المكرر منها وتهميشها Annotation (من هامش أو توضيح)، وتوضيح كل المعلومات المتوفرة لكل جينوم وذكر طرق التشذيب وبرمجة تواليات الجينوم (DNA) وترجمتها إلى بروتينات وإجراء المقارنات بين التواليات الجديدة مع تلك المودعة في قواعد البيانات ووضع الصور التوضيحية لها والرسومات الخاصة بأشجار العلاقات التطورية Phylogenetic trees، ثم بعد ذلك إيداعها في قواعد البيانات الخاصة بكل مركز. وفضلاً عن ذلك وضع روابط على شبكة الانترنت توصل المستخدم إلى مواقع الكترونية Web sites وأخرى إلى معلومات إضافية قد يحتاجها الباحث لإجراء الدراسة والسلالة التي حُددت تواليها وغيرها من المعلومات.

فمن الواضح أن الزيادة في المعلومات هي الدافع الرئيس وراء تطور المعلوماتية الحيوية، ولكن من الأسباب الأخرى المهمة أن الزيادة قد لا تعني شيئاً أن لم يكن بالإمكان الاستفادة منها، ومن هنا جاءت الحاجة للوصول إلى البيانات واستخدامها واستغلالها واستخلاص المفيد منها. فعملية البحث في البيانات المتكررة والكبيرة والتي يصل البعض منها إلى آلاف الصفحات تفوق القدرات البشرية فضلاً عن أنها مملة وتستغرق الوقت الطويل وتحتاج إلى عمليات رياضية معقدة لذلك كان لابد من إيجاد وسائل للاستغلال.

ومن جهة ثانية فإن وسائل المعلوماتية الحيوية Bioinformatic tools توفر الإمكانيات لتحليل النتائج وربطها واستخلاص العبر منها، وذلك مما حدا ببعض العاملين إلى توقع إحلال المعلوماتية الحيوية بشقها المتعلق بالشبكة العنكبوتية أن يحل محل التجارب المخبرية - وهذا بطبيعة الحال غير ممكن - لأن الكثير من تنبؤات الحاسوب يكون الفيصل فيها هو الواقع

التجريبي. فالعمل المختبري يحتاج الى طرق عمل وأدوات للتعامل مع المواد الحيوية وهي التي تمثل ارض الواقع ، لتدخل في مجالات المعلوماتية الحيوية التي تكون أدواتها لغات البرمجة وقواعد البيانات وتقنيات الحاسوب للتعامل مع ارض الواقع المذكور آنفاً.

تطور ومهام المعلوماتية الحيوية

مما ذكر أعلاه حول تعدد مشارب ومصادر المعلوماتية الحيوية ، يمكن القول ان كل مجال يساهم بجزء معين من المهام التي تقوم بها المعلوماتية الحيوية فلو تم البدء بالجوانب الحيوية وبأخذ نظرة تاريخية مبسطة في هذا الجانب يلاحظ ان زيادة عدد الجينومات التي تم تحديد تواليها يزداد باضطراد فمثلاً:

- في عام 1983 تم تحديد توالي العائلي T7 Bacteriophage الذي يضم 39,937 زوج من القواعد النتروجينية (base pair = bp) وينتج عدداً من البروتينات.
 - في عام 1995 تم تحديد توالي جينوم السلالات الحرة المعيشة للبكتريا *Haemophilus influenzae* RdW20 وكان مؤلفاً من 1,830,137 زوج قواعد.
 - في عام 1996 تم تحديد توالي جينوم الدودة *Caenorhabditis elegans* والتي تعد أحد النماذج الدراسية وكان مكوناً من 95,078,296 زوج من القواعد ويضم ما يقرب من 19,009 جين.
 - في عام 1998 تم تحديد توالي جينوم أحد سلالات البكتريا *Escherichia coli* K-12 الذي كان بطول 4,639,221 قاعدة ويضم 4293 من أطر القراءة المفتوحة Open reading frame اي المناطق المشفرة للبروتينات.
 - في عام 2000 تم تحديد توالي الجينوم الكامل لذبابة الفاكهة *Drosophila melanogaster* (النموذج الدراسي) ليكون عدد أزواج القواعد فيه 116,117,226 متضمناً 13,601 جين.
 - في آذار 2003 أعلن عن اكتمال تحديد الجينوم البشري *Homo sapiens* وتضمن الأرقام عرضت للتحريض والتعديل لتكون أقل من ذلك (كما سيأتي ذكره في موقع آخر من الكتاب) وكان قد بدأ تحديد الجينوم البشري Human genome project في نهاية ثمانينيات القرن المنصرم وكان المحفز الكبير لتطوير المعلوماتية الحيوية التي توسع استعمالها عبر الانترنت او الشبكة العنكبوتية العالمية في التسعينات هذا فضلاً عن الكم الهائل من المعلومات في مجال علوم الحياة الأخرى التي أسفرت عنها تجارب المختبر.
- والحقيقة ان مصطلح المعلوماتية الحيوية قد استعمل مؤخراً بشكله الرسمي اما واقعه العملي فقد بدأ من الستينات عندما وضعت الباحثة Margaret O. Dayhoff أول قاعدة معلومات لتوالي البروتينات في أطلس خاص بتواليات وتركيب البروتينات ضمن سلسلة من الإصدارات استمرت الى منتصف سبعينات القرن الماضي ، وكان عندها عدد التواليات يزيد قليلاً عن العشرة مقارنة بالتواليات في الوقت الحاضر التي تصل الى عشرات الآلاف .

ورافقت ذلك (من الناحية التاريخية) وضع برنامج عام 1970 من قبل Needleman و Wunsch لمقارنة التواليات البروتينية والتي كانت خطوة أساسية لمقارنة التواليات، تلتها بعد ذلك وضع برنامج عام 1974 للتنبؤ أو حدس Prediction تركيب البروتينات والذي كان منطلقا لعدد كبير من برامج التنبؤ بتركيب البروتينات.

ومن الناحية التاريخية شهدت ثمانينيات القرن الماضي تطورا "سريعا" في وضع قواعد البيانات ووضع برامج للبحث فيها مثل FASTA و BLAST التي سيتم تناولها في مواقع أخرى. ولذلك فإن العاملين في مجال الحاسوب والرياضيات أخذوا على عاتقهم تطوير الجوانب الخاصة بهم ليواكبوا التطورات في المجال الحيوي، وبذا أصبحت وسائل الحاسوب والرياضيات جزءا مهما من علوم الحياة الحديثة أي ان المعلوماتية الحيوية أصبحت جزءا مهما في مختبرات علوم الحياة ، وهذا يتضح من ان المعلوماتية الحيوية التي بدأت بسيطة في بداية الستينيات توسعت نتيجة للتوسع المتوازي بين علوم الحياة وعلوم الحاسوب. وقد كانت هناك تحديات كبيرة أمام علوم الحاسوب أدت في النهاية - ربما- الى جلب علوم الحاسوب بشكل أقرب الى مجال الاكتشافات العلمية والتي منها علوم الحياة، وهذا يتطلب ويتطلب التطبيق الأمثل فضلا عن وضع البرامج الملائمة للجوانب والحقائق في مجال علوم الحياة. ووضع الخوارزميات Algorithms والبرامج وإدخال الإحصاء بشكل فاعل فيها للتوصل الى العلاقات بين مجاميع البيانات.

مما تقدم يتضح ان المعلوماتية الحيوية فضلا عن الجوانب الحيوية تحتاج الى:

- خزن البيانات وهذا يتمثل بإنشاء قواعد البيانات على مختلف أنواعها.
- استرجاع البيانات المخزونة باستخدام وسائل مختلفة.
- استخدام وسائل للتلاعب Manipulation للمعلومات الموجودة عن جزئيات الحياة سواء كانت DNA ، RNA او البروتينات.

الانترنت والمعلوماتية الحيوية

منذ تسهيل مهمة استعمال الانترنت - ربما - للجميع في تسعينات القرن المنصرم زاد انتشار وتبادل المعلومات الحيوية والتي كان لها التأثير الكبير في تطور علوم الحياة الحديثة والجزئية على وجه الخصوص، فقد كان الانفجار الهائل في تحديد تواليات الجينومات الدافع وراء الزخم الكبير الى إيجاد برامج الحاسوب لتحليل الكم الهائل من المعلومات.

الإحصاء والمعلوماتية الحيوية

يشكل الإحصاء أحد الدعائم الأساسية للمعلوماتية الحيوية ، فالكم الكبير من المعلومات ومحاولة مقارنتها واستنباط الاستنتاجات منها يحتاج الى عمليات إحصائية، تكون في بعض الاحيان معقدة جدا.

وولوج الإحصاء الى مجال المعلوماتية الحيوية ضروري ولكنه جلب معه احتمالاته ومصفوفاته الترجيحية وغيرها من الجوانب والتي لا بد منها عند التعامل مع المعلومات الكثيرة، لذلك يلاحظ ان الإحصاء يدخل في كل مجال وكل برنامج او خوارزمية لغرض وضع المستخدم على بيئة

من أمره ويعيده الى الجوانب الأخرى لترجيح الاستنتاج، فالإحصاء يعطي الترجيحات النظرية والحسابية التي يمكن القبول بها في حالة تناغمها مع الحقائق في الجوانب الأخرى ، وربما كانت القيم الإحصائية المستخرجة او المستعملة هي المسبار الأفضل لما غفل عنه الباحث.

الحاسوب والمعلوماتية الحيوية

العديد من المهام في المعلوماتية الحيوية تحتاج الى الإعادة ربما ملايين من المرات، فمثلا مقارنة توالي Sequence جديد يحتاج ان يقارن بكل التواليات المخزونة في قواعد البيانات، فضلا عن الحاجة في بعض الأحيان الى مقارنة العديد من التواليات مع بعضها لمعرفة العلاقة التطورية Phylogenetic فيما بينها وهذه تكون أكبر من القدرات البشرية لذلك يستعمل الحاسوب في حل هذه العمليات، وكذلك يستعمل الحاسوب لحل مشاكل طوي البروتينات Protein folding بعد معرفة توالي الحوامض الامينية فيه، او استنتاج المسار الأيضي Metabolic pathway بعد تزويده بنسق التعبير بشكل RNA ، وغيرها من العمليات والتي سيأتي ذكرها لاحقا.

من هذا يبدو واضحا ان وسائل الحاسوب تعتمد بعد او ما وراء المعلومات الحيوية على المعرفة بالرياضيات وعلم الإحصاء وتقنيات المعلوماتية.

المعلوماتية الحيوية مقابل المنظومة البايولوجية

Bioinformatics vs systems biology

المعلوماتية الحيوية تتعامل مع الأنظمة الحيوية على المستوى الجزئي بشكل أساسي ولذلك عدها البعض كما ذكر أنفا أنها علم الحياة الجزئي الحاسوبي. وفي حين يتعامل مجال المنظومات البايولوجية مع مجالات أوسع تخص مجتمعات الكائنات وتداخلاتها مع ما يحيطها، فهي تشمل إيجاد الموديلات الرياضية في الأنظمة البيئية Mathematical modeling and ecosystems، فضلا عن دراسة ديناميكية المجتمعات وتطبيق تأثير الجينات في التصرفات التي يبديها المجتمع، وتشمل ايضا إيجاد وبناء العلاقات التطورية بالاعتماد على سجلات الحفريات باستعمال وسائل الحاسوب، دون الأخذ بنظر الاعتبار الجزيئات الحيوية العملاقة في إيجاد العلاقات والتي تكون مهمة وأساسية في نهج المعلوماتية الحيوية.

ومن النظر الى أعلاه يلاحظ ان كل من المعلوماتية الحيوية ودراسات المنظومات البايولوجية هي تهدف في النهاية الى هدف مشترك لفهم الأنظمة الحيوية وربما نقلها من المجال النوعي الى المجال الكمي. وربما شكلت دراسة المنظومات البايولوجية مستقبل المعلوماتية الحيوية وذلك لان الأولى هي الأوسع والاشمل، والذي يترتب عليه جهود كبيرة للعاملين في مجال المعلوماتية الحيوية Bioinformaticians الى بذل المزيد من الجهود في الجوانب الإحصائية والرياضية وأنظمة المحاكاة.

مرافقات المعلوماتية الحيوية

رافقت تطورات المعلوماتية الحيوية في الجوانب العملية ما يقابلها في الجوانب النظرية والمصطلحات التي تصف المعلومات، وتمثلت هذه بدراسة -omics ، والتي تدرس

المصطلحات الكاملة المقابلة -ome فظهر مصطلح Genome والذي يمثل كل المحتويات الوراثية للكائن من حيث التركيب والترتيب والوظائف وغيرها من المعلومات تدرس ضمن حقل دراسة الجينوم Genomics. وتفرعت عن ذلك دراسات ثانوية مثل دراسات الجينوم الوظيفية Functional genomics الذي يهدف الى الكشف عن الدور الوظيفي للجينات المختلفة وكيفية أدائها للوظائف وتداخلها مع بعضها البعض ضمن الشبكات الخلوية. وكذلك مصطلح الدراسات الجينومية التركيبية Structural genomics التي تهتم بالنواحي التركيبية ، ودراسات الجينوم المقارنة Comparative genomics وتعنى بمقارنة الجينومات لاستنباط التشابه والاختلاف وغيرها من الأغراض.

وعلى غرار الدراسات الجينومية ظهر مصطلح مكنون النسخ Transcriptome ومجال دراسته Transcriptomics ويبحث في دراسة ما ينتج عن الانتساخ Transcription وهو RNA للمعلومات الوراثية للتعرف على التعبير الجيني ودور بعض تواليات DNA التنظيمي. اما المكنون البروتيني Proteome والمجال الذي يعنى بدراسته دراسة مكنون البروتين Proteomics فهو يدرس منتجات الجينات وتواليات DNA التي يشفر لها عبر المرور بمكنون النسخ (RNA). وتمثل دراسات المكنون البروتيني العديد من الجوانب بدءاً من تحدد تواليات الحوامض الامينية في الببتيدات او البروتينات وما تضيفه من صفات وفعاليات على البروتينات الناتجة وغيرها من الجوانب.

ومصطلح مكنون الايض Metabolome ونطاق دراسته Metabolomics فيهتم بدراسة التحليل الكمي والنوعي لكل المتأيضات Metabolites في الخلايا والأنسجة والسوائل الحيوية، وهناك مصطلح قريب منه Metabonomics والذي يعمل في التوجه نفسه ولكن في الأنظمة المعقدة ومجال كشف الأدوية.

والحقيقة ما ذكر أعلاه يعد الجوانب الرئيسة في المستجدات التي رافقت المعلوماتية الحيوية. ولكن كل المصطلحات يمكن ان تذكر بشكل كامل مثل دراسة الفسلجة Physiome والذي يعنى بالجوانب الفسلجية و ORFome الذي يعنى بدراسة أطر القراءة المفتوحة (المناطق المشفرة للبروتينات من الجينوم) وهناك العديد من هذه المصطلحات لا مجال لذكرها.

تطبيقات المعلوماتية الحيوية

من الواضح ان تطبيقات المعلوماتية الحيوية واسعة وهي تتسع يوماً بعد آخر، وفي التالي ذكر لبعض الجوانب:

- في مجالات الدراسات الأكاديمية أصبحت ضرورة لابد منها، فهي يمكن ان تشارك في تحليل تركيب الحوامض الامينية والبروتينات، اذ يمكن دراسة التركيب بالمقارنة وعمليات الاصطاف Alignments لغرض تصنيفها والتنبؤ بوظائفها بالاعتماد على التركيب فضلاً عن تحديد التحليل الوظيفي مثل تحديد نسق التعبير الجيني Gene expression profile وتحديد

الممهدات Promoters والتنبؤ بتداخلات البروتينات مع بعضها، فضلا عن تحديد مواقع البروتينات في الخلايا، ومن كل هذه المعلومات يمكن تصميم المسارات الايضية ومحاكاتها. وهذا يكون منوطا بمعرفة ان المجالات المذكورة أعلاه لا يمكن ان تتم بمعزل عن بعضها وانما تتداخل مع بعضها وبالتالي تحتاج الى تصنيف العمليات وتجميعها مع بعضها مثل عنقدة التعبير الجيني Gene expression clustering والذي يحتاج الى إيجاد شجرة علاقات تطويرية معتمدة على تواليات الجينات.

وبالإضافة الى ذلك يلاحظ ان تداخل الفعاليات للبروتينات يحتاج الى التعرف على الممهدات والمناطق المشفرة وغير المشفرة (المنظمة) ومعرفة كل هذه التداخلات والشبكات يمكن ان يفضي الى مجالات للتطبيقات الأخرى.

• التطبيقات في مجال تصنيع الأدوية التي تعتمد على معرفة ارتباط البروتينات بالربائط Ligands وكذلك التراكيب ثلاثية الأبعاد للبروتينات 3D-structures وغيرها من المعلومات التي سهلت وقصرت الطريق لتطوير الأدوية وبكفاءة أعلى وبأقل ما يمكن من التأثيرات الجانبية. حيث تستخدم المعلوماتية الحيوية ببرامجها المختلفة لتحديد الأهداف غير المستغلة في الأحياء الممرضة لتكون أهدافا للأدوية كما في طريقة الطرح الجينومي Genomic subtractive approach ، اذ تتم عمليات اصطفاف لجينومات الأحياء المرضية مع جينومات المضيف، ومعرفة المناطق او التواليات غير المتشابهة ليتم العمل عليها، بدلا من استعمال الطرق التقليدية التي تطول الى سنوات لاكتشاف دواء واحد.

• استخدمت المعلوماتية الحيوية وتستخدم في مجالات الطب الجنائي والتي أصبحت مقبولة في المحاكم والجهات المختصة بعد إخضاع التواليات الى عمليات إحصائية صارمة، سواء في تحديد الجرائم او إثبات النسب.

• تستخدم المعلوماتية الحيوية في مجالات التصنيع الحيوي او التقنيات الحيوية ، لإيجاد أفضل السلالات المنتجة وأفضل الظروف لعمليات الإنتاج (كما مر ذكره).

• واستخدمت وسائل المعلوماتية الحيوية بشكل فاعل في مجالات الزراعة ، اذ ان وجود قواعد البيانات لجينومات النباتات وتحاليل التعبير الجيني لها لعبت دورا مهما في تطوير أنواع وضروب مختلفة من المحاصيل التي تكون لها إنتاجية عالية او تكون ذات مقاومة عالية للظروف البيئية القاسية او مقاومة للإصابة بالأمراض والآفات وغيرها من الصفات التي يرغب بها المستهلك.

• أدت المعلوماتية الحيوية الى ثورة في مجال العناية الصحية والشخصية، ومن التوقعات في هذا المجال ان يقوم الطبيب بتحليل تواليات جينوم المريض بالسرعة الممكنة واستعمال وسائل المعلوماتية الحيوية لتحديد وتشخيص الطفرات المؤذية ومثل هذا سيساعد في التشخيص المبكر ومن ورائه المعالجة المبكرة.

المحددات في المعلوماتية الحيوية

بعد ما ذكر من أطر عامة حول المعلوماتية الحيوية وقبل الخوض في التفاصيل التي سيتناولها الكتاب، هناك بعض الحقائق التي تحد من استثناء استعمال المعلوماتية الحيوية. فمجال المعرفة هذا لا يزال في بداية الطريق برغم التوسع الهائل والإمكانات البشرية التي تصب فيه والموارد المالية المرصودة له ، فالمعلوماتية الحيوية وإن كانت تعتمد بشكل كبير على التجارب المختبرية ولكن يجب أن يبقى في الذهن أن هذه التوجهات يعترضها بعض الاستقلال، فالتجارب المختبرية تزود بالمعلومات الخام والأولية لتجرى عليها المعالجات في الجانب الثاني الذي يقوم بإعطاء التفسيرات الملائمة للمعلومات التجريبية وبالتالي وفي معظم الأحيان تحت مخرجات التنبؤات إجراء المزيد من التجارب لإثباتها. كما أن مسألة الاحتمالية والتنبؤ أو الحدس التي فرضت واقع وجود الإحصاء بشكل كبير تمثل أحد الجوانب التي تؤدي إلى انبهات صورة المعلوماتية الحيوية فضلا عن حقيقة أن التنبؤات التي تقوم بها وسائل المعلوماتية الحيوية هي ليست برهان موثق إلى درجة كبيرة لأي فكرة وعليه فهي لا يمكن أن تحل محل طرق إجراء التجارب المختبرية التي تقوم باختبار الفرضيات.

وعليه فإن العديد مما تقدمه المعلوماتية الحيوية يكون غير قادرا على عكس الحقيقة، فبعض الأحيان تعطي تنبؤ لا يمت إلى مجال علوم الحياة بصلة. ويعكس هذا الدقة التي أجريت بها الاختبارات التجريبية وذلك لأن :

- الأخطاء التي تحصل في المجال التجريبي تنعكس على الاستنتاجات فأذن جودة الاستنتاجات تعتمد على جودة العمل التجريبي، فإذا كان هناك خطأ تجريبي في تحديد التواليات أو تهميشها وتوضيحها فإن نتائج Downstream analysis ستكون مضللة، لذا كان الحذر الشديد في العمل التجريبي الذي هو سيد الموقف.

- تعتمد جودة المعطيات والاستنتاجات من وسائل المعلوماتية الحيوية على الحواسيب المستعملة، فهناك العديد من البرامج التي تكون بطيئة التنفيذ باستعمال الحواسيب العادية لذا يتم اللجوء إلى استعمال برامج سريعة، وعلاقة السرعة العكسية مع الدقة معروفة.

- نوعية البرامج لها دخل كبير في التقليل من كفاءة المعلوماتية الحيوية فاليوم هناك عشرات وعشرات البرامج لمختلف الأغراض ولكنها يمكن أن تعطي مستوى من الأخطاء غير مسموح به.

ومن هنا كانت التوصيات صريحة باستعمال أكثر من برنامج لتوثيق النتائج، وكذلك وجوب دراسة الأسس التي يعمل بها البرنامج ومدى ثقلها الحيوي لتحديد الصفات ، فالعديد من البرامج المعلن عنها وعن صفاتها الخارقة قد لا تكون كذلك.

ومن هنا يتضح أن مجال المعلوماتية الحيوية في الوقت الحاضر هو ليس ناضجا بشكل كبير ولا يزال يحتاج إلى الجهد.

ما يحتاجه العامل في مجال المعلوماتية الحيوية (Bioinformatician)

نظرا لاختلاف الحقول التي تضمها المعلوماتية الحيوية كانت أنجح الكتب والمعطيات هي التي تشمل اختصاصات مختلفة لتساعد في التوصل الى لغة علمية مشتركة يمكن التفاهم بها، وهذا التوجه في جمع المفاهيم والآراء نابعا من مشاكل التواصل بين العاملين في علوم الحياة اللذين يعتمدون على التجارب المختبرية والعاملون في مجال الحاسوب اللذين يريدون ان يتم كل شيء وفق برامجهم على الحاسوب وهذا غير ممكن لطبيعة علوم الحياة الخاصة بالتفاعلية مع البيئة سواء الخارجية او الداخلية فضلا عن تعاملها مع التأريخ الحيوي للخلايا والكثير الكثير من المؤشرات التي لا تزال دراساتها في المهد.

ومتلما ذكر في مستهل الموضوع ان المعلوماتية الحيوية هو حقل يضم اكثر من مجال، ولكل مجال مختصوه ، فالعاملين في مجال البرمجة يحتاجون الى أسس في علوم الحياة ليكونوا على بينة عند وضع البرامج. اما العاملين في مجال علوم الحياة فلا بد ان يكونوا على اطلاع بسيط على لغات البرمجة التي تكتب بها البرامج وعمليات تحديد تواليات الجينومات والبروتينات وعمليات صف التواليات Sequence alignment ، والأسس المستعملة في تسجيل درجات الترجيح Scoring، وأسس عمل كل برنامج، فضلا عن الإلمام بالنواحي الفيزيائية والكيميائية وان لم تكن بعمق معرفته في مجال علوم الحياة. ومن الواضح ان العاملين في هذا المجال لا يمكن ان تكون معلوماتهم متكاملة في كل حقل من الحقول التي تشملها المعلوماتية الحيوية ولكن كل من جانبه يجب ان يلم بالجوانب الأخرى ، كما هو الحال مع التقنيات الحيوية المستعملة في إنتاج مضاد حيوي ما، فالمختص في مجال الأحياء المجهرية ليس بالضرورة ان يكون ملما بشكل كبير بالهندسة الكيميائية وبناء المخمرات ولكن لابد من المشاركة للوصول الى الانجاز متكامل.

الفصل الثاني

وسائل المعلوماتية الحيوية

محاوّر المعلوماتية الحيوية
مهام المعلوماتية الحيوية
أساسيات المعلوماتية الحيوية
البرامج Programs
صیغ الملفات المستعملة في المعلوماتية الحيوية
مراكز المعلوماتية الحيوية
قواعد البيانات Databases
أنظمة الاسترداد Retrieval systems

وسائل المعلوماتية الحيوية

ذكر في مقدمة الكتاب ان المعلوماتية الحيوية هي مجال دمج للمعلوماتية وعلوم الحياة والإحصاء الحيوي، وهذا يتطلب التنسيق والترتيب للوصول الى أفضل النتائج. والمحرك الرئيس في عمليات الدمج هذه قد يكون أكثر من هدف وتأتي الجوانب الاقتصادية في المقدمة لتشبع رغبات الإنسان في التحري عما يفيد وما يضره. والبحوث الحديثة في علوم الحياة تعتمد على تقنيات المعلوماتية في ظل عصر ثورة المعلومات، وهذه تحتاج الى تخطيط دقيق يوضع مسبقاً لغرض توحيد ووضع المقاييس لتقييم العمل التجريبي. فمثلا العاملون في مجال الوراثة يضعون في حساباتهم الاحتياجات اللازمة للقيام بالعمل ثم تقيمه ، ولذلك تبني الأسس للحاسوب منها Hardware والبرامج Software للمساعدة في العمل لان هذه المتطلبات هي تقنيات البحث او لوازم البحث المستعملة لأداء عدة أغراض منها عمليات الكشف او القياس او السيطرة او غيرها من الأغراض التي تهدف إليها البحوث.

محاور المعلوماتية الحيوية

تتدرج أهداف ومهام المعلوماتية الحيوية في ثلاث محاور رئيسية:

- المحور الأول: تنظيم البيانات بطريقة تسمح وتمكن الباحثون من الوصول إليها مع إمكانية التحديث والإضافة عليها. وبينما تعد تهيئة البيانات من المهام الأساسية، فان خزن المعلومات في قواعد بيانات تعد بدون فائدة ما لم يتم تحليلها.
- المحور الثاني: تطوير الأدوات والمصادر التي تساعد في تحليل البيانات. فمثلا عند مقارنة توالي بروتين معين مع بروتين آخر الذي قد يحتاج الى أكثر من برنامج، فان النتائج تحتاج الى التأمل في معنوية التطابق البايولوجي. ولغرض تطوير مثل هذه المهام يجب الاستناد الى رأي الخبير في الجانب النظري في الحساب ثم المرور بفهم الجانب البايولوجي.
- المحور الثالث: استخدام الأدوات لتحليل البيانات ولتفسير النتائج الخاصة بالجوانب البايولوجية وبأسلوب واضح ومفهوم.

مهام المعلوماتية الحيوية

كثيرة هي المهام المنوطة بالمعلوماتية الحيوية ومختلفة لذلك كانت طرق أدائها وبرامجها مختلفة ومتطلباتها من خزن المعلومات مختلف أيضا الذي يتمثل بقواعد البيانات الموضوعة أصلا او التي يستلزم إنشاءها لإغراض معينة. وفيما يلي بعض هذه المهام (التي ستذكر بشكل مختصر ثم يأتي تفصيل بعضها فيما بعد)

عمليات التجميع Assembly: تقوم بهذه المهمة بعض برامج الحاسوب التي تعمل على جمع المقروءات Reads الناتجة من عمليات تحديد توالي الجينومات بطريقة التشظية Shotgun sequencing وتوضع بشكل متجاورات Contigs متداخلة، ثم يصار الى تشذيبها وحذف المتكرر منها للوصول الى التوالي المقبول.

إيجاد القطيفات Motifs: وهذه تمثل في العادة تواليات غير مشفرة وانما تقع ضمن عناصر التنظيم للجينات والتعريف بها يكون مهما من ناحية تحديد وظائف الجينات اذ ترتبط ببعض البروتينات المنفذة للوظائف مثل عوامل الانتساخ.

صف التواليات Sequence alignment: تعد من المهام الرئيسية التي تقوم بها المعلوماتية الحيوية وتقع هذه المهمة ضمن مجال دراسات الجينومات أو دراسة البروتينات المقارنة وتؤدي أكثر من وظيفة.

عمليات التهميش والتوضيح Annotation: وتكون ضرورية لمعرفة الكثير عن العنصر الذي تطاله المعلوماتية الحيوية ، إذ أن تحديد التوالي لوحده لا يزود بأي معرفة ما لم يلحق به بعض التعليق والتهميش مثل نوعية التوالي تحت الدراسة ومكان وجوده والأحياء التي تحوي عليه وغيرها من المعلومات.

إيجاد العلاقات التطورية Phylogeny: وتكون مهمة لفهم العلاقات الوراثية والتطورية بين الأحياء ، إذ تساعد المعلوماتية في إيجاد الأبعاد الزمنية لتطور الأحياء بواسطة الرسوم أو إيجاد المسافات وغيرها من الوسائل التي سيتم تفصيلها في مواقعها الخاصة.

تحليل وإيجاد الجين: تقوم بعض برامج المعلوماتية الحيوية بإيجاد التوالي المقدم أو المستعلم عنه Query sequence وتحديد فيما إذا كان جين أو قطعة من جين، وكذلك إمكانية تحديد موقع الجين على الكروموسوم . فضلا عن إمكانية تحديد ان توالي الاستعلام قد يكون ممهدا" أو غيره من التراكيب بالاعتماد على توالي DNA.

التنبؤ بتركييب العناصر: وهذه تشمل التنبؤ بتركييب DNA أو RNA أو البروتينات وأصنافها. وتتضمن هذه المهمة تحديد مواقع القطع في تواليات الحوامض النووية للعديد من الأغراض أبسطها إنشاء خرائط التقيد Restriction maps. والبحث عن وجود المكررات والمناطق غير المشفرة.

وتتضمن أيضا إيجاد بعض التراكيب الخاصة مثل أطر القراءة المفتوحة **إجراء التحليلات للمواد الحيوية:** وتشمل البحث عن التواليات المتشابهة بين DNAs مختلفة أو RNAs أو بروتينات. أو البحث عن الحوامض النووية مقابل البروتينات التي تنتجها، أي تحليل تواليات DNA قبل الترجمة.

تصميم البوادي Primer design : وتعد من المهام الضرورية في الوقت الحاضر، إذ توفر وسائل المعلوماتية الحيوية مجالا واسعا لتحديد البوادي لإجراء عمليات كثررة DNA بطرق PCR مثلا.

التنبؤ بالتراكيب الثانوية والثلاثية ومواقع الجزيئات الحيوية: تحديد التراكيب الثانوية والثلاثية لبعض الجزيئات الحيوية مثل البروتينات أو الحوامض النووية تعد أساساً لتحديد وظائفها، وكذلك إيجاد مواقعها في الخلية يوفر الفرصة لمعرفة وظائفها.

استرداد أو استرجاع البيانات Retrieval of data : توفر المعلوماتية الحيوية الكم الهائل من البيانات المخزونة في قواعد البيانات ويكون في كثير من الأحيان استرجاع هذه البيانات لإغراض شتى، وخاصة استرجاع التواليات والتعرف على التواليات من حيث موقعها في الجين أو وظيفتها أو أي معلومات موجودة ضمن عملية التهميش المرافقة للتواليات.

مهام تتعلق بالنواحي الفيزيائية: تضطلع المعلوماتية الحيوية في إيجاد التراكيب الجزيئية وما يكملها مثل التنبؤ بالطاقة اللازمة لعمليات التداخل والتوزيع الفيزيائي أو المكاني للجزيئات وإمكانية الارتباط بين الجزيئات والتي تستخدم في العادة لتصميم الأدوية.

البحث عن المراجع Literature searching: توفر وسائل المعلوماتية الحيوية قواعد تضم العديد من المراجع المتعلقة بالجوانب الحيوية أو الإحصائية أو الدراسات الحاسوبية *In silico* studies.

وفضلا عما ذكر من المهام الرئيسية المذكورة أعلاه هناك بعض المهام الخاصة مثل البحث عن تغايرات النيوكلوْتيد المفرد SNPs واستبدال القواعد النتروجينية التي تحت الأمراض وغيرها من المهام التي ستدرج في مواقعها.

ويتفاوت استخدام الباحثين للمهام ، وفي أكثر من دراسة تضمنت مسح عن المهام التي تستخدم فيها المعلوماتية الحيوية (لغاية 2004) أشارت الى النسب الموضحة في الجدول 1.2:

جدول 1.2 : المهام المنوطة بالمعلوماتية الحيوية ونسبها

النسبة المئوية	المهام
8.9	تشابه الحوامض النووية
12.4	تشابه البروتينات
1.9	ترجمة الحوامض النووية الى بروتينات
2.9	البحث عن تواليات DNA غير المشفرة
9.3	البحث عن تواليات غير مخصصة
11.1	البحث عن القطيقات الوظيفية
8.6	استرداد التواليات
6.7	عمليات اصطفاف التواليات المتعددة
6.0	خرائط التقييد
4.4	التنبؤ بالتراكيب الثانوية
4.4	ترجمة DNA
3.8	تصميم البوادي
3.5	تحليل أطر القراءة المفتوحة
3.2	تحليل البروتينات
3.2	البحث عن المراجع
2.5	تجميع التواليات
2.2	إيجاد مواقع الجزيئات
5	دراسات مختلفة

وبعض المهام ولكثرة استعمالها أصبحت مهام ثانوية تؤدي عند تأدية مهام رئيسية ويتضح مما ذكر أعلاه ان المعلوماتية الحيوية تحتاج الى عمليات ترتيب وتنظيم وهذا يحدد أساسياتها المعلوماتية الحيوية.

أساسيات المعلوماتية الحيوية

الأساسيات المعنية تكون لازمة لإدارة التحليل والتعامل مع المجاميع الكبيرة من البيانات الحيوية، ويمكن ان تعرف هذه الأساسيات بثلاث من المكونات التي تكون لازمة في هذا المجال.

- تهيئة قواعد بيانات تسمح بخرن وإدارة وإضافة وتجديد البيانات الحيوية.
- تطوير الخوارزميات والبرامج والإحصاءات لتحديد العلاقات بين البيانات واتخاذ القرارات الخاصة عند التعامل مع مجاميع كبيرة من البيانات.
- التعريف بمفردات النقطة الثانية واستخدامها لتحليل وتفسير الأنواع المختلفة من البيانات الحيوية والتي تشمل في العادة تواليات DNA و RNA والبروتينات بمختلف تراكيبيها لغرض توظيفها للتعريف بهذه التواليات مثل تحديد نسق التعبير الجينية وغيرها من الأغراض.

الدوريات والنشرات الخاصة بالمعلوماتية الحيوية

نظرا لأهمية حقل المعلوماتية الحيوية فقد تخصصت بعض الدوريات او المجلات بالاهتمام بالموضوع ونشر كل ما يتعلق به ، فضلا عن ان المجلات العادية الأخرى تظهر أهمية المعلوماتية الحيوية في البحوث الخاصة بها وخاصة المجلات التي تعنى بالدراسات الجزيئية. وقد صدرت عشرات وربما أكثر من الكتب التي تتناول جوانب المعلوماتية الحيوية المختلفة .

البرامج Programs

يتم تنفيذ مهام المعلوماتية الحيوية بواسطة برامج الحاسوب، لذا فهي أحد الوسائل التي يتم بواسطتها تنفيذ مهمة معينة للحصول على نتائج مصنفة بعد إدخال البيانات المراد معالجتها . وتسمية البرامج في الغالب تدل على المهامة التي تقوم بها مثل PSORT يدل على ان البرنامج يتعامل مع البروتينات وتوزعها في المناطق المختلفة من الخلية، وبرنامج TMHMM (Transmembranous Hidden Markov Model) يشير الى ان البرنامج يبحث ويساعد في معالجة البروتينات التي تكون مرتبطة او مستعرضة الأغشية ، وفي العادة تشير إيقونات رمز البرنامج الى ما يدل عليه.

وبذا يلاحظ ان البرنامج سواء في المعلوماتية الحيوية او غيرها من برامج وأنظمة الحاسوب تستعمل الكتابة الصورية او الرمزية كما في الكتابات القديمة مثل الكتابة السومرية او ما مستعمل في حضارات مصر القديمة. وتظهر في تسمية الملفات ما يشير الى نوع الملف مثل ظهور doc للملفات النصية المعالجة بنظام او برنامج Word او xl عندما يكون البرنامج ناتجا من برنامج Excel الأكسل . ويفضل غير المتمرسين باستعمال الحاسوب والتنقل بين البرامج والواجهات الى استعمال برنامج او وسائل Software متعددة الأغراض، وتمثل هذه المجموعة القطاع الأكبر من المستخدمين المعلوماتية الحيوية وهم العاملين في مجال علوم الحياة ويكونون بمثابة المستخدم الأخير للبرنامج. وهناك العديد من البرامج قد لا تكون مفيدة نظرا لمحدودية الاختيارات فيها مثل الحاجة الى صيغة محددة لإدخال المدخلات (Inputs) وتكون مخرجاتها محددة ولا تحتوي على رسوم توضيحية، او تحتاج الى ساعات لغرض التمرين عليها او تنفيذها فضلا عن ان تنفيذ بعضها يحتاج الى خطوات تحضيرية للوصول الى الصيغة الملائمة كمدخل للبرنامج.

ولذلك يلاحظ ان Spreadsheet software هي الأكثر استعمالا وفائدة للعاملين في مجال علوم الحياة، فهي توفر إمكانية إجراء العمليات الإحصائية، وكذلك وسائل لفرز وترتيب البيانات فضلا عن توفيرها للرسوم والمخططات ضمن المخرجات (Outputs). ومن الأمثلة عليها برامج نظام الإكسل الذي يستعمل لحساب مؤشرات حركات الإنزيمات، وتلوين الحوامض الأمينية بألوان محددة وكذلك يدخل ضمن برامج التنبؤ بتركيب الدومينات Domains في البروتينات وغيرها من الأغراض.

ومن ناحية وقت أداء البرنامج لمهمة محددة فالبعض تكون سريعة التنفيذ والبعض الآخر يأخذ وقت طويل قد تصل الى ساعات او بعض الأحيان أيام وذلك اعتمادا على نوعية البرنامج والمهمة التي يقوم بها ، لذلك تدخل بعض البرامج الوسيطة لتعجيل سرعة الأداء. كما ان تنفيذ بعض البرامج يكون غير ممكنا باستعمال الحواسيب العادية او الشخصية ويحتاج الى حواسيب عملاقة وكفاءة.

لغات البرمجة في المعلوماتية الحيوية

يوجد العديد من لغات البرمجة وهي في تطور مستمر اذ يظهر البعض ويختفي البعض الآخر، لذلك لا يمكن تخطيطها بشكل كامل في كتاب يهدف الى استعمالها فقط كوسيلة وليس هدفا لتوضيحها. فالأساسيات او القواعد التي يقوم بواسطتها الحاسوب لأداء مهمة تسمى الخوارزمية Algorithm لإعطاء النتائج ، وهذه تزود للحاسبة بلغة برمجة والحاسبة تقوم بتفسير البرنامج المكتوب بلغة محددة لإعطاء نتائج علمية ، واللغة اللازمة للاستعمال تعتمد على عدد من العوامل في الاختيار وعليه تكون وسائل البحث والدراسة متباينة.

هناك العديد من لغات برمجة الحاسوب لكتابة البرامج ولكن أكثر هذه اللغات استعمالا في مجال علوم الحياة هي SQL (Structured query language) و C ، C# ، C++ ، Perl ، Python ، Java وغيرها. والمعروف لدى المبرمجين طبيعة كل من هذه اللغات، اذ ان بعضها يكون نصي كما في Python و Perl تسمى Scripting language في حين ان اللغات الأخرى تستعمل الرموز، وبعض الأحيان تكون اللغة بين الاثنين كما في SQL التي تحجز بعض الإشارات بها مثل (>) او (:). فضلا عن احتفاظها ببعض العبارات الشرطية عند تكوين قطع مختلفة من البرنامج او خلق العروات Loops في البرنامج.

وأهم المؤشرات لاختيار اللغة لكتابة برنامج ما هو الدقة والسرعة وما يمكن ان يحجز البرنامج من ذاكرة الحاسوب، فالسرعة مطلوبة نظرا لحجم مجاميع البيانات في قواعد البيانات لذلك احتاجت العملية الى السرعة، وفي هذه الحالة قد يكتب البرنامج بلغة، وجزء منه بلغة أخرى لغرض التعجيل.

اما المؤشر الآخر في تفضيل استعمال لغة على أخرى هو توفر المكتبات لها فمثلا هناك المكتبات تساعد اللغات المستعملة وتشمل على تعليمات حول إمكانية الحصول على البيانات مثل التواليات وكذلك إرشادات حول قراءة الملفات ذات الصيغ المختلفة مثل صيغ FASTA.

وفي العموم فان البرامج يعتمد أدائها على اللغة التي كتب بها البرنامج، ونظام التشغيل مثل استعمال نظام Windows او Linux ، لذا فمن المتوقع ان يكون البرنامج المكتوب بلغات مختلفة او ينفذ باستعمال أنظمة تشغيل مختلفة ان يظهر اختلافات مهمة في السرعة والكفاءة.

وكذلك يعتمد على قواعد البيانات ونوعيتها وحجمها وفي هذا المجال قد تكون قاعدة البيانات المشدبة (nr (non-redundant هي الأفضل نظرا لتشيبيها وإزالة المكررات منها. لذا فان اختيار اللغة يعتمد ايضا على المكتبات المتوفرة لها.

وقد تساعد اللغة التي يكتب بها البرنامج في زيادة دقته في مجال محدد فمثلا النصوص الكتابية Scripts تساعد في استخلاص البيانات من ملفات البيانات الكبيرة وبالتالي تحسن من التنفيذ كما في استعمال لغة Python و Perl فضلا عن استعمالها في ربط البرامج، فبعض البرامج لا يمكن ان تربط بسهولة لذلك توضع بعض النصوص الكتابية للعمل على التلاقي بين البرامج لأداء الأجزاء المتتابعة من عملية التحليل.

كما ان تنفيذ البرنامج يعتمد على إصدار البرنامج Version اذ غالبا ما يتم تحسين البرنامج وإجراء الإضافات عليه.

وكل من اللغات المذكورة لها طريقتها في العمل وتصلح لأداء بعض المهام دون غيرها، لذلك فان النتائج التي تظهر بأي لغة تعتمد على نوعية الملف المقروء. كما ان البرنامج الواحد يمكن ان يكتب بأكثر من لغة لغرض التسريع فمثلا برنامج BLAST يتلقى ملايين الاستعلامات يوميا لذا يحتاج الى سرعة فائقة لتنفيذ المهمات.

ويمكن التوقف قليلا عند مجموعة لغات Markup languages والاهم فيها هو (Hyper HTML Text Markup Language) والتي كتبت بها صفحات شبكات الانترنت. وكذلك XML (eXtensible Markup Language) وهذه لها أحكام او أوامر خاصة ومعطياتها تشبه HTML إلا ان لها تطبيقات عديدة اما HTML فتكون خاصة بوصف صفحات الانترنت وتكون الملفات الناتجة منها هشة وعرضة للتغيير.

وهناك لغات أخرى قابلة للاستعمال في مجال علوم الحياة لكل منها ميزات والتي تلبي احتياجات جوانب خاصة، لذلك فعند اختيار اي برنامج للاستعمال والتطبيق لابد من ملاحظة اللغة التي كتب فيها، ولا مجال لذكرها جميعا.

التعامل مع البرامج

أكثر المهام والمتطلبات المشمولة بالمعلوماتية الحيوية وضعت لها برامج خاصة وهذا ناتجا من زيادة المعلومات حول الجزيئات الحيوية مثل المناطق غير المشفرة في DNA او غيرها من الجوانب، وزيادة المعلومات رافقتها الزيادة في تعقيد الطرق والرياضيات والعمليات الإحصائية. وفي العادة ولحد الآن لا يوجد ما يسمى بالبرمجة المتوازية Intrinsically parallelizable ولذلك فأكثر المهام تحتاج الى إجراء عمليات مستقلة ثم تدمج في الأخير كما في عدم قبول نوعية من الملفات الا بعد تحويل محتوياتها الى صيغ يحددها البرنامج.

وتحتاج البرامج في العموم الى بيانات كمدخلات وهذه يجب ان تكون وفق الطريقة التي كتب بها البرنامج فبعض البرامج لا تقبل مدخلات الا اذا كتبت بصيغة محددة وأكثر الصيغ المقبولة في معظم البرامج هي صيغة FASTA للتواليات والأخرى تحتاج الى ملفات مفصلة مثل صيغ قاعدة بنك الجينات GenBank ، في حين تحتاج بعض البرامج الى أرقام التسجيل Accession numbers للتواليات لنقوم هي بالمعالجة.

اما المخرجات فهي الأخرى تكون مختلفة وفق البرنامج المستعمل، فالبعض منها تكون نصية وترافقها القيم الإحصائية ليكون المستخدم على دراية كاملة لقبول او رفض النتائج او إجراء

تحويل على الطريقة التي يستخدم بها البرنامج. كما ان بعض المخرجات او المعطيات من البرامج يمكن ان تكون بشكل صور او أشكال وترافقها القيم الإحصائية لتسويق القبول او الرفض وفق الخلفية البايولوجية لمستخدم البرنامج.

لذلك فان التعامل مع البرامج قد لا يكون مباشرا وقد يحتاج المستخدم الى عملية التسجيل في الموقع الذي يروم استخدام أحد برامجهم ، كما ان النتائج قد لا يتم الحصول عليها مباشرة Online وإنما تحتاج الى إضافة الاسم وكلمة مرور او العنوان الالكتروني للمستخدم كي يتم إرسال النتائج له بعد معالجتها بالبرنامج. ومن جهة ثانية فان بعض البرامج يمكن تحميلها من شبكة الانترنت بشكل كامل من ملفات رئيسية وأخرى إضافية ليتمكن الباحث من استخدامها بشكل مستقل وقت ما يشاء وهذه هي Offline programs ، ولكن في العادة تكون مثل هذه البرامج لمهام لا علاقة لها بقواعد البيانات العالمية والكبيرة لان مثل هذه القواعد لا يمكن تخزينها على أجهزة الحاسوب الشخصية لغرض الرجوع اليها عند الحاجة.

ومن جهة ثانية فان للاستعمال غير المباشر للبرامج صيغ ملفات مدخلات خاصة كما ذكر أعلاه ، فضلا عن نوعية المعلومات فبعض البرامج قد تكون مخصصة للتعامل مع تواليات DNA فلا يمكن ان تقبل تواليات حوامض امينية ، لذلك لابد من تحويل نوعية البيانات اي استعمال برنامج آخر ليقوم بعملية تحويل البروتينات الى صيغ تواليات للـ DNA ثم استعمال الأخيرة . كما ان بعض البرامج لا يمكن ان تقبل توالي محدد وإنما تتعامل مع أرقام التعريف ID لقاعدة معينة لذلك لابد من البحث عن تعريف التوالي في قاعدة محددة لغرض استعمالها في البرنامج. وبالرغم من كثرة البرامج وزيادة عدد المختصين سواء بالبرمجة او استعمال وسائل المعلوماتية الحيوية لازالت هناك هوة ومحددات لاستعمال وسائل المعلوماتية الحيوية لعل في مقدمتها الاحتياجات الشخصية والتفضيل الشخصي للبرنامج وذلك نابعا من عدم توثيق للطرق المستعملة Lack of procedural documentation ومن هنا كانت التوصيات المتكررة عند تنفيذ مهمة واحدة استعمال أكثر من برنامج وكذلك استعمال برنامج واحد مكتوب بأكثر من لغة للوصول الى أقرب الحالات من الحقيقة ، وعليه فانه لا تزال هناك تحديات كبيرة أمام المعلوماتية الحيوية في المجالات التي يتوخاها الباحثون للركون الى هذا المجال من المعرفة لتوضيح استنتاجاتهم.

ونظرا لتشعب البيانات الحيوية التي تغطيها البرامج لذلك تقسم البرامج في أدائها ويتم تنفيذ ذلك من الخيارات المستعملة في البرنامج ، فمثلا عند البحث عن نواتج ترجمة توالي حوامض نووية الى بروتينات يجب تخصيص نوع الكائن الحي هل هو في مجموعة خلايا حقيقية النواة او بدائية النواة والتي تقسم بدورها الى بكتيريا موجبة لصبغة كرام او سالبة للصبغة او هي من مجموعة الاركيا وذلك لاختلاف الأحياء في تفضيل استخدامها للشفرات الوراثية لإنتاج الحوامض الامينية وغيرها من الاختلافات.

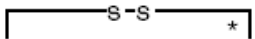
فعند تنفيذ برنامج ما (في بعض الأحيان) يصار الى تحديد تركيب الاستعلامات Queries كأن تجمع المعلومات التي لها صفات مشتركة لتستعمل كمدخل لبعض المهام. وفضلا عن وجود نظام التصفية او الترشيح (Filter) وهذه تحدد بالكلمات المفتاحية Keywords المستعملة مثل رقم التسجيل الذي يعد أفضل صفة انتخابية للمدخل المطلوب، او استعمال النوع Species او غيرها من المعلومات ولذلك فان تحديد المعطيات المطلوب معالجتها قد تدخل ضمن البرامج

ضمن ما يسمى بالمحولات Transformers التي تتعامل مع البيانات المدخلة لتعطي مخرجات محددة ، ولذا فان البيانات المدخلة والمشروطة يمكن ان تؤثر في مخرجات البرنامج وهذه المحولات يمكن ان تأخذ أكثر من مدخل من المعلومات وتمرره على ما يسمى بمرشح التحول وتعطي مخرج موحد ، ويتم ذلك بواسطة البرنامج الذي يحوي على أكثر من مقطع وكل مقطع يعمل على جزء من المعلومات المدخلة ثم توحد لتعطي مخرج واحد الذي يكون بأشكال مختلفة فضلا عن إمكانية تخزين النتائج بأشكال مختلفة ولمدد مختلفة. فالمعلومات الناتجة يمكن ان تستعمل للدراسة مباشرة او يمكن ان تخزن لمدة قصيرة للاستعمال في دراسات مستقبلية ولكن بعض البيانات الناتجة تزال بعد مدة وفي العادة يتم التنبيه الى مدة الخزن ، فالمعلومات الكبيرة يتم تخزينها لمدة دقائق وأخرى يمكن ان تحفظ بشكل دائم.

مخرجات البرامج Program outputs

تختلف البرامج في الأغراض والأسس التي تستند عليها ولذلك تختلف مخرجاتها ولكن الثابت بصورة عامة عند إنشاء البرامج هو الأخذ بنظر الاعتبار القيم الإحصائية كي يتمكن المستخدم من ترجيح مخرجات البرنامج بعد التنفيذ. كما ان البرامج التي تمثل فيها النتائج بصور تكون أفضل ومرغوبة أكثر لأنها تعطي ملخص واضح. ويستفاد واضعي البرامج من كل الإمكانيات المتوفرة في الحاسبات، اذ تستعمل الأشكال المختلفة والألوان للتوضيح، كما ان بعض البرامج يمكن ان تستعمل أكثر من وسيلة لتمثيل النتائج، ومن الأشكال المستعملة:

- الصيغ النصية التي توضع وتهمش بإشارات او ملحقات توضع عادة مع البرامج. الشكل التالي(1.2) يوضح مناطق تكون الأواصر الكبرى بين الثنائية بين ثملات السستين في تواليات التربين التابع للفأر وجراد البحر (Crayfish) وكذلك توضيح الثملات المتشابهة.

		
Mouse	IVGGYNCEENSVPYQVSLNS-----GYHFCGGSLINEQWVVSAGHCYK-----SRIQV	
Crayfish	IVGGTDAVLGEFPYQLSFQETFLGFSFHFCGASINYENYAITAGHCVCVYGGDDYENPSGLQI	

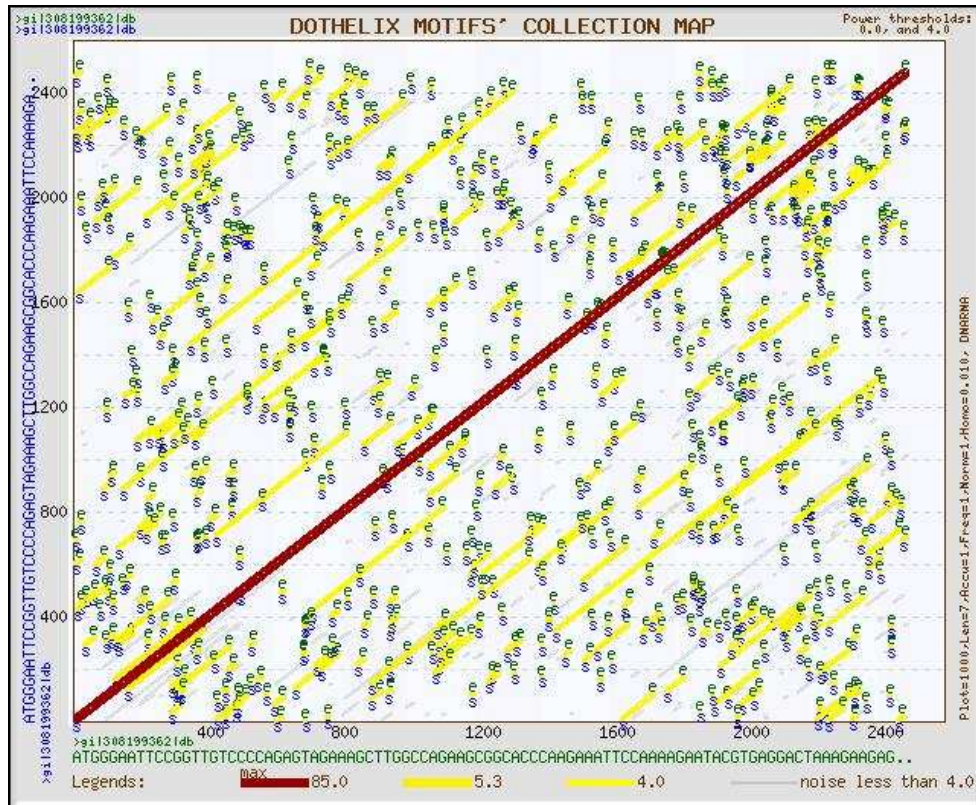
شكل 1.2 : توضيح مواقع تكون S=S ومواقع التشابه في بروتينين

ويمكن ان تكون نتائج الاصطفاف بشكل آخر كما موضح في الشكل التالي(2.2) الذي يوضح اصطفاف أحد البروتينات البشرية مع انزيم Quinone oxidoreductase من بكتريا *E. coli*

Human-ZCr	MATGQKLMRAVRVFEFGGPEVLKLRSDIAVPIPKDHQVLIKVHACGVNPNVETIYIRSGTYS
Ecoli-QOR	-----MATRIEFHKHGGPEVLQA-VEFTPADPAENEIQVENKAIGINFIDTYIRSGLYP
	* * * * *

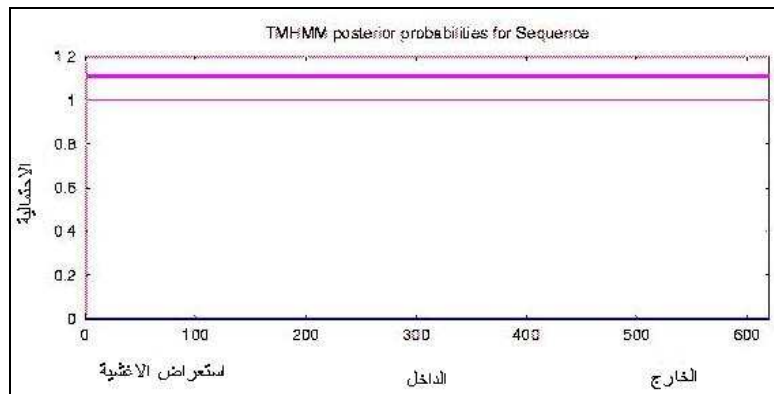
شكل 2.2 : طريق توضيح التشابه لتواليات مختلفة

- ويمكن ان تكون المعطيات بشكل رسم اذ يمكن ان تمثل عمليات الاصطفاف بخطوط قطرية بين التوالين ثم رسم المحصلة كما موضح في الشكل 3.2 :

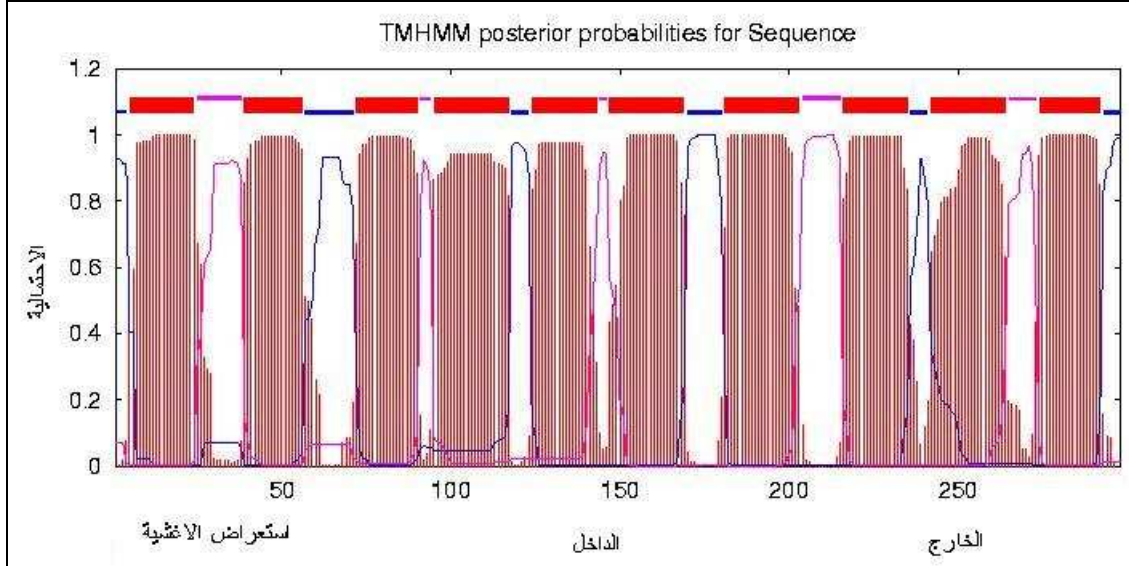


شكل 3.2 : إحدى طرق توضيح اصطفااف التواليات

وتستعمل الرسوم في تمثيل مكان وجود البروتينات ، الشكل (4.2) يمثل أحد البروتينات الدائبة التي ليست لها علاقة بالأغشية، اما الشكل (5.2) فتوضح وضعية البروتينات في الغشاء واستعراضها لكامل الغشاء مع توضيح القيم الإحصائية لترجيح هذه النتائج، في حين ان البروتين نفسه وعلاقته بالأغشية يمكن ان يمثل بصيغ نصية عند استعمال برنامج آخر الذي تظهر فيه احتماليات التنبؤ بالبروتين .



شكل 4.2 : مخرج يوضح احد البروتينات غير الغشائية(الدائبة)



شكل 5.2 : رسم توضيحي لأحد البروتينات المستعرضة للأغشية

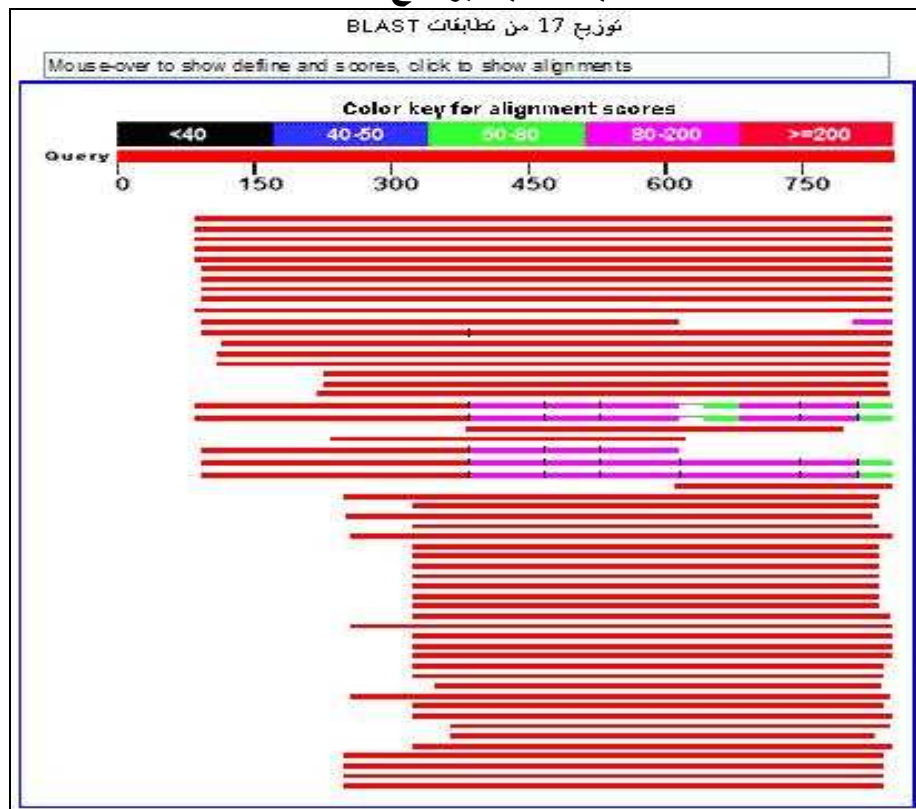
استعمال الألوان

يعد استعمال الألوان من الوسائل المفضلة لأنها تعطي معلومات أكثر كما تم إيضاحه أعلاه واختيار الألوان يجب ان يتم بعناية اذ انه يزيد من التأكيد على نقاط محددة للفهم . ولعمليات التلوين توجد برامج عدة تحوي ضمنا وسائل للتلوين مثل Excel و Matlab و Mathematica ، وفي بعض الأحيان تحتاج عمليات التلوين الى تحويلات تجميلية قبل عمليات النشر والطبع وهناك برامج يمكن ان تقوم بهذه المهمة منها Vector graphics مثل PDF و PostScript والنتائج في هذه الصيغ تعتمد على التمثيل الرياضي للخطوط والنقاط والمنحنيات وغيرها، اما الأنواع الأخرى من معالجات الصور والألوان فعلى سبيل المثال لا الحصر نظام Raster graphics لصور JPG و GIF فتعتمد قيم مصفوفات اللون Colormatrix values التي تسمى Pixels. والعمليات الثانوية مثل تصغير وتكبير الصورة وزيادة وضوحها فيعتمد على طريق المعالجة التي تمت على الصورة ففي حالة Vector graphics فان التكبير يمكن ان يزيد من توضيح الصورة او المحتوى الى حد ما ولكن هذا لا يحصل عند استعمال Raster graphics. وفي الحالتين يمكن خزن الملفات الناتجة بالصيغ الخاصة بكل نظام ويمكن طبعتها فيما بعد في معظم الحالات.

هناك برامج مهمة تستعمل صيغة الرسوم والصيغة النصية سوية مثل برنامج BLAST كما موضح في الشكل التالي(6.2) ويمكن توضيح اي منطقة من التأشير عليها:

Score = 5453 bits (2751), Expect = 0.0 Identities = 4036/4449 (90%), Gaps = 6/4449 (0%) Strand=Plus/Plus					
CDS: Putative 1	1	M Q R S P L E K A S V V S K L F F S W T			
Query	133	ATGCAGAGGTCGCCTCTGGAAAAGGCCAGCCTTGTCTCCAAACnnnnnnmCAGCTGGACC	192		
Sbjct	1	ATGCAGAGGTCGCCTCTGGAAAAGGCCAGCATCTTCTCCAAACTTTTTTCAGCTGGACC	60		
CDS:cystic fibrosis	1	M Q R S P L E K A S I F S K L F F S W T			
CDS: Putative 1	21	R P I L R K G Y R Q R L E L S D I Y Q I			
Query	193	AGACCAATTTTGAGGAAAGGATACAGACAGCGCCTGGAATTGTCAGACATATACCAAAATC	252		
Sbjct	61	AGACCAATTTTGAGGAAAGGATATAGACAGCGCCTGGAATTGTCAGACATATACCATATC	120		
CDS:cystic fibrosis	21	R P I L R K G Y R Q R L E L S D I Y H I			
CDS: Putative 1	41	P S V D S A D N L S E K L E R E W D R E			
Query	253	CCTTCTGTTGATTCTGCTGACAATCTATCTGAAAAATTGAAAAGAGAATGGGATAGAGAC	312		
Sbjct	121	TCTTCTTCTGACTCTGCTGACAATCTGCTGAAAAATTGAAAAGAGAATGGGACAGAGAA	180		
CDS:cystic fibrosis	41	S S S D S A D N L S E K L E R E W D R E			

أ : صيغة نصية لبرنامج BLAST



ب : صيغة مصورة لبرنامج BLAST

شكل 6.2 : الصيغ المختلفة لبرنامج BLAST

- ومن البرامج المتخصصة بمعطياتها الصورية هو برنامج الشعار الصوري WebLogo أو يختصر الى Logo ، والبرنامج يستعمل لعمليات اصطاف التواليات المتعددة (التي

تكون متساوية الأطوال) ليعطي الكثير من المعلومات وهو يوفر سجل واضح وأكثر دقة من الطرق الأخرى لتمثيل تشابه التواليات. ويستعمل لصف تواليات الحوامض النووية وكذلك البروتينات ويوفر ميزات لعمليات الاصطفاف التي تكون من الصعوبة إدراكها، وباعتماد معادلات خاصة وصفة الانتروبي Entropy ثم تحديد أعظم ثبوت للتوالي للموضع الواحد (الذي يمثل المحور الصادي) بالمعادلات

$$\text{Log}_2 4 = 2 \text{ bits}$$

هذه لتواليات DNA و RNA

$$\text{Log}_2 20 = 4.32 \text{ bits}$$

وتمثل الثمالات في هذا البرنامج بالنسبة للحوامض النووية

الكوانين: برتقالي ، الثايمين واليوراسيل: احمر ، السايتوزين: أزرق ، الأدنين: اخضر
اما ألوان الحوامض الامينية فتعطي الألوان اعتمادا على صفاتها الكيميائية ومنها:

- الحوامض الامينية القطبية تكون بلون اخضر وهي (N, Q, C, Y, T, S, G)
- الحوامض الامينية القاعدية وتكون بلون أزرق وتشمل H, R, K
- الحوامض الامينية الحامضية وتكون بلون احمر وتشمل E, D
- الحوامض الامينية الكارهة للماء وتكون بلون اسود وتشمل عدد كبير W, P, I, L, V, A, M, F

وللمستخدم ان يغير من الألوان او يستعمل الأبيض والأسود.

وكل سجل مصور يتكون من حروف الثمالات (استعمال الشفرة الأحادية) الموجودة في التواليات وكل كدس يوضح ثبوت الثمالة في ذلك الموقع المقاس بـ bits ، اما ارتفاع الرمز او الحرف في الكدس فيشير الى التكرار النسبي للثمالة في ذلك الموقع كما موضح في الشكل (7.2).

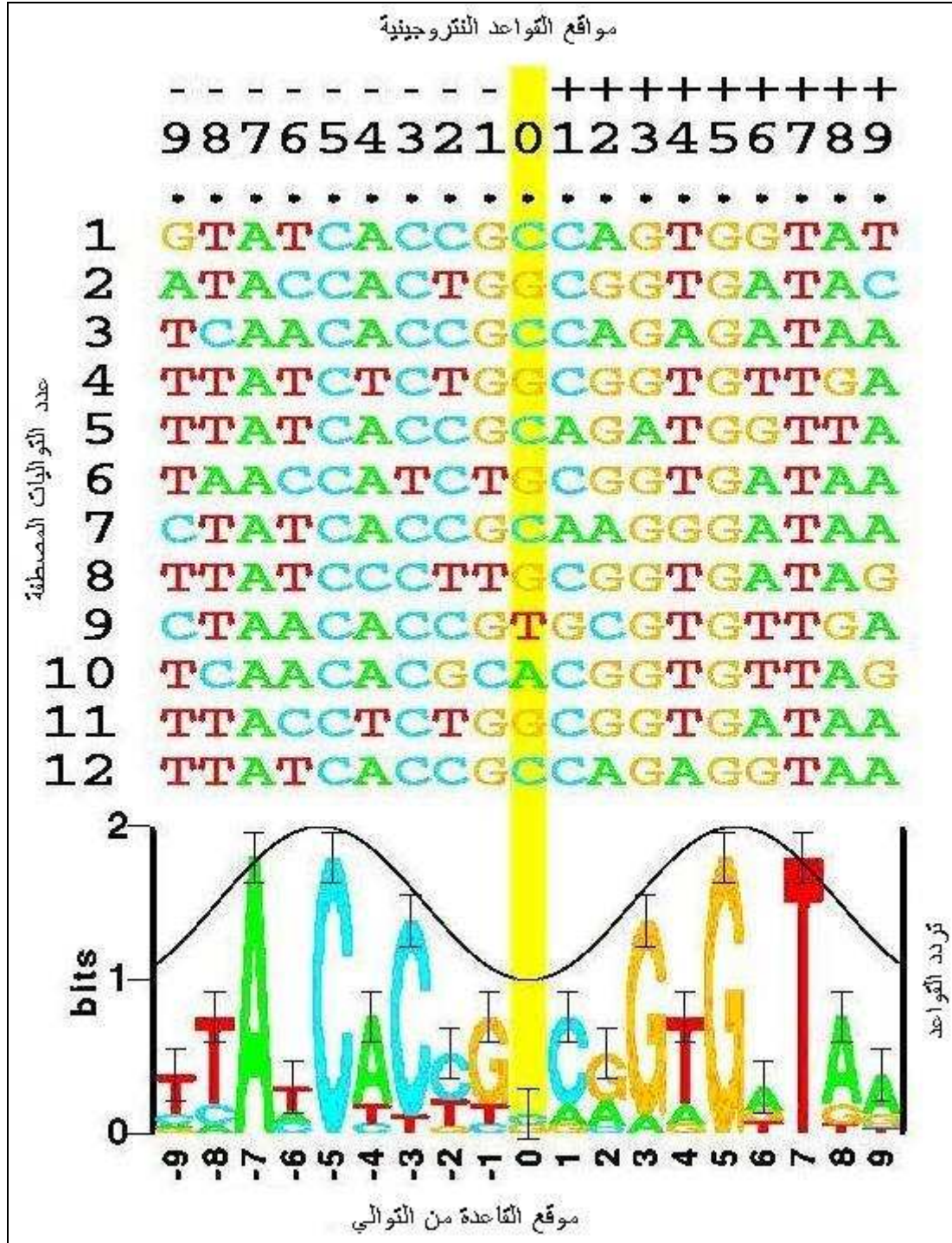
والملفات التي ينتجها البرنامج بصيغة الصور تكون مختلفة مثل Bitmap format او GIF او PNG المناسبة للظهور على شاشة الحاسوب، وكذلك يمكن ان تظهر الملفات بصيغ PDF و EPS التي تكون ملائمة أكثر عند الطبع والنشر.

وهناك العديد من الإمكانيات للتلاعب بالتقرير الناتج من حيث الحجم ودقة الوضوح وإظهار الإحداثيات والمؤشرات الإحصائية مثل إظهار Error bar ، يمكن الاطلاع عليها في واجهة البرنامج والإصدارات المختلفة له.

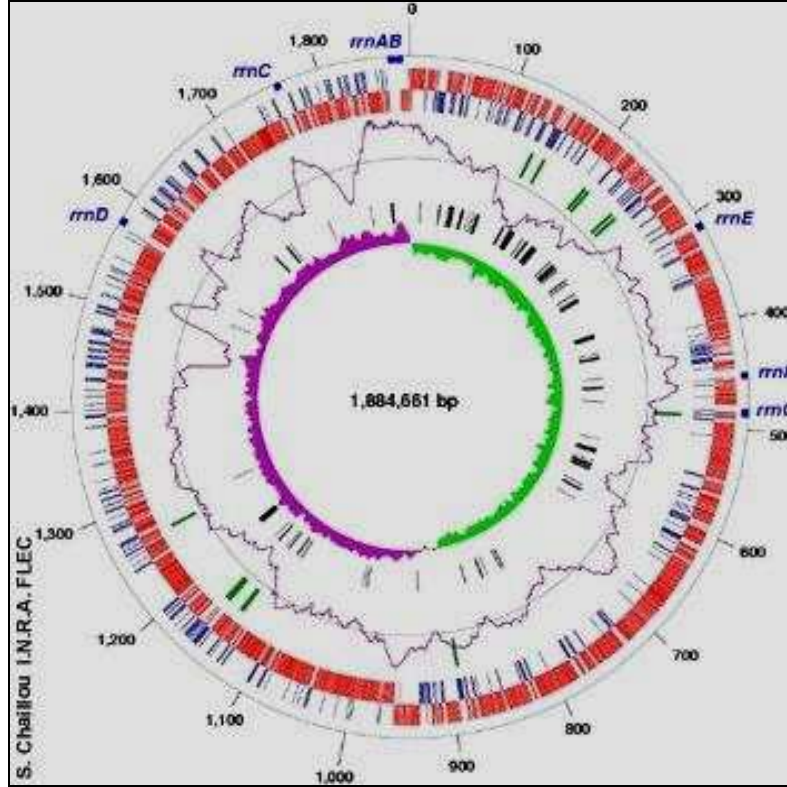
ومن الوسائل المصورة ذات الفائدة الكبيرة هو استعمال أطالس الجينومات

أطالس الجينومات Genome atlas وسيلة مفيدة لعرض كم كبير من المعلومات الخاصة بعناصر DNA الحلقية مثل كروموسومات بدائية النواة وبلازميدات وتتعتمد الأطالس على الألوان وتدرجها وعادة تظهر الأطالس مع مرفق بتدرج الألوان والقيم الإحصائية، ومن المعلومات المرفقة ومنها تقوس أشربة DNA الذاتية او الداخلية Intrinsic curvature مع القيم الإحصائية لها وكذلك طاقة الرص او التنضيد Stacking energy مع القيم الإحصائية فضلا عن توضيح المناطق المشفرة CDS وفيما اذا كانت معرفة ومهمشة ام لا، وكذلك تذكر المكررات في الجينوم وانحراف المحتوى من قواعد السايتوزين والكوانين GC skew

والمحتوى من القواعد AT وسيأتي ذكرها لاحقاً، الشكل (8.2) يوضح خريطة لأحد البلازميدات.



شكل 7.2 : مخرجات برنامج الشعار Logo المختلفة توضح عدد من تواليات العاثي λ



شكل 8.2 : أطلس لجينوم احد البلازميدات

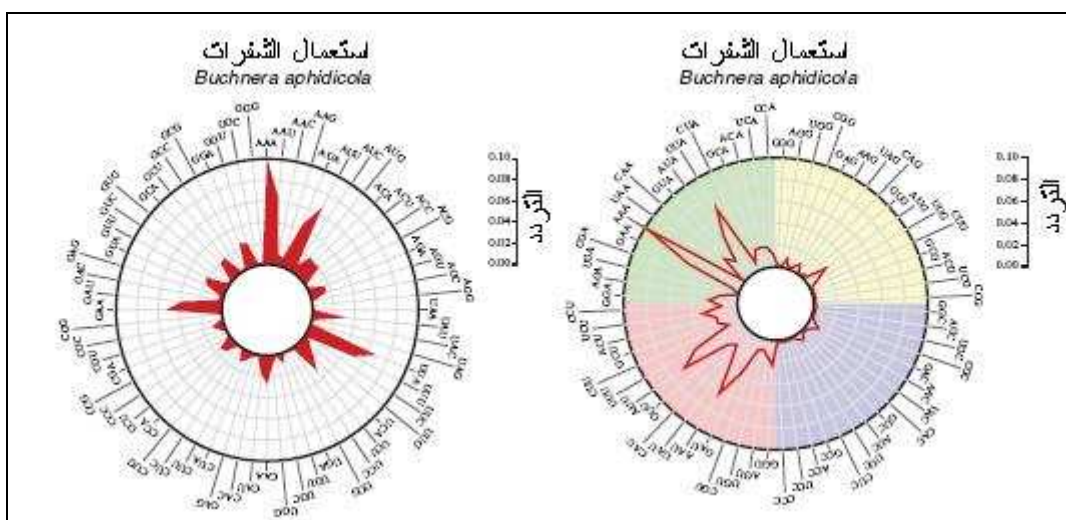
استعمال الشفرات هناك برامج ملونة تساعد في إظهار تفضيل السلالات لاستعمال شفرات معينة للحوامض الامينية ، ويمكن ان يستعمل ما يسمى برسم الوردة Rose plot لتمثيل توزيع الشفرات على الجينومات وتردد هذه الشفرات في الجينوم او يمكن ان تغير بعض مؤشرات البرنامج لتظهر حالة تفصيل الشفرات تحت ظروف مختلفة كما في الشكل التالي (9.2) الذي يوضح الحاليتين لجينومات *Buchnera aphidicola*.

خزن النتائج

يكون الخزن ضروريا للاحتياجات المستقبلية مثل عملية الأرشفة ، التعديل او حتى لإنشاء قواعد بيانات جديدة. وصيغ الخزن المستعملة تعتمد على البرنامج المستعمل، فبعض مالكي البرامج يقدمون صيغ غير شائعة بدلا من الصيغ القياسية او العامة او ملفات مفتوحة للاستعمال. ومثل هذه البرامج الخاصة بالمالك يمكن ان يؤدي الى الاعتماد على برنامج معين للوصول الى البيانات مما يصعب هذه المهمة وكذلك يصعب استرجاع البيانات وتتعدى العملية اكثر عند الحاجة الى ربط اكثر من برنامج حاسوب مختلف. ولذلك فان تحديد صيغ الملفات ضمن الشائع منها له فائدة كبيرة اذ يشجع بيع وانتشار البرامج ذات الأداء الجيد ، وعلى العموم يكون القرار لمالك البرنامج ويكون خاص بكل حالة.

الطريقة الأخرى لعمليات الخزن هي عندما يراد إجراء إضافة بيانات الى قواعد البيانات العلاقية وبهذا يفضل ان تكتب قواعد البيانات الأصلية بلغة SQL، وبمجرد إيجاد الجدول

المراد الإضافة عليه يمكن إدخال البيانات الجديدة وربما حذف غير المطلوب من القاعدة وهنا يجب التذكر ان مثل هذه العمليات تتم تحت أمره نظام إدارة القاعدة (Database DBMS management system)



شكل 9.2 : استعمال الشفرات الوراثية من الأحياء المختلفة تحت الظروف البيئية

التعاون العالمي والتقني في وضع البرامج وقواعد المعلومات

نظرا لضخامة حجم البيانات الحيوية وكثرة البرامج الموضوعة لمعالجتها حصل اضطراب في بعض النواحي لذا تم التفكير في إجراء عمليات التوحيد للمصطلحات والمسميات كما في توحيد تسمية الجينات Gene ontology واستعمال صيغ GFF (General file format) الموحدة فضلا عن اللجوء عن استخدام البيانات الإضافية Metadata وتشمل بعض الجوانب مثل الكائن المشتقة منه التواليات قيد الدراسة ، طول الجينوم محتواه من AT وغيرها من المعلومات الإضافية ، والأخيرة قد لا تكون لها معنى بدون وجود البيانات الأصلية التي تصفها البيانات الإضافية وتكون وثيقة الصلة بأصل البيانات. وان كانت البيانات يمكن ان تعطي معنى دون البيانات الأصلية ، الا ان البيانات الأصلية تكون عرضة للتغيير والتحديث لذا كانت البيانات الإضافية ضرورة للحفاظ على صلب الموضوع الخاص بالبيانات الأصلية ولذلك فان غياب البيانات الإضافية يمكن ان يؤدي الى التباس غير ضروري فمثلا التوالي AAAA يمكن ان يعني توالي الحامض الاميني الالانين في سلسلة بروتين ويمكن ان يعني توالي الأدينين في تواليات DNA لذلك فان وضع معلومة إضافية تشير الى ان التوالي قيد الدراسة هو DNA سيتضح ان "A" المكررة تعني الأدينين، وفي هذا المجال لابد من الرجوع الى قواعد IUPAC الخاصة بالرموز المستعملة.

وقد أتضح مما سبق ان هناك زيادة هائلة في قواعد البيانات التي تخزن مختلف البيانات وهذه تحتاج الى عمليات إدارة وتنسيق وإدامة ، والتفكير في دمج قواعد البيانات ليس بالعمل السهل وذلك لتنوع الدلالات Semantics المستعملة فيها وكذلك التنوع التركيبي لها . وتحتاج القواعد

الى تحديث المعلومات بشكل مستمر من ناحية التركيب او المحتوى وهو المقصود بالإدامة الأنفة الذكر، لذلك طورت بعض الوسائل والمؤسسات لجسر الهوة بين برامج الحاسوب وقواعد البيانات والتي تسمى العميل Agent مثل مؤسسة Agent Link III التي تقوم بترتيب أعمال الحاسوب، اذ تقوم بالتعريف بالبرامج وهويتها وحل المشاكل ذات الطابع المتعلق بالحاسوب، نظرا لما لهذا النظام من قابلية ومرونة لحل المشاكل على كافة المستويات دون تدخل المستخدم. كما ان هناك ما هو خاص بقواعد البيانات الأولية Primary database agents لإدارة القواعد الأولية للتواليات والحفاظ على المعلومات محدثة على طول الوقت وكذلك nr database agents التي تحافظ على هذه القاعدة محدثة ودون تكرار ومشذبة على طول الوقت، كما ان هناك أنواع خاصة من العملاء مثل Calculation agents وهي أنظمة حاسوب مسئولة عن إدماج المعلومات الموجودة أصلا ومسئولة عن الطرق والوسائل المستعملة في تحليل بيانات التواليات، اما Genome agents فهي الأنظمة المسؤولة عن إدارة معلومات الجينومات لكل كائن حي.

وفضلا عن ما ذكر من وظائف العميل فان هذه الوسائل يمكن ان تستعمل في تحديد العلاقات وتحليل الحالات المرضية خاصة اذا كانت متعددة الأسباب الوراثية Polygenic diseases ومشاركة SNPs في بعض الحالات. اذ ان مثل هذه الحالات المتعددة العوامل لا يمكن ترتيبها بواسطة المستخدم ولكن باستعمال مثل هذه الوسائل يمكن تحديد العلاقات لهذه الأمراض كما في استعمال BioAgent التي تؤدي مهام أخرى.

ويستعمل العميل في المجالات المعقدة مثل التنبؤ بالتركيب الثلاثي الإبعاد للبروتينات اذ تحتاج العملية الى الكثير من التداخلات الحيوية الفيزيائية والكيميائية ومعلومات عن التركيب الثانوي، ومسألة تحديد التركيب المجسم للبروتين يشكل أهم اهتمامات الباحثين وتستخدم فيه الشبكات العصبية الصناعية (ANN) Artificial neural network ويساعد العميل بشكل كبير في هذا المجال.

والآلية التي تعمل بها هذه الوسائط هي تجميع كل التهميشات للبروتينات او التواليات وتنشأ قواعد معلومات متوافقة اي تعمل بمثابة مدير لمختلف الآراء حول البيانات المجمعة وتعطي الموديل الكامل اللازم للاستنتاج.

ولذلك فان دمج برامج العميل مع المعلوماتية الحيوية يوفر فرص مضاعفة، فالمعلوماتية الحيوية وما تزخر به من معلومات توفر من جانبها الفرصة لتطبيق تقنيات العميل والتي تصب جميعها في صالح الباحث. ولا تزال الحاجة ماسة لتحسين أداء المعلوماتية الحيوية الموجودة الآن ولذلك يمكن للتوجهات المستقبلية باستعمال العميل ان تساعد في تطويرها.

وكان من ثمار دمج قواعد البيانات وغيرها من الإمكانيات هو تطور فكرة مزودات خدمة الانترنت Web services التي مكنت من دمج البيانات البايولوجية من مختلف المصادر باستعمال أنظمة دمج خاصة مثل MyGrid و BioMOBY وغيرها لتسهيل وتوفير وسائل لتبادل المعلومات بين مزود الخدمات والمستخدم. ومن الأمثلة عليها قاعدة البيانات SYSTOMONAS الخاصة بالبكتريا Pseudomonas التي تم فيها جمع البيانات من عدد كبير من قواعد البيانات مثل BioCyc و ENZYME و BRENDA (وهي قاعدة بيانات للايض او للإنزيمات) وقواعد بيانات أخرى فضلا عن قاعدة البيانات الخاصة بجنس

Pseudomonas وكذلك الاستعانة بالمعلومات الواردة في المراجع الخاصة بالجنس التي تصف الجينات والبروتينات ووظائفها. ووضعت في مخزن للبيانات ووفرت لها الوسائل للوصول إليها وتستعمل كمرجع للتنبؤ بشبكات الايض والتنظيم في الأحياء التابعة لهذا الجنس المعقد جدا.

صيغ الملفات المستعملة في المعلوماتية الحيوية

تحتاج صيغ الملفات المستعملة الى تحويل سواء في المعلومات المدخلة او المخرجات بعد إجراء عمليات المعالجة عليها بالبرامج ، لذلك وجدت برامج تحول صيغ الملفات بحيث تلاؤم استخدام بعض البرامج.

وأكثر الملفات شيوعا هو الملف العادي Flat file الذي يكتب بصيغة النص العادي Plaintext وتكون بصيغة قياسية محددة. اذ يتمثل الأخير بوجود صفات مشتركة تستعمل فيها الحروف والأعداد والتنقيط الموجودة على إحدى وسائل الإدخال الى الحاسبة وهي Keyboard ، وهذا لا يستعمل في كثير من الأحيان في المهام الخاصة بالمعلوماتية الحيوية وان كان يستعمل في بعض الأحيان في وضع قواعد البيانات. وتختلف النصوص العادية عن النصوص ذات الصيغ المختلفة Formatted texts وأفضل الأمثلة عليها HTML.

والملفات العامة تكتب بالنص العادي ويمكن ان تكتب بعدد من البرامج مثل Notepad الخاص ببرنامج نوافذ Windows او TextEdit في نظام ماكنتوش Mac او برنامج vi في نظام Unix وفي العادة تعطي التسمية txt للملفات الناتجة التي تساعد النظام في تحديد نوع الملف الذي يتم التعامل معه وهذا ما يحصل مع نظام التشغيل نوافذ ، اما في الأنظمة الأخرى المذكورة أعلاه (Unix, Mac) فان الزيادة في تسمية الملف (txt) فلا تعني شيء للنظام وإنما تستعمل لتذكير المستخدم بنوعها.

وقد تكون الصيغ خاصة بالتواليات فقط ومنها صيغة التوالي العادية Plain sequence format وتحتوي فقط التواليات للحوامض النووية او التواليات الحوامض الامينية بشفرة الحرف الواحد كما موضحة في الشكل 10.2:

```
ACAAGATGCCATTGTCCCCGGCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCCTGCTCCCTGCC
CCTGGAGGGTGGCCCCACCGGCCGAGACAGCGAGCATATGCAGGAAGCGGCAGGAATAAGGAAAAGCAGC
CTCCTGACTTTCTCCTGCTTGTGGTTTGAAGTGGACCTCCAGGCCAGTGC CGGGC CCCTCATAGGAGAGG
AAGCTCGGGAGGTGGCAGGCGGCAGGAAGGCGCACCCCCAGCAATCCGCGCGCCGGGACAGAATGCC
CTGCAGGAACCTTCTCTGG
```

شكل 10.2: صيغة عادية Plain sequence لتوالي ما

وتكون هذه الصيغ بدون اي معلومات.

ومن الصيغ المهمة في المعلوماتية الحيوية هي صيغة FASTA التي تستعمل بشكل خاص مع التواليات DNA او البروتينات وتستعمل في العديد من برامج التنفيذ. ويمكن للملف ان يحويوالي واحد او عدة التواليات منفصلة وعندها يسمى Multi-FASTA. وكتابة النص فيه تكون بان يحوي السطر الأول على بداية (تميز صيغ FASTA) وهي علامة اكبر من (>) يليها سطر او سطرين للتعريف بالتوالي (وهذه تكون بمثابة بيانات إضافية المذكورة في موقع آخر)

ثم يلي ذلك ذكر توالي الجزيئات المعنية سواء باستعمال الحروف الصغيرة او الكبيرة على ان تكون بشكل متواصل بدون فواصل او تنقيط وتستمر بشكل متواصل وعندما تكون طويلة تقسم الى عدة اسطر، وفي هذا النوع من التواليات يشار الى الحوامض الامينية بشفرة الحرف الواحد. ولذا فمن الملاحظ ان هذه الصيغة لا تعطي معلومات وافية عن التوالي المتعامل معه، وتقيد لبعض الأغراض غير استخلاص المعلومات الوافية.

واغلب صيغ الملفات يمكن تناقلها بين الحواسيب باستعمال طرق النقل مثل File (FTP) transfer protocol (الذي سيأتي ذكره لاحقا) الذي يعد طريقة قياسية لتحميل كميات من البيانات من مواقع مختلفة وهو في اغلب الأحيان يكون متاحا ودون الحاجة الى استعمال اسم المستخدم او كلمة مرور.

وفي بعض الأحيان يتم الحصول على ملفات مضغوطة والتي يشار اليها في نهاية اسم الملف بـ zip ، gz ، tar ، tgz او اي إشارة مشابهة ولغرض تحويل هذه الملفات الى شكل قابل للقراءة والاستعمال يمكن فك الضغط (او الاستخلاص) ببرامج مثل WinZip او PKZIP او غيرها من البرامج المتوفرة. ولكن في العموم يمكن الحصول على ملفات عادية دون الحاجة الى معالجتها ويمكن استعمالها مباشرة.

تحويل صيغ التواليات

تحتاج بعض البرامج ان تكون المدخلات بصيغ محددة فمثلا اغلب البرامج التنفيذية تحتاج التوالي ان يكون بصيغة FASTA في حين برامج أخرى تحتاج التوالي بـ EMBL format لذلك أوجدت برامج كثيرة للقيام بهذه المهمة مثل Readseq ، وبعض البرامج يمكن ان تحول اي صيغة للتوالي الى الصيغة المطلوبة وذلك بوجود الخيارات المرفقة في البرنامج.

وبعض الأحيان يحتاج المستخدم الى تحويل التواليات من تواليات للـ DNA الى تواليات بروتين او بالعكس ولهذه برامج خاصة ولكن هناك شروط تقييد في هذه البرامج، لذلك تلحق بها بعض الخيارات مثل تحديد التوالي المراد تحويله هل هو من أحياء حقيقية النواة او بدائية النواة التي ربما تكون اركيا او بكتريا موجبة لصبغة كرام او سالبة لصبغة كرام وذلك بعد الأخذ بنظر الاعتبار وجود فروق في تفضيل استخدام الشفرات بين الأحياء لغرض الوصول الى الغرض المنشود، ولذلك تربط بعض هذه البرامج بجداول تفضيل استعمال الشفرات الخاصة بالكائنات الحية ولعل الوسائل الموجودة في واجهة قاعدة البروتينات السويسرية SWISSPROT وخاصة ExPASy (Expert Protein Analysis System) يوفر مثل هذه البرامج المفيدة.

مراكز المعلوماتية الحيوية

هناك العديد من المراكز التي تعنى بالمعلوماتية الحيوية وتقدم خدمات كثيرة عبر شبكة الانترنت، وتتوزع المراكز على أنحاء مختلفة من العالم. وتختلف المراكز من حيث سعة شموليتها لمتطلبات المعلوماتية الحيوية وإمكانياتها ولعل أهمها

NCBI (National Center for Biotechnology Information).

EBI (European Bioinformatics Institute) البريطاني

والفروع الأخرى مثل EMBL (European Molecular Biology Laboratory) في ألمانيا/الدول الأوروبية

DDBJ (DNA Database of Japan) في اليابان .

وكل من هذه المراكز متواصلة مع بعضها على مدار الساعة، وأهمها NCBI الذي هو أحد منظومات حكومية تعنى بالشؤون الطبية، وموقع المركز يحوي إمكانات هائلة تتطوي تحتها المعلوماتية الحيوية، إذ يضم عدد من قواعد وبنوك البيانات ووسائل تنفيذ البرامج والروابط مع المواقع خارج نطاقه. ويليه في الأهمية المركز الأوروبي الذي هو الآخر يمتلك إمكانات كبيرة تساهم في دراسات المعلوماتية الحيوية، ثم يليهم بالأهمية المركز الياباني. فضلا عن ذلك فهناك العديد من المراكز الأخرى الخاصة ببعض البلدان أو الجامعات المشهورة التي تقدم الوسائل والإمكانات المستخدمة في مجال المعلوماتية الحيوية ولكنها تكون على مدى ضيق مقارنة بما مذكور أعلاه .

والمراكز العالمية الثلاث (المهمة) المذكورة أعلاه ونظرا لتواصلها مع بعض فهي متوائمة في أغلب الجوانب ومختلفة في أخرى، فمثلا تتفق في استعمال رقم تسجيل موحد لكل جزيئة حيوية وتعدّها الشفرة الخاصة بكل جزيئة، ومن الاختلافات الملاحظة انه في DDBJ تستعمل وسائل لا تتوفر في NCBI مثل برامج FASTA المستعملة لصف التواليات واستعمال عائلة برامج CLUSTAL لصف تواليات البروتينات (التي سيأتي ذكرها لاحقا)، والمركز الأخير يكون التعامل معه اقل تفضيلا للمستخدم مقارنة بـ NCBI، في حين في رأي الكاتب يكون التعامل مع EMBL هو الأفضل إذ يوفر روابط الى مواقع تزود بالمعلومات، ويضم وفرة من الجينات البكتيرية ولكل مركز درجته من المصداقية، ولكن التفضيل ليس هو الحد الفاصل في مجال استخدام مواقع هذه المراكز نظرا لتفاوت الخدمات التي تقدمها.

المركز الدانمركي أنشأ عام 1993 وهو CBS (Center for Biological Sequence Analysis) ويقوم المركز بعدة أبحاث في مجال علوم الحياة المتعلقة بالمعلوماتية الحيوية. ويوفر المركز الكم الكبير من البرامج اللازمة في دراسات المعلوماتية الحيوية واغلب برامجه بسيطة سهلة الاستخدام ومخرجاتها سهلة التفسير وتشمل كل النواحي التي يحتاجها المستخدمون في مجال المعلوماتية الحيوية. والمركز مدعوم من قبل الجهات الرسمية في الدانمرك .

والمواقع الالكترونية تكون عرضة للنقاع والاختفاء ولكن هناك بعض المراكز التي تكون مدعومة بمؤسسات رسمية وتكون ثابتة والموضح بعضها وعناوينها في الآتي :

Genbank (<http://www.ncbi.nlm.nih.gov/>)
 Protein Data Bank (<http://www.rcsb.org>)
 ExPASy Proteomics Server (<http://us.expasy.org/>)
 European Bioinformatics Institute (<http://www.ebi.ac.uk/>)
 SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>)
 CATH (<http://www.biochem.ucl.ac.uk/bsm/cath/>)
 PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>)
 PubMed Central (<http://www.pubmedcentral.nih.gov/>)

قواعد البيانات Databases

قواعد البيانات إحدى الركائز التي تعتمد عليها المعلوماتية الحيوية التي نشأت نتيجة لتفاقم البيانات التي احتاجت إلى ترتيب و تخزين وإيجاد الوسائل الملائمة للوصول إليها ، فقواعد البيانات أصبحت ضرورة للعاملين في مجال علوم الحياة وبمثابة وسيلة الاستخدام اليومي ، لذلك كانت عمليات التحديث المستمر ضرورية. والمتداول لمعنى قاعدة البيانات Databases أكثر من معنى فهي:

- تعني تجميع مرتب للبيانات وهو المعنى الأكثر شيوعا في الاستعمال.
- يمكن ان تعني برنامج حاسوب تستعمل للبحث في السجلات واسترجاع بعض المعلومات وبهذا المعنى تكون قاعدة البيانات ممثلة للـ Database management system (DBMS) مثل MySQL وهنا هي تمثل البرنامج الذي يدير قاعدة البيانات وليس قاعدة بيانات.
- المعنى الثالث وهو يستعمل بكثرة أيضا وهو وصف لمجموعة من الملفات مخزونة في الحاسوب وموجودة في مطوي Folder يحوي عدد من التواليات يعتبرها البعض قاعدة بيانات. وهناك عدة أسباب ودواعي لاستخدام قواعد البيانات ومنها:
 - ❖ تحديد هوية التواليات الجديدة، وعند تحديد توالي جديد المطلوب معرفة فيما اذا كان مسجل مسبقا سواء كليا او جزئيا لذلك يتم البحث عنه في قواعد البيانات وكذلك معرفة فيما اذا كان التوالي الجديد موجود في أحياء أخرى.
 - ❖ البحث عن تواليات وامتدادات معينة خاصة من تواليات DNA غير المشفرة مثل المكررات او العناصر التنظيمية.
 - ❖ إيجاد وظائف الجينات او التواليات غير المعروفة بالمقارنة مع ما يقابلها من التواليات الموثقة في قواعد البيانات.
 - ❖ البحث عن البوادي Primers لتواليات تحت الدراسة لأغراض التضخيم كما في استعمال تقنيات PCR ، او لغيرها من الأغراض وذلك بإجراء دراسات مقارنة مع ما مسجل في قواعد البيانات.

التعريف بقواعد البيانات

قواعد البيانات عبارة عن جداول او أرشفة الكترونية تستخدم ل تخزين وتنظيم البيانات بطريقة يمكن معها استرجاع المعلومات المطلوبة بسهولة باستخدام معايير بحث مختلفة. تتألف قواعد البيانات من البرمجيات والجزء الصلب للحاسب الالكتروني لغرض إدارة البيانات. ان الهدف الرئيس من تطوير قواعد البيانات هو لتنظيم البيانات في مجموعة هيكلية من القيود لتسمح بعد ذلك من استرجاع المعلومات. كل قيد يسمى أيضا مدخل ، ويحوي عدداً من الحقول التي تحتوي على عناصر البيانات الحقيقية، على سبيل المثال حقول مثل الأسماء، أرقام الهاتف، العناوين والتواريخ. ولغرض استرجاع قيد معين من قواعد البيانات، فان المستخدم يمكن ان يخصص قطعة او جزء معين من المعلومات ، تسمى قيمة Value لغرض إيجادها بصورة خاصة في حقل معين فالحاسب الالكتروني يسترجع جميع قيود البيانات التي لها علاقة بقطعة المعلومة المطلوبة. وتسمى هذه العملية بالاستعلام Querying.

على الرغم من ان استرجاع البيانات هو الهدف الرئيس لجميع أنظمة قواعد البيانات، فان أنظمة قواعد البيانات البايولوجية تمتلك مستلزمات ذات مستويات أعلى ، تعرف باستكشاف المعرفة Knowledge discovery التي تشير الى التطابق في الارتباطات بين قطع المعلومات غير المعروفة عندما تدخل المعلومات في قواعد البيانات لأول مرة. على سبيل المثال، تحتوي قواعد البيانات على بيانات التوالي الخام Raw sequence ، وهي بحاجة الى أعمال حسابية إضافية للتعريف بالتماثل الجذري للتوالي Homology ، هذه الصفات والخصائص تسهل اكتشاف المعلومات البايولوجية الجديدة المنبثقة من البيانات الخام والأولية .

أنواع قواعد البيانات

تستخدم جميع قواعد البيانات صيغة الملف العادي الذي هو ملف نصي طويل يحوي العديد من المدخلات، مفصولة بعدة مؤشرات او الرموز الخاصة مثل العمود الرأسى (1). في كل مدخل وهناك عدد من الحقول مفصولة بواسطة الفواصل او الفراغات. يحتوي كل حقل على بيانات خام، في حين لا يحتوي الملف النصي على أية تعليمات مخفية للحاسب الالكتروني لكي تستخدم للبحث عن معلومات خاصة او تهيئة التقارير استنادا الى حقول معينة لكل قيد. يمكن النظر الى الملف النصي على انه جدول منفرد وهكذا فعند البحث في الملف العادي عن جزء معين من المعلومة فان على الحاسب الالكتروني قراءة الملف بأكمله، وبالتالي فانه من الواضح هي عملية غير فعالة. ويمكن استخدامها لقاعدة بيانات صغيرة، لكن بزيادة حجم قاعدة البيانات او عند تنوع البيانات المستخدمة فان العملية تصبح أكثر تعقيداً، وهذا النوع من قواعد البيانات يكون الاسترجاع منه صعباً" للغاية عند استرجاع البيانات المخزونة فيه وفي الواقع فان البحث خلال هذه الملفات غالبا ما يسبب تحطيم او تعطيل نظام الحاسب الالكتروني بسبب طبيعة العمل المركز الذي يجري في ذاكرة الحاسب الالكتروني.

ولغرض تسهيل الوصول واسترجاع البيانات، يتطلب ذلك برنامج معقد للحاسب الالكتروني للتنظيم والبحث والوصول الى البيانات التي تم تطويرها وتدعى أنظمة إدارة قواعد البيانات Database management systems تحتوي هذه الأنظمة ليس على قيود البيانات الخام فقط ولكن ايضا تعليمات العمل للمساعدة في التعريف بالارتباطات الخفية خلال قيود البيانات. ان الهدف من تأسيس هيكل للبيانات هو سهولة تنفيذ البحوث ولتوحيد القيود المختلفة لتشكيل التقارير النهائية للبحث اعتمادا على أنواع هياكل البيانات. ان أنظمة إدارة قواعد البيانات يمكن ان تصنف الى نوعين:

❖ أنظمة إدارة قواعد البيانات العلاقية Relational database management systems .

❖ أنظمة إدارة قواعد البيانات المواضيع-الموجه Object-oriented database management systems .

وبالتالي، فان ما ينتج من تنفيذ أنظمة قواعد البيانات هذه تعرف بقواعد البيانات العلاقية او قواعد الموضوع-الموجه.

مثال 1.2:

لتوضيح ما تقدم يفترض وجود معلومات توضح أسماء الطلبة في فصل دراسي معين ولعدد مختلف من الجامعات خزنت في ملف عادي يحتوي على خمسة قيود طلبة من عدة جامعات

مختلفة، كل طالب أخذ موضوع فصل دراسي مختلف كما موضح في الشكل التالي (11.2).
 بفصل كل قيد بعامود رأسي عن القيد الآخر، ويحتوي القيد الواحد على أربعة حقول تمثل،
 الاسم واسم الفصل الدراسي (المادة)، رقم الفصل الدراسي واسم الجامعة.

الاسم، اسم الفصل الدراسي، الدرجة، اسم الجامعة|اسعيد حامد محمد، معلوماتية حيوية، 89،
 جامعة النهرين|مريم علي عبد الحافظ، برمجة، 98، جامعة بغداد|وليد فيصل علي، كيمياء
 حيوية، 64، جامعة الموصل|منى عامر ليث، الكيمياء العضوية، 78، جامعة البصرة|أسماء طارق
 عبد الله، الرياضيات، 88، جامعة النهرين

شكل 11.2: قاعدة بيانات ذات ملف عادي تحتوي معلومات قيود خمسة طلبة.

قواعد البيانات العلاقية Relational databases

تستخدم قواعد البيانات العلاقية مجموعة من الجداول بدلاً من استخدام جدول منفرد كما في قاعدة البيانات ذات الملف العادي لتنظيم البيانات. يدعى كل جدول "علاقة"، يتكون من عدد من الأعمدة والصفوف. تشير الأعمدة إلى الحقول المنفردة، في حين تمثل الصفوف قيم الحقول في القيود. تفهرس الأعمدة في الجدول طبقاً إلى ميزة عامة تدعى بالصفة المميزة Attribute وهذا يمكن الانتقال المتقاطع مع الجداول الأخرى. ولغرض تنفيذ الاستعلام في قاعدة البيانات العلاقية، يختار النظام عناصر ربط البيانات من عدد من الجداول المختلفة ثم يوحد المعلومات في تقرير واحد وبذا يمكن الحصول بسرعة أكبر على معلومات محددة من قواعد البيانات العلاقية مقارنة مع قاعدة البيانات ذات الملف العادي.

يمكن إنشاء قواعد البيانات العلاقية باستخدام لغة برمجة خاصة تدعى لغة الاستفسار الهيكلية Structured query language (SQL) وبعد إنشاء قاعدة البيانات الأصلية، تمكن اللغة يمكن إضافة أصناف جديدة من البيانات بسهولة إلى الجداول الموجودة، وبهذا فإن البحث في قواعد البيانات وجمع البيانات للحصول على التقارير المطلوبة يمكن أن تجري بصورة سهلة ومباشرة نسبياً.

مثال 2.2: يوضح الجدول التالي (2.2) طريقة عمل قاعدة البيانات العلاقية باستخدام المعلومات نفسها الواردة في المثال السابق. تتلاءم البيانات مع كل جدول، بحيث يمكن ربط الجداول المختلفة بواسطة عمود مشترك من البيانات يدعى المفتاح الرئيس Primary key للوصول إلى المعلومات المطلوبة وهو في المثال ممثل برقم الطالب.

فمثلاً، إذا كان السؤال الآتي: أي الفصول الدراسية التي أخذها طلاب جامعة النهرين؟ فإن محرك البحث في قواعد البيانات سيجد أو يحدد حقل الجامعة من جدول رقم (1) ثم البحث عن اسم النهرين-ضمن حقل الجامعة- وعند نهاية البحث في هذا الحقل ستكون النتيجة الحصول على الطالبين بالتسلسلين 1 و 5 والذي يناظر الدرجة 89 و 88 على التوالي من الجدول رقم (2)، كذلك يناظر اسم الفصل الدراسي معلوماتية حيوية ورياضيات من الجدول رقم (3). وأخيراً فإن التقرير سيتضمن -استناداً إلى جامعة النهرين- اسمي الطالبين والدرجتين التي تقابلان اسمي الفصلين الدراسيين. ويلاحظ أن عملية البحث والاستفسار عن المعلومات المطلوبة نفسها في قاعدة البيانات ذات الملف العادي ويتطلب البحث خلال جميع المعلومات الموجودة في الملف -كلمة بعد كلمة وخزن المعلومات في مجال خزن مؤقت ثم تأشير القيود

التي تحتوي على كلمة النهرين، هذه العملية ستكون سهلة في حالة كون قاعدة البيانات ذات الملفات العادية صغيرة الحجم، اما في حالة البحث في قاعدة بيانات — من هذا النوع — ذات حجم كبير فان المهمة ستكون شاقة وصعبة.

جدول 2.2 : جدول بيانات لقواعد البيانات العلاقية

جدول (1)		جدول (2)		جدول (3)	
رقم الطالب	اسم الطالب	اسم الجامعة	رقم الطالب	الدرجة	اسم الفصل الدراسي
1	سعيد حامد محمد	النهرين	1	89	معلوماتية حيوية
2	مريم علي عبد الحافظ	بغداد	2	98	برمجة
3	وليد فيصل علي	الموصل	3	64	كيمياء حيوية
4	منى عامر ليث	البصرة	4	78	كيمياء عضوية
5	أسماء طارق عبد الله	النهرين	5	88	رياضيات

قواعد بيانات الموضوع-الموجه Object-oriented databases

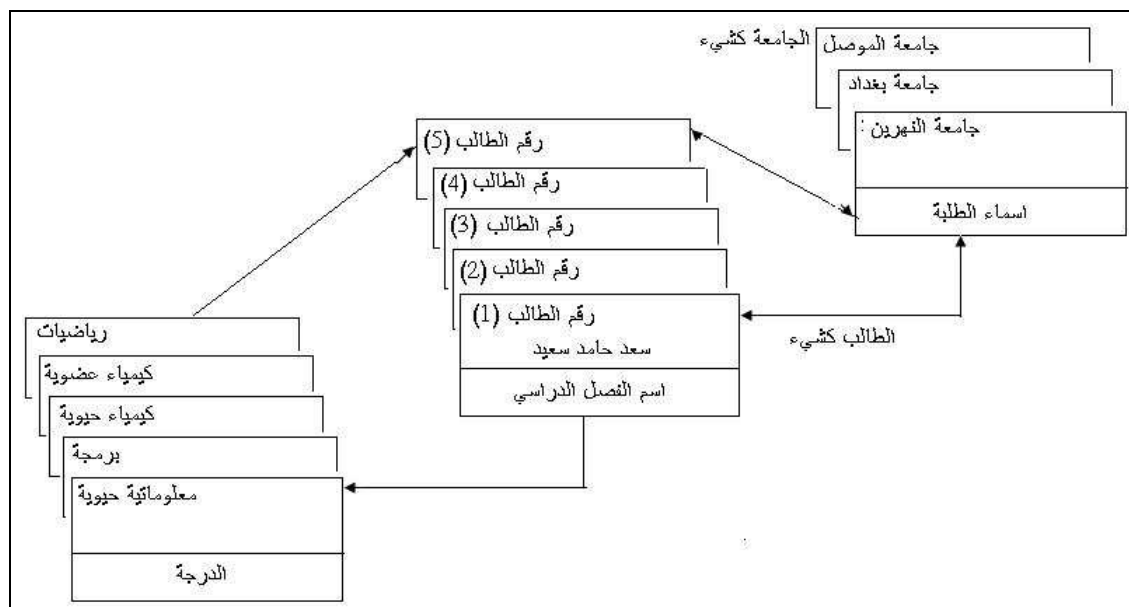
أحدى المشاكل التي تواجه قواعد البيانات العلاقية هي ان الجداول المستخدمة لا تصف العلاقات الهرمية المعقدة بين مواد البيانات. وللتغلب على هذه المشكلة، طورت قواعد بيانات الموضوع الموجه عملية خزن البيانات كأشياء. في لغة برمجة الشيء الموجه، يمكن النظر الى الشيء على أساس انه الوحدة التي تضم البيانات والبرامج الفرعية التي تعمل على البيانات. ان قواعد البيانات يتم هيكلتها بحيث ان المواضيع او الأشياء تربط بواسطة مجموعة من المؤشرات التي تعرف مجموعة من العلاقات بين الأشياء. والبحث في قاعدة البيانات يستلزم الإبحار خلال الأشياء بمساعدة المؤشرات التي تربط بين الأشياء المختلفة. وتستخدم لغات البرمجة مثل C++ لإنشاء قواعد بيانات المواضيع الموجهة .

تعد قواعد بيانات المواضيع-الموجهة أكثر مرونة، فالبيانات يمكن هيكلتها اعتمادا على العلاقات الهرمية وبهذا فان مهمة البرمجة هو إمكانية تبسيط التعامل مع البيانات المعروفة بان لها اي البيانات علاقات معقدة. على أية حال فان هذا النوع من قواعد البيانات يفتقر الى الأساس الرياضي المتين الموجود في قواعد البيانات العلاقية. كذلك هنالك بعض الخطورة لان بعض العلاقات بين الأشياء او المواضيع قد تكون غامضة.

تمتلك بعض قواعد البيانات الحالية خصائص وميزات مشتركة لكلا النوعين من برمجة قواعد البيانات وهي قواعد البيانات العلاقية وقواعد بيانات المواضيع — الموجهة بإنشاء نظام إدارة قاعدة البيانات علاقية — الشيء.

مثال 3.2: استنادا الى البيانات المذكورة في المثال السابق والتي يمكن استخدامها لبناء قاعدة بيانات الموضوع-الموجه. يمكن تصميم أربعة كارتات مختلفة اسم الطالب واسم الفصل الدراسي والدرجة واسم الجامعة ويرمز للعلاقات المتداخلة بين هذه الأشياء (العناصر) بواسطة الأسهم والخطوط. يوضح الشكل التالي (12.2) الإجابة على السؤال نفسه — ما هو اسم الفصل الدراسي الذي درسه طلاب جامعة النهرين. بهذا فان البحث سيبدأ من جامعة النهرين من خلال

اسم الجامعة - الذي يرمز له كشيء- وله مؤشرات تقود الى الطالبين 1 و 5 في اسم الطالب- الذي يرمز له كشيء- . هنالك مؤشرات إضافية تؤشر اسم المواد الدراسية المعطاة ومن ثم درجة كل طالب. وبالتالي فان الانتقال خلال روابط الأشياء تجهز التقرير النهائي.



شكل 12.2 : قاعدة بيانات الموضوع-الموجه لخمس طلاب

طبيعة البيانات المخزونة في القواعد

اغلب قواعد البيانات الخاصة بعلم الحياة تحوي على بيانات أولية وهي اما تواليات للـ DNA او RNA او البروتينات. ولذلك كانت هذه بحاجة الى بيانات إضافية Metadata ولذا وجب وضع استراتيجيات مهمة عند تخزين المعلومات وأهمية التفريق بين البيانات الأولية والبيانات الإضافية . فالبيانات هي اما تواليات حوامض نووية او تواليات بروتينات، ففي الحالة الأولى تكون البيانات بمثابة مكررات من 4 او 5 حروف او أكثر عند إضافة الرمز N ليعني اي نيوكليوتيد. اما في حالة البروتينات فتكون مكونة من 21 حرف وجعل X لأي حامض غير معروف الهوية وبذا تكون مفردة البيانات الأصلية محددة . اما مفردة البيانات الإضافية فهي غير محددة ، وقد تكون البيانات الإضافية هي الهدف المطلوب من البحث في قاعدة البيانات دون الحاجة الى البيانات الأصلية ، مثلا يراد عمل تجميع لكل الأحياء المحبة للحرارة الموجودة في القاعدة او معرفة الدرجات الحرارية التي تعيش فيها الأحياء المحبة للحرارة، فالمتوقع هنا عدم الحاجة الى التواليات ومن غير المتوقع ان توجد هذه في الجزء الخاص بتواليات الجينوم. ويمكن تقسيم البيانات الإضافية الى:

- البيانات الإضافية الأولية او البدائية والتي تسمى ايضا الحقائق الأولية وهي التي لها علاقة بالتوالي نفسه او بالطريقة التي أدت الى الحصول عليه ومنها ذكر الكائن الذي اشتق منه التوالي والتوزيع الهندسي للجزيئة مثلا تكون حلقية او مفتوحة، وكذلك معلومات عن تركيب

الكروموسوم او البلازميد او الفيروس وطرق تحديد تواليهم في المختبر، فضلا عن احتواءها على الصفات الفسلجية للكائن والموقع الجغرافي الذي يوجد فيه ونوع النموذج والأعراض السريرية عند الإصابة في حالة الأحياء الممرضة وغيرها من المعلومات.

- البيانات الإضافية المشتقة Derived metadata، يتم الحصول عليها من عمليات الحاسوب التي تجري على البيانات وتشمل ايضا التمهيش اليدوي Manual annotation ، وهذه تأتي اما من التوالي او اي معلومات أولية أخرى فمثلا تحديد محتوى الجينوم من AT الذي يؤدي بدوره كتحصيل حاصل الى معرفة المحتوى من GC. ويكون بعض الأحيان من الصعوبة تحديد صنف البيانات الإضافية فيمكن ان تتحول البيانات من الصنف الأول الى الثاني. ومن أهم البيانات الإضافية هي أطالس الجينومات Genome atlases التي تعد بيانات إضافية للتواليات الأصلية ويكون الأطلس من البيانات المشتقة.

أهمية البيانات الإضافية

تكون البيانات الإضافية مهمة جدا في بعض الأحيان وبدونها يصبح من غير الممكن استخدام التواليات للجزيئات الحيوية وذلك لعدة أسباب:

- تكون البيانات الإضافية ضرورية لعملية الفهرسة Indexing وهذا يسهل الوصول الى مجاميع خاصة من المراجع المهمة بتوالي معين.
- كما يمكن باستعمال البيانات الإضافية الوصول الى تواليات محددة مثل البحث عن التواليات الحاوية على AT بنسب محددة فان ذلك يسهل باستخدام البيانات الإضافية .
- يمكن ان تساعد في معرفة أصل التوالي المستعلم عنه مثلا توالي مشتق من كروموسوم او بلازميد، وكذلك يمكن تحديد الكائن او المجموعة التصنيفية التي ينتمي اليها. وتساعد في تحديد البيئة التي أخذ منها الكائن.

وغيرها من الأسباب. ويمكن الاستنتاج ان البيانات الإضافية ستحسن او على الأقل تقلل من اللبس الحاصل حول تفسير النتائج، وذلك لكونها تتضمن بالإضافة الى ما ذكر أعلاه من المعلومات وصف للتقنيات المستعملة والتعريف بوحدات القياس المستعملة ووصف لظروف أخذ النماذج وطرق إجراء التجارب، وعليه فان غياب البيانات الإضافية يمكن ان تجعل البيانات المخزونة في قواعد البيانات بدون فائدة، لذا كان للبيانات الإضافية دورا "كبيراً" في المعلوماتية الحيوية.

تصنيف قواعد البيانات

تختلف المؤشرات المستعملة في تصنيف قواعد البيانات وذلك لاختلاف:

- الطريقة التي تكتب بها القاعدة.

- البيانات التي تحتويها

ونظرا لاحتمالية تغيير قواعد البيانات او دمجها مع بعضها سيتم التطرق الى ذكر القواعد الخاصة بجوانب معينة دون الخوض في ذكر مواقعها الالكترونية لانه في بعض الأحيان قد يتم حذف قاعدة معينة والدخول اليها واذا كان إلغائها حديثا يذكر الموقع ذلك ويرشد الى الوسيلة الموصلة الى البيانات التي احتوتها القاعدة. فضلا عن وجود حالة دمج قواعد البيانات وبالتالي إنتاج قاعدة واحدة.

قواعد البيانات العامة للجزيئات الحيوية

وهذه يمكن ان تضم قواعد بيانات خاصة بمجموعة معينة من الجزيئات كقواعد أولية ثم مشتقات الجزيئات كقواعد ثانوية وفقا لأهمية الجزيئات:

- القواعد المتعلقة بتواليات البروتينات وأهمها قاعدة البروتينات السويسرية وهي موجودة على الموقع الأوروبي وتابعة لمعهد المعلوماتية الحيوية في جنيف. وتعد من أهم القواعد الثانوية التي تمثل قاعدة بيانات تحوي على تواليات الحوامض النووية المترجمة الى البروتينات والمخزونة في قواعد المركز الأوروبي EMBL. وتوفر القاعدة بيانات وتهميشات عن التواليات البروتينية مثل التركيب والوظيفة والعائلة البروتينية وتركيب البروتينات ومواقع الفعالية المساعدة Catalytic sites ومواقع ارتباط العوامل المرافقة في حالة الإنزيمات. فضلا عن التحويلات التي تمت بعد الترجمة ومعلومات عن المسارات الايضية التي تدخل فيها البروتينات والعلاقة بالأمراض والتشابه مع التواليات الأخرى. ومعظم هذه المعلومات تم الحصول عليها من المراجع العلمية ويتم إدخالها من قبل القائمين على القاعدة Database curators وهم أشخاص مدربون لمثل هذه المهام لذلك تكون المعلومات ذات نوعية جيدة، اذ ان عملية التهميش والتشذيب تضيف قيمة علمية لسجل التوالي الأصلي. ومن الأشياء الأخرى التي جعلت هذه القاعدة مرغوبة من قبل البايولوجيين هو قلة التكرار الناتج عن عمليات التشذيب المذكورة آنفاً وكذلك مستوى التكامل العالي مع قواعد البيانات الأولية والثانوية الأخرى. وقد تم دمج القاعدة مع TrEMBL وكذلك Private information retrieval resource (PIR) لتكوين قاعدة موحدة أوسع من القواعد الثلاث وهي UniProt database وبقيت محافظة على صفاتها من قلة التكرار والمستوى العالي من التهميش.

- قواعد بيانات خاصة بالكربوهيدرات المسماة CarBank وكانت سابقا تسمى Complex Carbohydrate Structure database (CCSD) ولكنها توقفت.

- قواعد بيانات خاصة بالتركيب الثلاثي 3D للجزيئات منها الخاصة بتركيب البروتينات مثل بنك بيانات البروتينات Protein Data Bank (PDB) وهي قاعدة أمريكية مستعملة بكثرة وتعد القاعدة من القواعد المركزية في هذا المجال، وتحوي أرشيفات القاعدة على ترتيب الذرات في الجزيئات الحيوية العملاقة سواء كانت البروتينات او مكوناتها من الحوامض الامينية التي حددت تراكيبها بواسطة دراسات الأشعة السينية وطريقة الرنين النووي المغناطيسي NMR ، وتستعمل القاعدة صيغة الملف البسيط وتشمل معلوماتها على اسم البروتين، اسم او أسماء الباحثين، التفاصيل والوصف المختبري، والتركيب الثانوي في حالة البروتينات والعوامل المرافقة Cofactors وغيرها من الصفات، وتوفر واجهة القاعدة وسائل للتوضيح Viewing tools فضلا عن الروابط المساعدة.

- هناك قواعد خاصة بالجزيئات العملاقة بشكل عام كما في Macromolecular structure database (EBI-MSD) الموجودة على موقع المركز الأوروبي.

- قواعد بيانات لتراكيب الحوامض النووية Nucleic acid structure database (NDB) التابعة لإحدى جامعات نيوجرسي الأمريكية.

قواعد البيانات الخاصة

ويمكن ان تكون هذه القواعد خاصة بالجزيئات الحيوية العملاقة او أجزاء خاصة فيها والبيانات في هذه القواعد اما ان تكون تواليات او أنواع أخرى من المعلومات ويمكن ان تتداخل بياناتها مع

البيانات في قواعد البيانات الأولية (وخاصة التواليات) ولكن يمكن ان تحوي على بيانات جديدة قدمت مباشرة من قبل الباحثين. ونظرا لانها تعالج وتشذب من قبل ناس متخصصين فقد يكون لها ترتيب خاص وتهميشات إضافية لها علاقة بالتواليات والعديد من قواعد بيانات الجينومات ذات العلاقة بالتصنيف تقع ضمن هذه المجموعة وهي في العادة تخدم مجاميع خاصة من البحوث ومنها:

• القواعد الخاصة بالـ DNA وملحقاته

◆ dbEST وهي خاصة بالبيانات المتعلقة بطمغات التواليات المعبر عنها Expressed sequence tags (ESTs) وتوجد في موقع NCBI وتشابهها في ذلك قواعد أخرى مثل Mendel ESTs و Body map الخاصة بالـ ESTs في الإنسان والفار (وهذه تابعة لجامعة طوكيو اليابانية).

◆ dbGSS الخاصة بالبيانات للـ Genome survey sequences توجد في موقع NCBI.

◆ dbSTS قاعدة بيانات خاصة بـ Sequence tagged sites في موقع NCBI.

◆ كما توجد قواعد بيانات خاصة بتغايرات النيوكلوثيرد المفرد SNP مثل dbSNP في موقع NCBI وكذلك SNPper الخاصة بمستشفى الأطفال في مدينة Boston الأمريكية.

وفي السياق ذاته هناك قواعد بيانات لتغايرات النيوكلوثيرد المفرد ولكن في الأغلب غير المترادفة Nonsynonymous SNPs التي تحوي على الوفير من المعلومات الخاصة بالاضطرابات والأمراض الناتجة عن هذه الحالات اذ ان معظم تغايرات النيوكلوثيرد المفرد تكون غير مؤثرة.

فضلا عن وجود قواعد بيانات يطول ذكرها خاصة بالـ DNA.

وضمن قواعد البيانات الخاصة هناك قواعد أخرى :

• قواعد البيانات خاصة بالطفرات التي تحصل في الإنسان Human gene (HGMB) mutation database والتي تساعد في إيجاد الطفرات ذات العلاقة بالأمراض.

قواعد بيانات الوراثة اللاجينية Epigenetics

قواعد البيانات الخاصة بالوراثة اللاجينية واسعة ونظرا لتوسع هذا المجال من علم الحياة ظهرت العديد والكثير من قواعد البيانات، وفيما يلي مجمل للتوجهات التي وصفت قواعد البيانات والتي توفر وسائل وبرامج للعمل والدراسة ولا يمكن ذكرها بالتفصيل نظرا لكثرتها وكذلك توسعها المستمر.

➤ قواعد بيانات خاصة بالكروماتيدات ومفصل عن تراكيبيها ومزودة بالوسائل لدراسة الجوانب المتعددة لمكونات الكروماتيدات.

➤ قواعد البيانات خاصة بجزئيات RNA غير المشفرة والمشاركة في الوراثة اللاجينية وتشمل قواعد كثيرة

○ قواعد بيانات خاصة بجزئيات RNA مثل siRNA وغيرها.

○ قواعد بيانات خاصة بمثيلة DNA وملحقاته كأحد الأحداث المهمة في الوراثة اللاجينية.

○ قواعد بيانات عن مثيلة DNA

○ قواعد بيانات للجينات التي يكون DNA حاويا" على المثلث Methylated genes لأنواع مختلفة في السرطانات.

○ قاعدة بيانات لإنزيمات المثيلة DNA methyltransferases وهي مشتقة من قاعدة البيانات العامة GenBank.

- قواعد بيانات لمثيلة الجينات للعديد في سرطانات الإنسان.
- قواعد بيانات للبواقي لتحليل مثيلة DNA للإنسان والفيضان، وغيرها من الكائنات.
- قواعد بيانات خاصة بمثيلة كروموسومات الإنسان رقم 21 و 22.
- قواعد بيانات خاصة بالمثيلة في الخلايا الجنسية في الإنسان وتحتوي على أنواع من نسق المثيلة للـ DNA في التوائم العادية Dizygotic twins والتوائم المتماثلة Monozygotic twins.
- قواعد بيانات خاصة بحالات المثيلة المرتبطة بأمراض الزهايمر Alzheimer's disease.
- قواعد بيانات وبرامج لإيجاد جزر CpG المرتبطة بالممهدات.
- قواعد بيانات وبرامج لإيجاد المناطق الغنية بـ CpG .
- قواعد وبرامج تظهر عمليات تحليل وتحديد التواليات بطريقة Bisulfite.
- قواعد بيانات وبرامج تبحث عن CpG اعتمادا على مؤشرات معينة.
- قواعد وبرامج لرسم علاقات CpG.
- قواعد بيانات وبرامج لإيجاد الممهدات في جزر CpG.
- قواعد بيانات لإيجاد مثيلة DNA تحت الظروف البيئية المختلفة.
- قواعد البيانات الخاصة بالجينات الأساسية لحياة الكائن Essential gene database وفيها يتم ضم الجينات المهمة بعد تحديد هذه الأهمية بناء على العديد من التجارب واستعمال المراجع، والقاعدة مقسمة الى جينات خاصة بالأحياء بدائية النواة والأخرى حقيقية النواة، وتحميل المعلومات منها سهل والمعطيات لها تكون بصيغة ملفات بنك الجينات الحاوية على وفرة من المعلومات فضلا عن توالي الحوامض النووية وترجمتها وتستخدم للعديد من الأغراض مثل تحديد الأهداف الوراثية للأحياء الممرضة او غيرها .
- فضلا عن القواعد والبرامج اعلاه التي تخص DNA هناك قواعد بيانات للبروتينات وما يرتبط بها مثل الهستونات ومنها:
- قواعد بيانات للهستونات وتحتوي على التواليات لكل من H₂A ، H₂B ، H₃ ، H₄ لأنواع مختلفة من الأحياء .
- قواعد بيانات خاصة بالعوامل المحورة للكروماتين في تواليات الأحياء حقيقية النواة.
- قواعد بيانات خاصة بالجينومات
- ومن أكبرها قاعدة الجينومات المباشرة Genomes online database (GOLD) وتهتم بالتواليات الجديدة التي يرام إدخالها الى قواعد البيانات العامة وتدرج لكل الجينومات التي تم تحديد تواليها وهي تابعة لجامعة Illinois الأمريكية وقاعدة الجينومات التابعة لمركز NCBI الأمريكي.
- وهناك قاعدة بيانات خاصة بالأحياء Organism specific database الأمريكية والأخرى (TIGR database) TDB التابعة لمركز بحوث الجينوم الأمريكي كما توجد قاعدة تحوي بيانات لجينومات الأحياء المجهرية MBGD (Microbial genomes database) وهذه تمثل جزءا " من قاعدة بنك الجينات وتشمل جينومات الأحياء المجهرية التي تم تحديد تواليها.
- Ensemble قاعدة بيانات مهمشة تابعة للمركز الأوروبي EBI .
- وتوجد قواعد بيانات أخرى كثيرة تعني بتهميش وتوضيح الجينات.

وفي هذا المجال هناك قواعد بيانات بجينومات العضيات الخلوية مثل المايتوكوندريا MITOP وتحوي بيانات عن الجينومات والجينات ذات العلاقة وبيانات حول البروتينات الخاصة بالمايتوكوندريا والأمراض.

وضمن هذه المجموعة تدرج قاعدة البيانات الخاصة باستعمال الشفرات ومنها Codon usage database وتحوي بيانات حول استعمال الشفرات وتفضيلها في الأنواع المختلفة من الأحياء ولكل الجينات الموجودة في GenBank (NCBI) كما تدرج مع المجموعة قاعدة البيانات الخاصة بالدراسات الجينومية المقارنة مثل GeneCensus التابعة لجامعة Yale الأمريكية.

قواعد البيانات الجينومية الخاصة بالأحياء

وتتطوي تحت هذه المجموعة قواعد بيانات متفرقة على أنحاء العالم من حيث المنشأ ومنها الخاصة بالأحياء المجهرية

1. بكتريا *Escherichia coli* كنموذج دراسي
2. قواعد البيانات الخاصة بالأحياء الممرضة الخطرة التي تشد انتباه الباحثين ومنها قاعدة البيانات الخاصة بالبكتريا *Mycobacterium tuberculosis* المسببة للتدرن الرئوي والذي ظن العالم انه تخلص منه ولكنه عاد وبشكل شرس هذه المرة. والقاعدة الخاصة بهذه البكتريا MYcDB وهي قاعدة بيانات علاقية تم إنشاؤها لإجراء التحاليل المقارنة لجينومات السلالات المختلفة من البكتريا التي تم تحديد توالي جينوماتها. الأساس في القاعدة هي *My. tuberculosis* H37RV و *My. tuberculosis* CDC 1551 فضلا عن *My. bovis* AF 2122/97 و *My. leprae* و *My. avium* subsp. *paratuberculosis* K10 و *My. smegmatis*. MC2155 و TN وهذه القاعدة توفر بيانات عن البروتينات ومواقعها داخل الخلايا وتحدد عناقيد الجينات المتناظرة Orthologous groups مع روابط للعديد من قواعد البيانات الأخرى ذات العلاقة، وتوفر واجهة القاعدة العديد من الإمكانيات الدراسية.
3. قاعدة خاصة بخميرة الخبز *Saccharomyces cerevisiae*
4. قاعدة البيانات الخاصة بفيروس الايدز HIV sequence database .
5. قواعد خاصة بالنباتات سواء المهمة من النواحي الدراسية مثل القاعدة الخاصة برشاد الصخر *Arabidopsis thaliana* ، او المحاصيل الاقتصادية مثل الرز *Oryza sativa* .
6. قاعدة بيانات Worm Base الخاصة بالدودة *Caenorhabditis elegans* النموذج الدراسي للحيوانات.
7. قاعدة بيانات FlyBase خاصة بجينوم حشرة ذبابة الفاكهة *Drosophila melanogaster*.
8. قواعد بيانات للأسماك باعتبارها أحياء مائية وهي خاصة مثل *Danio rerio* السمكة المخططة Zebrafish.
9. قاعدة البيانات الخاصة بالفيران *Mus musculus* تسمى Mouse genome informatics (MGI)
10. ويأتي الإنسان *Homo sapiens* بالدرجة الأولى من حيث قواعد البيانات الخاصة ويأتي رشاد الصخر بعده اذ توجد 5 قواعد بيانات معلنة. اما للإنسان فهناك 7 قواعد بيانات متوفرة على شبكة الانترنت مناشئها مختلفة وتتناول جوانب مختلفة من الجينوم البشري.

قواعد البيانات الخاصة بـ RNA

وهذه القواعد تهتم بخزن البيانات الخاصة بكل من rRNA و tRNA و جزيئات RNA العاملة في التنظيم. وفي العموم تكون قواعد البيانات الخاصة بالـ RNA اقل من تلك الخاصة بالبروتينات و DNA ومع هذا فهناك قواعد متخصصة بالـ RNA الرايبوزومي فهناك أكثر من 100,000 توالي لجزيئات البكتريا الخاصة بـ 16S rRNA. وهناك قواعد خاصة بـ tRNA و tmRNA فضلا عن جزيئات ncRNA غير القابلة للترجمة ، توجد في قاعدة عوائل RNA (RNA families) Rfam ، وتحوي القاعدة على تواليات كاملة للبكتريات التي أعلنت وهمشت جزيئات tRNA و 5S rRNA لها ويمكن البحث في هذه القاعدة باستعمال رقم التسجيل، وتوفر إمكانيات التنبؤ بالتركيب والمراجع ذات العلاقة وكذلك توجد قواعد أخرى منها rRNA SSU & ISU البلجيكية و RDP الأمريكية و miRNA Registry البريطانية.

وبعض قواعد البيانات مزودة بالعديد من الوسائل للتعامل مع التواليات ومنها على سبيل المثال لا الحصر :

- وسائل للتنبؤ بـ miRNA
- وسائل للتنبؤ بعمليات طوي miRNA
- وسائل للكشف عن فعاليات miRNA
- وسائل للكشف عن علاقة miRNA مع الجينات
- وسائل للكشف عن miRNA في الإنسان بصورة خاصة

قواعد البيانات الخاصة بالبروتينات

وتهتم قواعد البيانات هذه بجوانب متعددة متعلقة بالبروتينات، وتمثل اغلبها قواعد ثانوية ومنها Proteome analysis وهي قاعدة تابعة للمركز البريطاني EBI وتضم بيانات إحصائية وافية وتحليلات مقارنة للبروتينات المنتبأ بها للأحياء التي تم تحديد تواليها. اما باقي القواعد فتصنف وفق الغرض الذي وضعت من أجله ومنها:

● **قواعد البيانات الخاصة بالعوائل البروتينية الخاصة:** وهذه قد تكون قواعد بيانات قائمة بذاتها او تكون ضمن قواعد بيانات اكبر مثل القاعدة الخاصة بعوامل الانتساخ TRANSFAC التي تكون موجودة ضمن BioBase.

- **ProWeb** من القواعد المستقلة التي تكون بمثابة دليل لعوائل البروتينات.
- قواعد خاصة للبروتينات العاملة في مجال التحسس او مسارات نقل الإشارات مثل تلك الخاصة ببروتينات G مثل G-Protein Coupled Receptor Database (GPCRDB).
- **PKR** قواعد خاصة بإنزيمات الفسفرة العاملة في نقل الإشارة وغيرها من الفعاليات.
- قاعدة **ENZYME** تابعة للمركز الأوروبي وتكون ضمن الموقع ExpASY الذي يوفر وسائل وإمكانيات كبيرة لدراسة البروتينات من حيث تفاعلاتها وإيجاد المواقع بواسطة برامج توضح مواقعها الخلوية وغيرها مما تتطلبه دراسة الإنزيمات.
- يمكن ان تخصص قواعد البيانات بأنواع معينة من الإنزيمات مثل Peptidases كما في القاعدة البريطانية **MEROPS**. وتوجد قواعد خاصة بالبروتينات التي تحتاج Prosthetic groups مثل الايونات لأداء فعاليتها منها **PROMISE** الأمريكية، او الأخرى التي تحتاج الى الكربوهيدرات **CAZy database** الفرنسية.
- هناك قواعد بيانات خاصة بالمشبطات الإنزيمية مثل مشبطات البروتينات Prolysis الفرنسية.

- هناك قواعد بيانات مثل **ReLiBase** التابعة للمركز الأوروبي وتمثل نظام بيانات لتحليل معقدات المستلمات مع الربائط **Receptor /ligand** للبروتينات الموجودة في المركز.
- بروتينات المناعة هي الأخرى حظيت بقاعدة بيانات مثل **ImMuno GeneTics database** للـ **Ig** (IMGT) والمستلمات **TcRs** وجزئيات **MHC** وهي قاعدة تابعة لجامعة **Alberta** والتي سيأتي ذكرها في فصول قادمة .
- ويمكن إدراج القاعدة **REBASE** الخاصة بإنزيمات القطع التي تعمل على تقطيع الحوامض النووية.
- **قواعد البيانات الخاصة بتصنيف البروتينات:**
هذه القواعد كثيرة ويمكن ان لا تكون مستقلة لوحدها، كما ان بياناتها التي تقدمها يمكن ان تخدم في مجال آخر غير التصنيف مثل **ENZYME** المارة الذكر ضمن محور عوائل البروتينات. ولذا سيتم ذكر المهم منها والمستعملة بكثرة في دراسات المعلوماتية الحيوية ومنها:
• **Pfam** (Protein families database) ويكون تصنيف البروتينات في هذه القاعدة معتمدا على عمليات الاصطاف **Alignment**.
- **SCOP** (Structural classification of proteins) تضم بيانات لتصنيف البروتينات اعتمادا على العوائل الكبيرة للبروتينات وكذلك عمليات طوي البروتينات وغيرها من المعلومات.
- **CATH** (Class, architecture, topology and homologous superfamilies) من قواعد البيانات المستعملة في البحوث والدراسات لتصنيف البروتينات.
- **CDD** (Conserved domain database) تستعمل هذه لإيجاد الدومينات الثابتة وتساعد في إيجاد العلاقات التطورية او فعالية الإنزيمات وهي منضوية تحت **NCBI**.
وفضلا عن ذلك توجد العديد من قواعد البيانات الخاصة لتصنيف البروتينات يمكن اعتمادها لحالات خاصة في دراسات البروتين.
- **قواعد خاصة بتركيب البروتينات:** تهتم قواعد البيانات هذه بتوفير البيانات حول تداخل البروتينات مع غيرها من الجزيئات الحيوية مثل الحوامض النووية مثل **Protein-Nucleic Acid Recognition Database** وقواعد أخرى تهتم بتوفير البيانات الخاصة بتركيب البروتينات ومواصفاتها ووظائفها وحركتها.
- **قواعد البيانات الخاصة بالمسارات الايضية**
توفر قواعد البيانات هذه العديد من المعلومات الخاصة حول تفاعلات الإنزيمات وعلاقات الجزيئات الحيوية مع بعضها وفعاليات الجينات والتعبير عنها. ومنها:
قاعدة **KEGG** (Kyoto Encyclopedia of Genes and Genomes) و **LIGAND** وهي قاعدة بيانات كيميائية لتفاعلات الإنزيمات وتابعة ايضا الى جامعة كيوتو اليابانية.
- ومنها قواعد خاصة بالمسارات الايضية الميكروبية مثل **BioCyc** وكذلك المسارات الايضية الميكروبية الخاصة مثل تفكيك الملوثات كما في قاعدة البيانات الخاصة بجامعة **Minnesota** الأمريكية **UM-BBD**.
- **قواعد بيانات المصفوفات الدقيقة**
بما ان تقنيات المصفوفات الدقيقة **Microarrays** دخلت بثقل كبير في مجال دراسات علوم الحياة لذلك احتاجت الى قواعد بيانات خاصة بها ومنها **ArrayExpress** الموجودة ضمن المركز

الأوربي/البريطاني EBI ، وكذلك GEO (Genome expression omnibus) التي تعد مستودعا" للبيانات التي يتم الحصول عليها من المصفوفات الدقيقة حول التعبير الجيني الموجودة ضمن NCBI وكذلك توجد قواعد بيانات من هذا النوع خاصة بالشركات المنتجة للمصفوفات.

قواعد بيانات الترحيل الكهربائي

نظرا لكثرة استعمال هذه التقنية في التحليل للجزيئات الحيوية أنشأت لها قواعد بيانات مثل Gel Bank التابع الى مركز الطاقة الأمريكي وكذلك SWISS-2DPAGE التابع الى المركز السويسري.

وعلى العموم فهناك مواقع الكترونية متخصصة بإعداد قوائم بقواعد البيانات وأنواعها ويمكن الرجوع اليها لمعرفة التفاصيل الدقيقة عن كل قاعدة بيانات. ومن الجدير بالذكر ان واجهات معظم قواعد البيانات تحوي روابط توصل الى معلومات أكثر، او تحوي إيقونات تساعد في الدراسة ولذلك فان قواعد البيانات في الوقت الحاضر الموجودة على شبكة الانترنت هي ليست فقط درج للبيانات الأولية وانما تحوي وسائل مهمة تسهل الدراسة من خلال واجباتها.

قواعد البيانات الشخصية

بالرغم من كثرة قواعد البيانات الهائلة الا انها في بعض الأحيان لا تخدم غرض بحثي معين لذلك يتم اللجوء الى إنشاء قواعد بيانات خاصة بمشكلة البحث، ويقوم الباحث بكتابة قاعدة البيانات الخاصة به بأي لغة ممكنة وقد توفر لغة SQL سبيلا ملائما لإيجاد القاعدة ومن ثم إجراء الاستعلام عن البيانات ويمكن تطوير برامج خاصة بها.

قواعد البيانات الحيوية Biological databases

أوضح مما ذكر سابقاً ان قواعد البيانات الحيوية تتوزع في مشارب شتى، ولكن هناك بعض المظاهر التي تجمع قواعد البيانات الخاصة بالنواحي الحيوية ، فهي تعتمد على كل ما يتعلق بالجزيئات الحيوية وتراكيبها وتفاعلاتها وتداخلاتها التي سيمر معظمها في الفصل الخاص بالنواحي الحيوية والتي تستعمل في إنشاء قواعد البيانات وكذلك في كتابة برامج الاسترداد او المعالجة والتنفيذ.

و قواعد البيانات الحيوية تستعمل كل أنواع القواعد المذكورة سابقا وهي Flat-file databases و Relational databases و Object-oriented database . وبالرغم من معوقات استعمال النوع الأول الا انها لا تزال مستعملة خاصة عندما يكون هناك القليل من البيانات خاصة وان معطيات هذه القواعد تكون سهلة الفهم من قبل العاملين في مجال علوم الحياة فضلا عن ان هذه القواعد تتضمن اقل مقدار من التصميم مقارنة بقواعد البيانات الأخرى ، وتكون هناك بعض الصعوبات في ربط هذه القواعد نظرا لعدم توافق الصيغ المكتوبة بها البيانات. وفي هذا المجال يمكن ان تقسم قواعد البيانات الحيوية الى قسمين رئيسية اعتمادا على محتوى القاعدة من البيانات الحيوية وهي:

- قواعد بيانات أولية وتحوي على المعلومات الحيوية الأصلية فهي بمثابة أرشيف للتواليات الخام Raw sequences او معلومات حول التركيب الذي تم إيداعه من قبل الباحثين ومن أمثلتها بنك الجينات وبنك بيانات البروتينات PDB (Protein Data Bank) والقواعد الموجودة في المركز الأوروبي EMBL والمركز الياباني DDBJ.
- قواعد البيانات الثانوية تمثل القواعد الحاوية على المعلومات التي عولجت يدوياً او بالحاسوب المعتمدة على البيانات في القواعد الأولية والمعالجات التي جرت على البيانات الأولية لتنتقل الى

قواعد البيانات الثانوية وتختلف قواعد البيانات الثانوية فيما بينها بشكل كبير، فالبعض منها يكون بمثابة أرشيفات بسيطة من التواليات المترجمة من أطر القراءة المفتوحة المعروفة في DNA، في حين تعطي قواعد أخرى إيضاحات وتهميشات إضافية تخص مستويات أعلى مثل معلومات عن التركيب والوظيفة وغيرها وقد تحتاج عملية نقل البيانات الى عمليات حسابية تختلف في درجة تعقيدها ومن هذه القواعد Translated protein sequence database، وتحوي على التهميشات الخاصة بالوظيفة ومن أمثلتها قاعدة البروتينات السويسرية والقواعد الخاصة بالتواليات المسجلة من قبل Margaret Dayhoff.

ومن الجدير بالذكر ان التواليات المودعة في الثمانينات أدخلت الى الحاسوب من قبل الكوادر العاملة في إدارة قواعد البيانات، اما في الوقت الحاضر فان التواليات تخضع لعمليات تهيئة لتكون صالحة للنشر في المجالات العلمية وكذلك تسهيل الوصول اليها خاصة في المراكز NCBI و EMBL و DDBJ التي تتبادل المعلومات على مستوى يومي، والتي يمكن الوصول اليها عبر شبكة الانترنت، وبالرغم من احتواء المراكز الثلاثة قواعد البيانات ذاتها إلا ان الصيغ Format تكون مختلفة قليلا. وكما ذكر أعلاه فان ربط قواعد البيانات المستعملة في مجال العلوم الحيوية ليس بالسهل نتيجة لاختلاف الصيغ والتركيب، ولذلك عمدت بعض الجهات الى استعمال طرق لربط قواعد البيانات بالشبكات باستعمال Specification language تسمى Common object request broker architecture (COBRA) والتي تسمح لبرامج قواعد البيانات من مواقع مختلفة بالتواصل عبر الشبكات باستعمال ما يسمى بوسيط الواجهة Interface broker دون الحاجة الى فهم تركيب كل قاعدة بيانات. كما ان بروتوكول Extensible markup language (XML) يساعد في جسر الهوة بين قواعد البيانات، والآليات التي تعمل بها هذه البرامج هي في تطور مستمر.

المعوقات المتعلقة بقواعد البيانات الحيوية

من أهم المشاكل المتعلقة بهذه القواعد هو اعتمادها الكبير على المعلومات الخاصة بالتواليات والتهميشات المتعلقة بها اذ يمكن ان يكون تهيمش الجين غير صحيح او غير متكامل اي لا يكون على درجة عالية من الدقة فضلا عن تجاهل الأخطاء التي يمكن ان تحدث أثناء تحديد التواليات. كما ان هناك تكرار كثير في قواعد البيانات الأولية، كل هذه الحالات تؤدي الى انتقال الأخطاء من قاعدة الى أخرى مؤدية الى تضخيم الأخطاء.

وأكثر الأخطاء خطورة هي التي تحدث أثناء تحديد تواليات النيوكليوتيدات، بعض هذه الأخطاء تؤدي الى حرف أطر القراءة Frameshift التي تجعل من تعريف او تحديد الجين في غاية الصعوبة او تجعل من ترجمته مستحيلة، كما ان في بعض الأحيان تتلوث التواليات من نيوكليوتيدات تعود في الأصل الى نواقل الكلونة المستعملة لتضخيم قطع DNA، وكل هذه الأخطاء قلت وزادت نوعية التواليات جودة بعد عقد التسعينات من القرن المنصرم، لذا كان من الأشياء الواجب الانتباه اليها تواريخ تحديد التواليات.

ومن المشاكل الأخرى زيادة وفرة التواليات التي تؤثر في قواعد البيانات الأولية التي تمثل المركز الرئيس للمعلومات المخزونة وكذلك المركز لتوزيع التواليات الخام والمعلومات حول التركيب على باقي قواعد البيانات البايولوجية، وهذه الوفرة تعود الى أسباب عدة منها:

- الإيداع المتكرر لتواليات متشابهة او متداخلة من قبل الباحث نفسه او باحثين مختلفين.
- تنقيح التهميشات وإعادة صياغتها.

● الإدارة السيئة لقواعد البيانات. وكل هذه وغيرها تؤدي الى جعل قواعد البيانات كبيرة جدا وغير ملائمة لاسترداد المعلومات منها. لذلك اتخذت بعض الجهات المعنية خطوات لمنع التكرار فمثلا مركز NCBF انشأ قاعدة مشذبة nr (non-redundant) وبعض الأحيان تسمى بقاعدة التواليات المرجعية (Refseq) اذ تم فيها دمج تواليات النيوكليوتيدات المتماثلة من الكائن نفسه في مدخل واحد وكذلك الحال مع البروتينات. والتغيرات في التواليات من الكائن نفسه والتي تحوي على اختلافات بسيطة وقد تكون ناتجة عن أخطاء أثناء تحديد التواليات عوملت على أساس انها مداخل ذات علاقة ولكن مميزة الواحدة عن الأخرى وهذه القواعد المنقحة يمكن ان تعد على انها قواعد بيانات ثانوية واتخذت إجراءات مماثلة مع قواعد بيانات أخرى ، وتمثل قاعدة البروتينات السويسرية قاعدة حاوية على القليل من التكرار مقارنة بالقواعد الأخرى .

ومن المشاكل الأخرى هي التهميش الخاطئ، فبعض الأحيان يكون توالي جين معين موجود تحت مسميات مختلفة مما يؤدي الى وجود عدة مداخل وحصول إرباك في المعلومات ولازالت هذه المشكلة قائمة وان كانت هناك بعض الخطوات التي اتخذت لتوحيد تسمية الجينات عالمياً وكذلك إيجاد مفردات لغة موحدة لوصف الجينات والبروتينات وهو ما أدى الى نشوء Gene ontology (الذي ذكر في موضع آخر) ، وتضاف الأخطاء الطباعية ايضا والاختلاف في وجهات نظر المؤلف او المؤلفين حول وظيفة الجين او البروتين المودع في قواعد البيانات ، ولذلك يلاحظ ان الأخطاء في التهميش تكون خطرة جدا خصوصا في الأونة الأخيرة التي اعتمدت طرق التشابه ومطابقة التواليات لما موجود في قواعد البيانات . وفي حالة تطابق جين او بروتين جديد مع آخر موجود في قواعد البيانات والحاوية على أخطاء، فان الأخيرة ستسحب على التواليات الجديدة ، بعد ان هجر الباحثون اعتماد الطرق المختبرية .

أنظمة الاسترداد Retrieval systems

يعد استرداد البيانات من المهام الرئيسة التي تنطاط بالمعلوماتية الحيوية فضلا عن استرداد البيانات يحتاج العامل في مجال المعلوماتية الحيوية الى استرداد معلومات أخرى سوى التواليات. ولعملية الاسترداد مشاكل كثيرة يمكن ان تعيق عمليات البحث وذلك ناتجا من عدة أسباب ويمكن توضيح بعضها في الآتي :

- عند البحث يجب ان يكون التعريف بالشئ المراد البحث عنه واضحا وهذا يكون معتمداً على البيانات الإضافية ولكن مثل هذه قد لا تكون معروفة بالتحديد، فمثلا ما هو الرقم الهيدروجيني الذي يحدد البيئة حامضية ام لا، ما هي درجة حرارة الغرفة، ما هو تعريف جين الضراوة وغيرها من المفردات.
- اللغة المستعملة في البحث هي ليست قياسية ، وحتى عند استعمال اللغة الانكليزية كلغة تواصل فهناك اختلافات في الإملاء مثلا النمط الانكليزي الأمريكي يكتب اللون Color في حين النمط البريطاني يكتبها Colour وكذلك الحال مع الكبريتات التي تكتب Sulfate او Sulphate وتمثل هذه مشكلة في تحليلات الحاسوب والحاسوب كآلة لا يمكنه التعرف على المترادفات او الخلافات الإملائية. وكذلك تفسير المقاطع اللغوية، لذلك فان النص او السياق اللغوي يؤثر في الفكرة فمثلا Media تعني الإعلان ووسائله في العموم في حين تعني الوسط الزراعي لتنمية الأحياء المجهرية للعاملين في مجال الأحياء المجهرية. لذلك كان لابد من وجود بيانات إضافية كي تتمكن الماكنة (الحاسوب) من إرجاع الكلمات المستعملة ضمن النطاق المطلوب.

ونظرا لتعقيد البيانات الإضافية والعلاقات المتداخلة بينها نشأت مشكلة المعنى والتصنيف للأفكار وهذه هي مشكلة دائمة في علوم الحاسوب التي تعد مشكلة عامة، والزيادة في طلب استرجاع المعلومات وتصنيفها وعمل روابط لها أدت الى نشوء تخصص علمي جديد هو Ontology والذي يعني في علوم الحاسوب هو تمثيل أساسي وواضح اي مقبول ومحدد المقاييس لمجموعة من الأفكار من حيث معانيها وارتباطها مع بعضها بعلاقات واضحة، كما في التصنيف الهرمي Hierarchical classification والذي يعني ترتيب الأفكار من العامة الى أفكار خاصة. و Ontologies يمكن ان يستعمل من قبل الحاسوب والمبرمج في عمليات:

- البحث
- الوصف
- التصنيف

• تعريف العلاقات

وتوجد صيغ وأرقام خاصة لكل موضوع معروفة يمكن للمبرمج الرجوع اليه على صفحات الانترنت.

والحقيقة ان هذا المجال (Ontology) يساعد في وضع صيغ أكثر ملائمة للبيانات الإضافية وتقوم بعض المؤسسات مثل Genomic standards consortium بالعمل على وضع الصيغ العامة فيما يخص التواليات لتكون عامة للجميع والتي تتمثل بـ Minimum information about a genome sequence/meta genomic sample (MIGS/MIMS) التي نشرت عام 2008 ، وفيها اقل ما يمكن من المعلومات لوصف التواليات المشتقة من الكائن او الفيروسات او العضيات او النماذج البيئية (Metagenomic samples) وهي الآن مستعملة في صيغ ملفات الجينوم ضمن ما يسمى بـ Genomic contextual data markup language (GCDML) وأنظمة الاسترداد تسهل الدخول الى عدة قواعد بيانات للحصول على نتائج البحث، وعند استخدام استعلامات معقدة في قاعدة ما لا بد من استخدام العلاقات الجبرية Boolean operators وفيه يتم استعمال عبارات الربط بين الكلمات المفتاحية (الكلمات الرئيسية) ومن العبارات المنطقية (و) AND (لتضييق عمليات البحث) ، (او) OR (لتوسيع عمليات البحث)، (كلا) NOT (لتضييق عمليات البحث) لغرض توضيح العلاقة بين الكلمات المفتاحية المستعملة في البحث، اذ تعني AND ان نتائج البحث يجب ان تحوي الكلمتين و OR تعني ان نتائج البحث يجب ان تحوي إحدى الكلمتين او الاثنين معا اما NOT فتعني استبعاد إحدى الكلمتين او الاثنين معا. كما يمكن في بعض الأحيان استعمال الأقواس () لتحديد الفكرة فيما اذا كان هناك عدة كلمات وعلاقات متضمنة في عملية البحث، وتفهم الحاسبة اي جزء من البحث يجب تنفيذه أولا حيث يتم تنفيذ البحث أولاً على المفردات داخل الأقواس كما يمكن استعمال المزدوجات او علامات الاقتباس " " لتحديد مقطع ما، وأغلب قواعد البيانات البايولوجية تستعمل بعض هذه العلاقات الجبرية .

الأغراض من الاسترداد

تعد عمليات استرجاع او استرداد المعلومات من المهام الأساسية للمعلوماتية الحيوية كما ذكر آنفاً، وهذا الاسترداد سيساعد في التعامل والتلاعب ما مسترد واشتقاق المعلومات والعلاقات التي هي في النهاية لها علاقة بتواليات البروتينات او DNA والتي يمكن ان تستغل في التجارب المخبرية او الدراسات الحاسوبية . وهذا الاستغلال سيؤدي الى اكتشافات جديدة وزيادة المعلومات حول البيانات الأصلية والى إيجاد بيانات إضافية . فبعد استرداد المعلومات تأتي مرحلة تحليل البيانات

مثل أجراء التحويلات عليها واغناءها ومدها لتشمل آفاق جديدة باستعمال الحاسوب وبالتالي تمثيلها وتفسيرها، ومرحلة التفسير هي التي تحتاج الى تدخل المستخدم الذي له خلفية بايولوجية تساعد في تفسير النتائج، ثم خزن البيانات والعلاقات المستنبطة للاستعمال المستقبلي كما يحصل في اكتشاف أهداف جديدة للأدوية في الأحياء الممرضة التي أصبح معظمها مقاوماً للمضادات الحيوية والأدوية قيد الاستعمال. ولكن المهم ان تبقى حقيقة ان العمل في محيط الحاسوب لا يجعل العلم أكثر دقة او اقل صعوبة ومع أخطاء قليلة فهو مثل الحال مع العامل في المختبر.

استرداد البيانات لإجراء تحليل الحاسوب

يعتمد استرداد واستخلاص البيانات من أماكن وجودها اعتماداً كلياً على ماذا يريد المستخدم؟ فالبيانات موجودة في ملفات وأوعية يجب ان تستخلص منها، ومن الأوعية والملفات التي توجد فيها هي ملفات النصوص Textfiles وهي ملفات حاسوب تحوي على نص مكتوب قد يمثل توالي لجزيئة مرتبة بحروف متسلسلة، فضلاً عن معلومات أخرى قد توجد في الملفات الإضافية والتي تكون مفصلة عن البيانات الأولية. وملف النص تنتهي تسميته بـ txt وان كانت هناك بعض الملفات تنتهي بـ fas للـ FASTA و gb او gbk لبنك الجينات و doc (Word TM document) وهذا الأخير لا يكون فايل نصي وانما يعطي إشارات للحاسوب حول تنسيق الملف مثل الخط وإعداد الصفحة وغيرها، وعند عدم وجود هذا الملف لا يمكن قراءة الملفات على الحاسوب.

وعلى العموم هناك طرق وبرامج أخرى لاستخلاص المعلومات من الجداول او المخططات يمكن ان ترتب من قبل المبرمج لتسهيل على المستخدم الوصول الى المعلومات المطلوبة.

برامج وأنظمة الاسترداد

تعتمد هذه على البنية التركيبية لقواعد البيانات المراد الاسترداد منها، فبعض البرامج تكون خاصة باسترداد البيانات من القواعد العلاقية ويمكن ان تستعمل مع هذه لغة SQL والتي يجب ان يكون الاستعلام فيها على دقة عالية. ويمكن ان تستعمل برامج او قواعد بيانات تتوسط في استرداد المعلومات مثل My SQL او Microsoft Access او غيرها.

والطريقة الأخرى استعمال صفحات الانترنت والتي يمكن الوصول اليها باستعمال المتصفحات Browsers، ومنها Firefox ، Safari ، Internet Explorer ، Opera وغيرها. ووجود قواعد البيانات على صفحات Web والوسائل المتعلقة بها مهمة جداً ولكن لها مشاكلها ففي بعض الأحيان يصعب الحصول على البيانات منها. لذلك استعملت مزودات الخدمة وهو يشبه تصفح Web (الشبكة) ولكنها تعتمد في استعمالها على الحاسوب وليس بواسطة المستخدم وتسهل استخلاص البيانات من صفحة الشبكة باستعمال الأوامر التنفيذية مثل النقر (Click) او مليء الحقول الفارغة وتشغيل البرامج.

Entrez

نظام استرداد تابع للـ NCBI ويمثل نقطة البداية في البحث في قواعد NCBI ويكون البحث من خلاله سهلاً ولكن محدود. يكون النظام مدخلاً للبحث في قواعد الحوامض النووية مثل بنك الجينات وكذلك قواعد البروتينات، فضلاً عن وجود روابط تخص المراجع PubMed (التي سيأتي ذكرها لاحقاً) و رابط لعمليات التصنيف وغيرها، لذلك لا يحتاج المستخدم زيارة عدة مواقع بشكل منفصل وانما يمكنه استخدام الروابط. وتزداد كفاءة البحث باستعمال هذا النظام عند معرفة وفهم المعالم الرئيسة لهذه العملية فمثلاً وجود المحددات Limits عند اختيارها يساعد في تحديد

البحث في أجزاء معينة من قاعدة بيانات كبيرة أو البحث في قواعد البيانات الثانوية. كما انها تساعد في تحديد مجال البحث في قاعدة البيانات مثل البحث عن معلومات في حقل معين أو باحث معين أو تاريخ محدد أو البحث عن معلومات تخص أجزاء معينة مثل البلاستيدات الخضراء أو RNA أو DNA.

أما الخيار الآخر خاصية الإظهار والفهرسة Preview/Index ويربط عمليات البحث بالعلاقات الجبرية (التي مر ذكره في موقع آخر) الذي يربط الكلمات المفتاحية ليقوم بعمليات بحث جديدة ويمكن البحث عن جين معين يذكر اسم الجين الذي يتعرف عليه النظام أو رقم التسجيل، ويوفر النظام خدمة أخرى وهي التاريخ إذ يزود الاختيار بسجل عن عمليات البحث السابقة ليتمكن المستخدم من مراجعتها واستخدامها لإغراض أخرى. فضلا عن ذلك يوفر النظام مخزن للنتائج Clipboard لإظهارها فيما بعد ولكن لوقت محدد جدا.

ويوفر نظام الاسترداد Entrez مواقع مهمة مثل Online Mendelian inheritance in man (OMIM) التي تقوم بتوفير البيانات حول الأمراض الوراثية والاضطرابات الوراثية في الإنسان وكل مدخل فيها يحوي على ملخص للمعلومات المتوفرة حول مرض معين والجينات ذات العلاقة ومواقعها على الكروموسومات، لذلك تعد منطلقا جيدا لدراسة علاقة الجينات بالأمراض. فضلا عن توفير روابط الى قواعد بيانات أخرى مثل تغايرات النيوكليوتيد المفرد والتراكيب الوراثية لأحياء معروفة مثل الفيران والكلاب والنباتات المهمة.

Search / retrieval system SRS

نظام استرجاع تابع للمركز الأوروبي EBI وهو أكثر شمولية من نظام Entrez. وتوفر واجهته إمكانيات كبيرة للحصول على المعلومات مثل الدخول الى أكثر من 80 من قواعد البيانات التي تتمثل بقواعد بيانات للتواليات وكذلك يوفر مدخلا للمسارات الابضية وعوامل الانتساخ والجينومات الكاملة (المحدد تواليها) والخرائط الوراثية والطفرات ومواقعها، فضلا عن إمكانيات للتنبؤ بالتركيب الثلاثي للبروتينات الموجودة ضمن EMBL.

ولكل قاعدة منضمة تحت النظام وصف بسيط لها مثل آخر إصدار لها وغيرها من المواصفات. وبطبيعة الحال تحوي الواجهة على روابط تؤدي الى مواقع أخرى قد تكون خارج النظام الأوروبي. ويحوي الموقع على القواعد الحديثة تحت مسمى EMBLNEW، إذ إن إطلاق وتجديد القواعد يتم بشكل دوري لكل ثلاثة أشهر. وتستخدم في النظام طرق خاصة للحصول على المعلومات فمثلا مدخل All Text يعد استعلام عام لكل الكلمات المفتاحية المدخلة من قبل المستخدم، في حين إن إدخال ID يعد رقم التسجيل الرئيس للتوالي. وكما هو الحال مع Entrez يحوي النظام على وسيلة للبحث عن الأقسام Division التي لها علاقة وثيقة مع التصنيف، وفي هذا النظام يمكن البحث عن التوالي المحور أي عن الإصدارات الجديدة (Seq version). فضلا عن إمكانية البحث عن الجزيئات مثل RNA، DNA وغيرها. ويمكن البحث فيه باستعمال الكلمات المفتاحية والأخيرة يتم إعدادها من قبل العاملين على القاعدة لزيادة المعلومات لذلك قد لا توجد في كل المداخل ويمكن البحث أيضا عن طريق استعمال اسم الكائن الحي. أما البحث عن المراجع فله طريقة خاصة يتم بذكر اسم الشهرة للكاتب أو المودع للتوالي وذكر الأحرف الأولى وبدون فواصل. كما يمكن البحث في هذا المجال بذكر اسم المجلة بالأحرف الكبيرة فضلا عن ذكر المجلد والصفحة الأولى وسنة النشر بين مزدوجتين " ". أما عند استخدام التاريخ للبحث فيكون ذكر

السنة الرباعي الأرقام والشهر واليوم بشكل متواصل YYYYMMDD مثل 20100712 ويمكن ايضا استعمال صيغ أخرى لكتابة التاريخ.

وفي هذا النظام يمكن البحث عن التواليات بتخصيص طولها "Seqlength" وهناك إمكانيات أخرى للبحث عن البروتينات. فضلا عن ان واجهة النظام توفر الإمكانية لإيداع التواليات بشكل مباشر.

وطرق البحث المذكورة أعلاه تقع ضمن أحد السياقات المستخدمة في النظام فهي اما ان تكون سريعة Quick text search وهذه تستخدم نافذة استعلام واحدة Query box، او تكون ضمن عمليات البحث القياسية Standard text research وتسمح باستعمال أربعة حقول ترتبط العلاقات الجبرية. اما السياق الآخر Extended text research فيسمح باستعمال مؤشرات متباعدة وحقول متباعدة.

تكون النتائج حاوية على التوالي المستعلم عنه والتهميش الخاص به وروابط الى المراجع الخاصة ومسارات الايض وقواعد البيانات الأخرى . وتكون ملفات التقارير الناتجة حاوية على تفاصيل ومداخل كثيرة كلها تؤدي الى زيادة المعلومات الممكن الحصول عليها.

وينتج نظام الاسترداد هذا صيغ ملفات تعرف بـ EMBL format خاصة بالمركز الأوروبي تحوي معلومات تخص التواليات المستعلم عنها منها التعريف ID وترتب التواليات كل 10 سوية كما موضح في الشكل 13.2:

ID	AB000263 standard; RNA; PRI; 368 BP.
XX	
AC	AB000263;
XX	
DE	Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.
XX	
SQ	Sequence 368 BP;
	acaagatgcc attgtccccc ggccctcctgc tgctgctgct ctccggggcc acggccaccg 60
	ctgccctgcc cctggagggt ggcccccaccg gccgagacag cgagcatatg caggaagcgg 120
	caggaataag gaaaagcagc ctccctgactt ttctcgcttg gtggtttgag tggacctccc 180
	aggccagtgc cgggcccctc ataggagagg aagctcggga ggtggccagg cggcaggaa 240
	gcgcaccccc ccagcaatcc gcgcgccggg acagaatgcc ctgcaggaaac ttcttctgga 300
	agaccttctc ctctgcacaa taaaacctca cccatgaatg ctacagcaag ttaattaca 360
	gacctgaa 368
//	

شكل 13.2 : صيغة ملفات المركز الأوروبي EMBL

وهناك العديد من الصيغ التي يمكن ان تخص كل برنامج يتم توضيحها بمرفقات البرامج ولا يمكن درجها بالتفصيل.

البحث في قواعد البيانات

ذكر سابقا الأغراض التي تستعمل لها قواعد البيانات ويكون البحث فيها اما:

- البحث عن التواليات سواء البروتين او DNA او RNA وهذا يكون بواسطة برامج خاصة، وقد تكون وسيلة البحث مزودة مع القاعدة نفسها.
- البحث عن النصوص وهنا يكون البحث عن التهميش والمعلومات الإضافية ويمكن استعمال أنظمة البحث الشائعة مثل Entrez او SRS او Lookup او غيرها وتوفر المواقع روابط لمداخل

ملائمة للمؤشرات المراد البحث عنها، واعتيادي ان تختلف الأنظمة في طريقة البحث او الروابط التي تستعملها.

والمعروف ان قواعد البيانات الخاصة بـ DNA هي اكبر من تلك الخاصة بالبروتينات، والأولى تنمو بشكل اكبر، وما يخص DNA تكون التواليات من مكررات لأربعة (n) حروف هي القواعد النتروجينية (T, G, C, A) وما تترجم له من بروتينات تمثل (n/3) من حروف الحوامض الامينية البالغة 20 حرف. وعند البحث عن توالي DNA يتم مقابل القواعد المتوفرة ولكن يمكن ترجمة تواليات DNA الى تواليات بروتين ثم البحث عن الأخيرة في قواعد البروتينات ، ولو ان هذا يمكن ان يؤدي الى فقد بعض المعلومات، فضلا عن ان تواليات البروتينات هي الأكثر ثباتاً من النواحي البايولوجية مقارنة بتواليات DNA ، ولذلك فان الترجمة تعطي نتائج أفضل وهذا ما يوصى به عن البحث عن تواليات DNA مشفرة. وعمليات البحث تخضع لمؤشرات تقييم وذلك لان البحث يهدف الى الحصول على بيانات او معلومات ذات علاقة وتجنب البيانات التي ليس لها علاقة، ومؤشرات التقييم تتمثل بـ :

- الحساسية Sensitivity والتي تعرف على انها الموائمة الحقيقية True matching او True positive ، فالطرق الحساسة جدا هي التي تجد كل التوائمات الموجبة الحقيقية ولكنها قد تحوي على عدد من النتائج الموجبة الكاذبة False positives.
- التخصصية Specificity وهي القابلية على رفض التوائمات الموجبة الكاذبة وأكثر عمليات البحث المتخصصة تنتج من التوائمات الحقيقية ولكن قد تحوي على نتائج سالبة كاذبة False negatives.

نصائح لعمليات البحث في قواعد البيانات

هناك بعض الإرشادات التي تساعد في عمليات البحث منها:

- استخدام آخر وأحدث إصدار لقاعدة البيانات.
- استخدام وسائل البحث الموجهة والسريعة Heuristic ثم الانتقال الى وسائل البحث الدقيقة مثل البرامج المعتمدة على البرمجة الديناميكية Dynamic programming والتي تتوفر لكل منها برامجها الخاصة.
- عند وجود إمكانية يفضل استعمال تواليات البروتينات بدلا عن ما يقابلها من تواليات DNA.
- الانتباه الى القيم الإحصائية المستعملة مثل E value (التي سيأتي ذكرها لاحقا) ويفضل ان تكون اقل من 0.05 وهذا المستوى ذات أهمية إحصائية مهمة في الاهتمامات البايولوجية.
- الانتباه الى التراكم غير الطبيعية لتواليات الاستعلام عند إدخالها للبحث عن التوائمات المشابهة لها في قواعد البيانات لان التراكم غير الطبيعية تؤدي الى انحياز في الدرجات المسجلة Biased scoring ، والدرجات هي التي ترجح صحة البحث.
- يفضل البحث باستعمال تواليات استعلام طويلة مثلا للـ DNA تكون اكبر من 1000 من القواعد النتروجينية وللبروتينات بحدود 200 حامض أميني.
- وفي حالة احتواء توالي الاستعلام على المكررات Repeats فيجب إزالتها ولهذا توجد برامج خاصة، ثم معاودة البحث في قواعد البيانات.

وفي جميع الأحوال سواء كان بحث في قاعدة بيانات فإن البحث يفضل ان يكون باستعمال أكثر من قاعدة، او كان تنفيذ برنامج ما فيفضل عندها دراسة البرنامج والأسس التي بني عليها ومن ثم استعمال أكثر من برنامج للوصول الى نتائج معينة.

أهم قواعد البيانات ومتعلقاتها

قواعد البيانات كما ذكر سابقا يمكن ان تضم في الوقت نفسه أكثر من نوع من البيانات مثل التواليات لجزيئات مختلفة، فضلا عن البيانات الإضافية وهناك بعض القواعد التي لا يمكن تجاوزها عند التعامل في المعلوماتية الحيوية والانترنت ومنها:

بنك الجينات GenBank : قاعدة بيانات تابعة لـ NCBI تحوي على التواليات المختلفة وقد أنشأت عام 1982 وتتضاعف محتوياتها كل 18 شهر تقريبا وتم تحويلها الى مركز NCBI عام 1992.

وتحوي على مجمع كبير للحوامض النووية التي تم تهميشها لمعظم الكائنات المدروسة ومحتوياتها تشمل DNA الجينومي و mRNA ، cDNA ، ESTs وقواعد لتواليات البروتينات وخاصة المترجمة من DNA ويوفر المركز روابط لعدد من قواعد البيانات المشابهة الأخرى خارج نطاق NCBI التي تكون مهمة للعاملين في مجال علوم الحياة.

وقاعدة بنك الجينات المنضمة الى مركز NCBI لها بعض الخصوصيات مثل GenBank Accession number الذي يختصر الى Accession number رقم التسجيل وهو الرقم الذي يعطي للجزيئة الحيوية عند إيداعها في البنك وهو يعد Primary key اي المفتاح الرئيس الذي يستعمل عند بناء قواعد البيانات وهذا الرقم يكون مشتركا مع المركز الياباني DDBJ والمركز الأوروبي EMBL ولذلك يكونه خاصا ومفردا في تعريف الجزيئة ويستعمل في استرداد التواليات او الحصول على المعلومات. وهناك طريقتان للبحث عن التواليات في هذه القاعدة وهي اما:

- استعمال الكلمات المفتاحية

- استعمال التوالي حيث يتم البحث عن ما يشابهها باستعمال برنامج BLAST.

وللقاعدة صيغة خاصة للملفات التي تمثل نتائج البحث، وتكون الصيغة من أكثر الصيغ احتواءا على المعلومات وبشكل ملفات نصية عادية سهلة القراءة والملف الناتج يحوي على ثلاث مقاطع هي الترويسة Header، الخصائص Features والتوالي الأصلي كما موضح في الشكل 14.2:

وهناك عدد من الحقول في الترويسة والخصائص ولكل منها تعريف Identifier خاص (لسهولة الفهرسة بالنسبة للحاسبة)، ومعرفة تركيب الملفات يساعد في تصميم طرق فعالة للبحث.

مقطع الترويسة:

- السطر الأول : Locus يحوي على التعريف الخاص بالتوالي في قاعدة البيانات وموقع التوالي في القاعدة (وهو ليس الموقع الخاص بالكروموسوم) وبعض المعلومات الأخرى مثل طول التوالي، فاذا كان DNA يكون الرقم مثلا 411bp لعدد القواعد النتروجينية اما اذا كان بروتين فيكون 411aa ، ثم يذكر نوع الجزيئة مثل mRNA او اي شيء آخر، وكذلك صفة الجزيئة هل هي مفتوحة Linear او حلقية او صفة أخرى ثم يليه مقطع مكون من ثلاث حروف كبيرة تمثل الشفرة المستعملة في أقسام بنك الجينات وفي هذا المجال تحوي القاعدة على حوالي 17 مرتبة لإغراض الخزن وليس لها علاقة قوية بالنواحي العلمية ومنها المذكورة في الشكل 14.2

Header
الترجمة

Features
المعاصر

الترجمة
الأصلي

LOCUS Q9ZGE9 440 aa linear BCT 15-JUN-2002
DEFINITION Light-independent protochlorophyllide reductase subunit N (LI-POR subunit N) (DPOR subunit N).
ACCESSION Q9ZGE9
VERSION Q9ZGE9 GI:18203677
DBSOURCE swissprot: locus BCHN_HELMO, accession Q9ZGE9;
class: standard.
created: Oct 16, 2001.
sequence updated: Oct 16, 2001.
annotation updated: Jun 15, 2002.
xrefs: gi: 3820536, gi: 3820556
xrefs (non-sequence databases): InterProIPR000510, PfamPF00148
KEYWORDS Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
SOURCE Heliobacillus mobilis
ORGANISM Heliobacillus mobilis
Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae; Heliobacillus.
REFERENCE 1 (residues 1 to 440)
AUTHORS Xiong,J., Inoue,K. and Bauer,C.E.
TITLE Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from Heliobacillus mobilis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (25), 14851-14856 (1998)
MEDLINE 29061957
FUTMED 2943979
REMARK SEQUENCE FROM N.A.
COMMENT -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <http://www.expasy.ch/sprot> and <http://www.ebi.ac.uk/sprot>

[FUNCTION] Uses Mg-ATP and reduced ferredoxin to reduce ring D of protochlorophyllide (Pchlde) to form chlorophyllide a (Chlide) (By similarity). This reaction is light-independent.
[PATHWAY] Light-independent bacteriochlorophyll biosynthesis.
[SUBUNIT] Protochlorophyllide reductase is thought to be composed of three subunits; bchL, bchN and bchB. Could form a heterotetramer of two bchB and two bchN subunits.
[SIMILARITY] BELONGS TO THE BCHN / CHLN FAMILY.
FEATURES Location/Qualifiers
source 1..440
/organism="Heliobacillus mobilis"
/db_xref="taxon:28064"
gene 1..440
/gene="BCHN"
Protein 1..440
/gene="BCHN"
/product="Light-independent protochlorophyllide reductase subunit N"
/EC_number="1.18.-.-"
ORIGIN

```
1 merverengc fhtfcpiasv awlhrkikds fflivgthtc ahfigtaldv mvyahsrfgf
61 avleesdlvs aspteelgkv vqgvvdewhp kvifvlstcs vdilkmldlv sckdlatrfg
121 fpvlpastsg idrsitggd avlhallpfv pkeapavepv eekprwfsf gkesekae
181 parnlvlvlg vtdstiqglg welkqlglpk vdvfpdgdv kmpvinegtv vvpigpylnd
241 tlatirrrrr akvlstvipi gpdgtarfle aiclfegldt srikekeaga wrdleplqi
301 lrgkkimflg dnllelpfar fltscdvqv eagtpyihsk dlqgelellk erdvrviesp
361 dftkqlrmq eykpdvrvag lgicnpleam gfttawsief tfaqingfvrn aidliklftk
421 pllkrqalme hgwaegwle
```

شكل 14.2 : صيغة ملفات قاعدة بنك الجينات GenBank

PLN خاصة بتواليات النباتات، الفطريات، الطحالب

PRI خاصة بالتواليات المشتقة من رتبة الثدييات

MAM خاصة باللبائن

BCT خاصة بالبكتريا

EST خاصة بالتواليات المعبر عنها EST

ROD تواليات القوارض

VRT تواليات الفقريات

INV تواليات اللافقريات

VRL تواليات الفيروسات

PHG تواليات العاثيات

SYN تواليات صناعية

UNA تواليات غير مهمة

PAT تواليات مسجلة ببراءات اختراع

STS تواليات Sequence tagged sites
 GSS تواليات Genome survey sequences
 HTG High-throughput genomic sequences
 HTC تشبه تواليات HTG ولكنها غير مكتملة
 ENV تواليات من نماذج بيئية

ويليها التاريخ الذي أعلن فيه التوالي للاستعمال العام، إذ إن إيداع التوالي لا يعلن عنه مباشرة وإنما يحتاج إلى مدة للتأكد والتشذيب والمقارنة والمسجل في العادة تاريخ الإعلان له .

● السطر الثاني التعريف Definition يتضمن ملخص المعلومات عن سجل التوالي ويشمل الاسم العلمي للكائن المشتق منه التوالي فيما إذا كان معروفا وكذلك فيما إذا كان التوالي كاملا أو جزئي من تركيب معين مثل توالي كامل لشفرات قابلة للترجمة Complete CDS أو جزء منها، ويستعمل الرمز CDS للتواليات المشفرة في قاعدة بنك الجينات.

● السطر الثالث: رقم التسجيل وهو الرقم الذي يعطى للجزيئات عند إيداعها لأول مرة في القاعدة ويبقى الرقم مرافقا للتوالي بشكل دائم وهو الذي يذكر عند نشر البحوث ويكون الرقم بصيغتين الأولى مكونة من حرفين مع خمس أرقام أو من حرف مع ست أرقام.

● السطر الرابع: ويذكر فيه رقم التسجيل فإذا كان التوالي لم يحصل عليه تغيير فيكون مثلا AF00027.1 أما إذا حصل تحديث أو تغيير فيبقى رقم التسجيل ولكن يصبح AF00027.2 ، ومعه يذكر Geninfo identifier (gi) ، وعند حصول أي تحديث أو تغيير على التوالي يتغير رقم gi . والبروتينات المترجمة من توالي معين لها gi مختلف عن توالي DNA الذي اشتقت منه. ● السطر الخامس: يمثل المصدر الذي أخذت منه البيانات إذ تذكر قاعدة البيانات وبعض المعلومات عنها.

● السطر السادس: يحوي الكلمات المفتاحية التي يمكن أن تساعد في البحث عند إعادة الكرة، وهي تعطي معلومات عن الجزيئة التي أعطيت نتائجها.

● المصدر وفيه يذكر اسم الكائن الذي اشتقت منه الجزيئة التي هي قيد البحث والدراسة. يلي ذلك سطر خاص بالكائن إذ يذكر الاسم العلمي وبعض المعلومات الخاصة عن التصنيف أو يذكر النسيج أو العضو الذي اشتقت منه الجزيئة المستعملة في الاستعلام.

● سطر المرجع تذكر المراجع الخاصة بدءاً من المؤلف، عنوان المقالة، والمجلة التي نشرت فيها المقالة ، المجلد وأرقام الصفحات وسنة النشر. ويذكر للمقالة رقم Medline وكذلك PubMed (قاعدة المراجع) وهو رابط لقواعد المراجع ليسهل الوصول إلى المقالة، وبعض الأحيان تكون المعلومات تجريبية وغير منشورة فتبقى هذه الحقول فارغة.

ثم تذكر بعض التعليقات التي تخص عمليات معالجة التواليات قبل إعلانها. وفي نهاية الجزء الأول من ملف تقرير بنك الجينات تذكر الوسيلة التي يمكن بواسطتها التواصل مع الشخص الذي أودع التوالي في البنك.

مقطع الخصائص Features

ويمثل المقطع الثاني من ملف التقرير يحوي على معلومات حول الجين ومنتجاته والمنطقة من التوالي الذي يشملها التقرير، مثل CDS المناطق المشفرة وكذلك توالي الحوامض الأمينية التي تشفر لها، وفي حالة الخلايا حقيقية النواة يحوي هذا المقطع على مواقع الاكسونات. وربما تعطى معلومات أخرى اعتماداً على توفرها.

مقطع التوالي: وتذكر في بدايته Origin وفيه تذكر التواليات وفي ملف تقرير بنك الجينات تذكر بحروف صغيرة وعلى شكل اسطر تحوي على 60 ثمانية ترقيم بداية السطر فقط وتفصل كل عشر ثمالات بفراغ وفي حالة كون التواليات خاصة بالحوامض النووية يذكر عدد كل من القواعد النتروجينية الموجودة في التوالي في نهاية المقطع. وينتهي التقرير بخطين مائلين //

ويمكن ان تختلف ملفات بنك الجينات بعض الشيء ولكن المذكور أعلاه يمثل أحد التقارير التي قد يزداد عليها او ينقص منها وفق المعلومات المتوفرة. وملفات هذه القاعدة تعد أكثر الملفات تفصيلا وتوفر واجهة NCBI إكسكسكس أخرى فضلا عن ملفات بنك الجينات منها إمكانية الحصول على التوالي بصيغة FASTA او صيغ ممثلة بالرسوم او غيرها.

وعند استرجاع تواليات DNA او البروتينات من القاعدة فان البحث يمكن ان يحدد بحقول مختلفة من التهميش مثل اسم الكائن الحي او رقم التسجيل وبعض الأحيان باستعمال اسم المؤلف او المقالة المنشورة او التاريخ، ويمكن ان استعمال إمكانية مثل المحددات والإظهار والفهرسة المذكورة او يمكن استعمال ما يسمى بالتصنيفات GenBank qualifiers التي تحدد حقل واحد في ملفات بنك الجينات وهي الموضحة في الجدول 3.2

جدول 3.2 : التصنيفات المستعملة في قاعدة بنك الجينات GenBank

الرمز	اسم الملف	التعريف
[ACCN]	رقم التسجيل	رقم التسجيل للتوالي او سجل النيوكليوتيد او البروتين او سجل الجينوم
[ALL]	كل الحقول	كل المفردات من حقول قاعدة البيانات
[AUTH]	اسم المؤلف	أسماء المؤلفين لكل المراجع في سجلات قاعدة البيانات
[ECNO]	رقم الإنزيمات	رقم الإنزيمات المحددة من قبل Enzyme commission او خدمات Chemical abstract
[FKEY]	مفتاح الخصائص	تمثل الخصائص الحيوية المعلنة لتواليات النيوكليوتيدات
[GENE]	اسم الجين	الاسم القياسي والاسم العام للجينات الموجودة في سجل قاعدة البيانات
[JOUR]	اسم المجلة	المجلة المرجع الذي نشرت فيها المعلومات عن الجزيئات
[KYWD]	الكلمات المفتاحية	تمثل فهرس للمفردات ذات العلاقة بينك الجينات و EMBL و DDBJ وقاعدة البروتينات السويسرية ، PIR وغيرها من قواعد البيانات
[MDAT]	تاريخ التحوير	تمثل آخر تحديث أدرج في Entrez بصيغة السنة والشهر واليوم YYYY/MM/DD
[MOLWT]	الوزن الجزيئي	الوزن الجزيئي للبروتينات بالدالتون بالطريقة المذكورة في Entrez
[PROP]	الصفات	ذكر للصفات لتواليات البروتينات مثل نوع الجزيئة ، موقع الجزيئة ، القسم التابعة له في بنك الجينات
[PROT]	اسم البروتين	الاسم القياسي للبروتين في سجلات القاعدة
[PDAT]	تاريخ النشر	تاريخ إطلاق السجل في Entrez بشكل YYYY/MM/D
[SQID]	تعريف التوالي	تعريف التوالي Sequence identifier
[SLEN]	طول التوالي	ذكر لطول كل توالي
[TITL]	عنوان الكلمات	يشمل الكلمات الموجودة في سطر التعريف من السجل
[ORGN]	الكائن	يشمل اسم الكائن العلمي والاسم الشائع الذي يحوي توالي الحوامض النووية او البروتين

وهذه المختصرات تكون مشتركة في بعض الأحيان مع قاعدة المراجع PubMed

Refseq: تمثل قاعدة بيانات للتواليات المدخلة الى بنك الجينات ولكن بعد إجراء التعديلات والتشذيب اليدوي ولذلك فان هذه التواليات تعطي أرقام تسجيل جديدة ثم تعاد الى البنك المذكور، والعكس ليس صحيح. رقم هذه القاعدة يتكون من حرفين مع شارحة للأسفل Underscore ثم تليها ستة أرقام مثل AB_123456. والحرفين المستعملة في الترقيم تشير الى نوع التوالي باستعمال شفرات خاصة ومنها:

NT تشير الى المتجاورات Contigs

NM لتواليات cDNA التي بنيت من mRNA

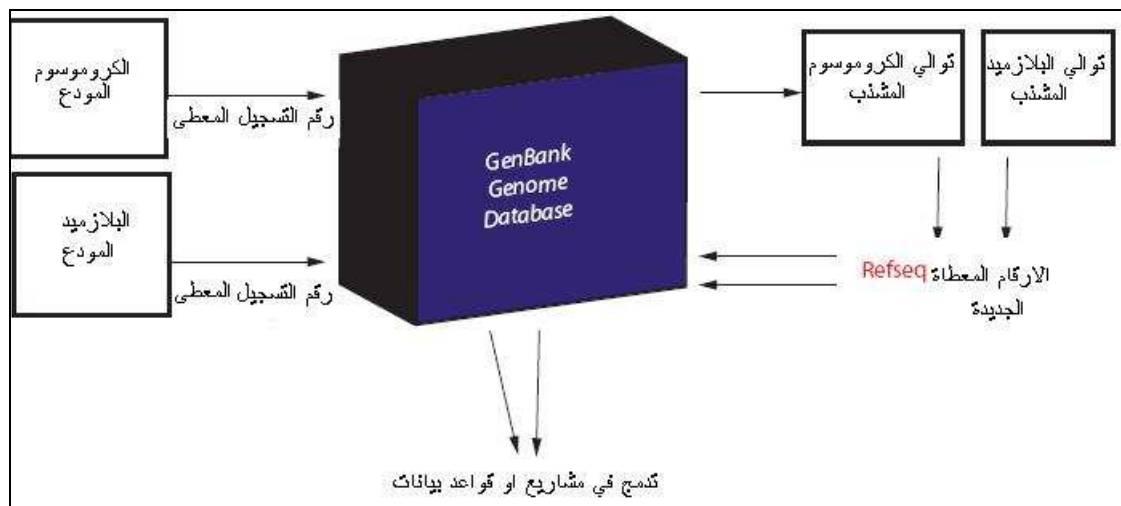
NP تواليات البروتينات

NC لتواليات الكروموسومات والبلازميدات

NZ لتواليات حددت بطريقة التنشيطية Shotgun غير المتكاملة، وتختلف هذه بعض الشيء في صيغها اذ ان لها أربعة حروف وثمانية أرقام مثل NZ_ABCD 12345678.

والعلاقة بين المدخلات الأولية وتصفياتها بـ Refseq موضحة في الشكل (15.2) وتلحق بقاعدة بنك الجينات قواعد خاصة بتركيب الجينات وخاصة لجينات الخلايا حقيقية النواة. ومنها Exon-intron database (EID) وتضم الآلاف من الجينات المشفرة للبروتينات، والبيانات في قاعدة بنك الجينات أنتج معظمها من تحليلات الحاسوب *In silico analysis* ، ولذلك أنشأت EID لتصحيح الوضع واعتمدت في الأغلب على الدراسات التجريبية، وتوفر القاعدة لكل جين معلومات من تواليات البروتينات فضلا عن توالي DNA نفسه وكذلك المعلومات المشتقة من ترويسة تقرير بنك الجينات وملفات القاعدة كلها بصيغة FASTA لتسهيل استعمالها في تطبيقات أخرى .

اما القاعدة الأخرى في هذا المجال فهي ExInt والتي تضم الاكسونات والانترونات للجينات المودعة في بنك الجينات ، والقاعدة هي قاعدة علاقية واجري عليها التنقيح ولكن بدرجة اقل من EID المذكورة أعلاه ولقاعدة ExInt خصائص أخرى تستعمل في حالات الضرورة فضلا عن احتوائها على تواليات للبروتينات التي تشفر لها الاكسونات الموجودة في القاعدة وتمثل بشكل خاص فمثلا عند وجود الانترونات في وسط شفرة أحد ثملات البروتين او قبلها فان تلك الثمالة تعطى حرف صغيرة اما باقي التقرير فيكون بالأحرف الكبيرة .



شكل 15.2 : تصفية وتشذيب التواليات لإنتاج Refseq

قواعد بيانات البروتينات

بما ان البروتينات تعد من الجزيئات الحيوية المهمة لذلك وضعت لها عدة قواعد للبيانات وبمختلف المستويات والتراكيب. ولكن ما يجب تذكره ان قواعد البيانات سواء للبروتينات او غيرها من الجزيئات لا تكون دائمة وذلك لانها عرضة للدمج مع بعضها لغرض توفير خدمات أفضل او حذفها عند وجود ما يغني عنها.

ومن قواعد البروتينات التي لا يمكن إغفالها هي Protein information resources (PIR) والتي تم دمجها مع قاعدة البروتينات السويسرية (كما ذكر في موقع آخر). وهي قاعدة أنشأت بداية الستينات من قبل مؤسسات تعنى بالنواحي الطبية وقد تطورت بشكل كبير وما يميز هذه القاعدة انها تقسم الى أربعة مستويات اعتمادا على نوعية المعلومات المتوفرة والتهميش على التوالي ومنها:

- المستوى الأول PIR I وفيه تكون المداخل مصنفة بشكل كامل ومهمشة بشكل جيد ولكن المعلومات قد تكون قليلة حتى في هذا المستوى الذي يعد أفضل المستويات.
- المستوى الثاني PIR II يحوي مداخل أولية لم يتم مراجعتها بشكل دقيق.
- المستوى الثالث PIR III المداخل غير مصدقة ولم تتم مراجعتها.
- المستوى الرابع PIR IV ومداخلها لا تزال افتراضية لعمليات الترجمة.

قاعدة البروتينات السويسرية قاعدة أنشأت عام 1986 (وقد مر ذكرها في موقع آخر) تتم إدامتها وتحديثها من قبل المعهد السويسري للمعلوماتية الحيوية والاتحاد الأوروبي، تزود بمعلومات عالية الجودة وتشمل وصف البروتين ووظيفته وتركيب الدومينات فيه والتحويلات الحاصلة بعد الترجمة والتغايرات وغيرها ، وتكون عادة قليلة التكرار ولها ارتباطات مع قواعد البروتينات الأخرى . والواجهة الخاصة بالقاعدة توفر العديد من البرامج الخاصة للتعامل مع البروتينات.

حول البروتينات التابعة لقاعدة البروتينات السويسرية وتشمل كل تواليات الحوامض النووية

الموجودة في قواعد المركز الأوروبي التي تم ترجمتها . والبروتينات التي لم يتم تهميشها او مراجعتها في هذه القاعدة هي البروتينات الافتراضية Hypothetical proteins.

وفضلا عن ذلك هناك العشرات من قواعد البيانات الخاصة بالبروتينات التي تختلف الأهداف من وضعها فهي قد تكون خاصة بالدومينات او تختص بإعطاء معلومات عامة او خاصة بعوائل البروتينات وغيرها لا يمكن ذكرها جميعا، ولكن من الموصى به انه عند البحث ودراسة بروتين ما يجب استعمال أكثر من قاعدة (وسينم ذكرها عند الحاجة) .

قواعد البيانات الخاصة بالمراجع

تتوفر العديد من قواعد البيانات الخاصة بالحصول على المراجع، البعض منها يكون مجاني بالكامل والآخر يقتصر أدائه على إعطاء خلاصات البحث. ولعل البداية تكون بـ

PubMed قاعدة بيانات للمراجع واسعة جدا تابعة للمركز NCBI توفر بعض الكتب والعديد من المقالات العلمية في مجالات مختلفة، ولكل مقال رقمه التعريفي في القاعدة PMID او PMCID الذي يكون بمثابة مفتاح أولي ويكون خاص بالبحث لوحده ويمكن استعمال رقم التعريف PMID في محركات البحث مثل Google او غيرها للوصول مباشرة الى المقالة، لذلك تضع العديد من المراجع الحديثة الرقم التعريفي في نهاية البحث في حالة كونها مدرجة ضمن هذه القاعدة .

ويمكن الوصول الى قاعدة بيانات المراجع اما بطريق مباشر عن طريق الصفحة الرئيسية للـ NCBI او من خلال نظام البحث Entrez. واغلب البحوث التي تغطيها القاعدة هي ذات العلاقة بالنواحي الطبية ولها ارتباط مع Medical subject heading (MeSH) وهو نظام يتكون من مفردات لغة مقيسة تستعمل في فهرسة المقالات وتساعد بالتزويد بمفردات أخرى تخص الموضوع نفسه.

ويستعمل في قاعدة بيانات المراجع ما يسمى بـ Word weight algorithm لتحديد المقالات ذات العلاقة بالكلمات المفتاحية المستعملة والتي لها علاقة ، او كلمات متشابهة في العنوان او الخلاصة او MeSH. وفي حالة البحث المعقد تستعمل العلاقات الجبرية او الجمع بين المحددات Limits وصفة الإظهار والفهرسة Preview/Index ، او يمكن استعمال الطمغات Tags لتحسين عملية البحث مثل [Au] تختصر البحث تحت اسم المؤلف او المؤلفين و [JID] وتختصر البحث تحت اسم المجلة ولذلك فان مفردات البحث يمكن ان تخفض باستعمال الطمغات التي تربط بالعلاقات الجبرية. وبعض هذه الطمغات او البطاقات المستعملة في قاعدة بيانات المراجع موضحة في الجدول 4.2 :

جدول 4.2 : الطمغات المستعملة في قاعدة بيانات المراجع PubMed

الوصف	الاسم	الطمغة
الخلاصة	Abstract	AB
المنتسب(هنا يشار الى المؤسسة التي يعمل بها الباحث الأول وعنوانه ورقم المنحة التي تم بموجبها اجراء البحث او الدراسة).	Affiliation	AD
معلومات عن الناشر والتعريف به او التعريف الالكتروني (digital object identifier) doi	Article identifier	AID
أسم المؤلف او المؤلفون	Author	AU
تاريخ نشر المقالة او البحث	Date of publication	DP

Journal ID	JID	الرقم التعريفي الخاص بكل مجلة في مصنف المكتبة الوطنية القومية للكتب، والمجلات وغيرها من وسائل الإيضاح
Language	LA	اللغة التي كتبت بها المقالة
Place of publication	PL	البلد الذي تصدر فيه المجلة
Publication type	PT	نوع المادة التي تتناولها المقالة
EC/RN number	RN	الأرقام الخاصة بالإنزيمات
Source	SO	حقل مركب يحوي على المعلومات الخاصة بالمراجع Bibliographic information
Journal Title Abbreviation	TA	المختصرات المحددة عالميا لأسماء المجلات
Title	TI	عنوان المقالة
Volume	VI	رقم المجلد للمجلة الحاوية على المقالة

وتتوفر سواء عبر قاعدة بيانات المراجع مباشرة أو Entrez أو روابط الى مراجع خارج نطاق مركز NCBI وأهمها Linkout.

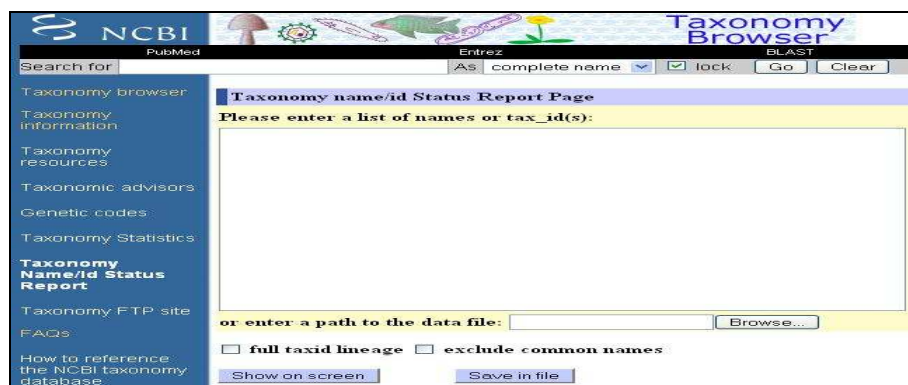
ومن المواقع الالكترونية الأخرى التي توفر العديد من المراجع هو موقع منظمة الصحة العالمية WHO المسمى HINARI والمراجع المتوفرة في هذا المركز معظمها مجانية ويمكن الوصول إليها بسهولة. وهناك موقع آخر هو موقع الجمعية الأمريكية للعاملين في مجال الأحياء المجهرية American society of microbiology (asm) الذي يوفر آلاف المراجع مجانية، فواجهة الموقع توفر المجلات الرئيسة الصادرة عن الجمعية كخطوة أولى والبعض من هذه المراجع يعود الى القرن التاسع عشر اي منذ صدور بعض المجلات، والخطوة التالية توفر آلاف المجلات التي ترتبط إليها القاعدة. وتوفر المقالات بصيغة pdf أو بصيغة HTML والأخيرة تكون أكثر مرونة إذ ان أي مرجع مستعمل في المقالة يمكن الرجوع إليه للحصول على المقالة الكاملة أو على الأقل المختصرات من خلال ارتباطها بالموقع MedLine ، وفضلا عن ذلك فإن الملفات الناتجة تشير في نهاية المقالة الى المراجع الحديثة التي استعملت فيها المقالة قيد البحث وهذه تكون ضمن المجلات التي لها علاقة مباشرة مع asm وبذلك يكون من السهولة الحصول على نصوص المقالات الحديثة الكاملة بصيغة HTML أو pdf.

وهناك العديد من المواقع الالكترونية المهمة بتوفير المراجع العلمية من مقالات أو كتب وكذلك توجد مواقع خاصة ببراءات الاختراع وقد تحتاج بعض هذه المواقع الى التسجيل أو تكون لقاء ثمن يدفع عن طريق الانترنت، والمواقع الأخيرة تعد قليلة مقارنة بالمواقع المجانية. ويحصل في بعض الأحيان ان تكون المراجع الحديثة جدا لقاء ثمن ولكنها تطلق بعد ستة أشهر أو سنة الى السياق المجاني.

قواعد التصنيف Taxonomy databases

يعد تصنيف الأحياء أحد الجوانب المهمة في دراسات علوم الحياة، لذلك أخذ هذا الجانب نصيبه من وضع قواعد البيانات. وأكثر المواقع اهتماما بعمليات التصنيف هو NCBI، إذ يحوي على قاعدة بيانات للتصنيف ، ويوفر شجرة للعلاقات التطورية بين الأحياء بالاعتماد على RNA للوحدات الصغيرة للرايبوزوم. ويتم البحث في مجال التصنيف اعتمادا على الاسم

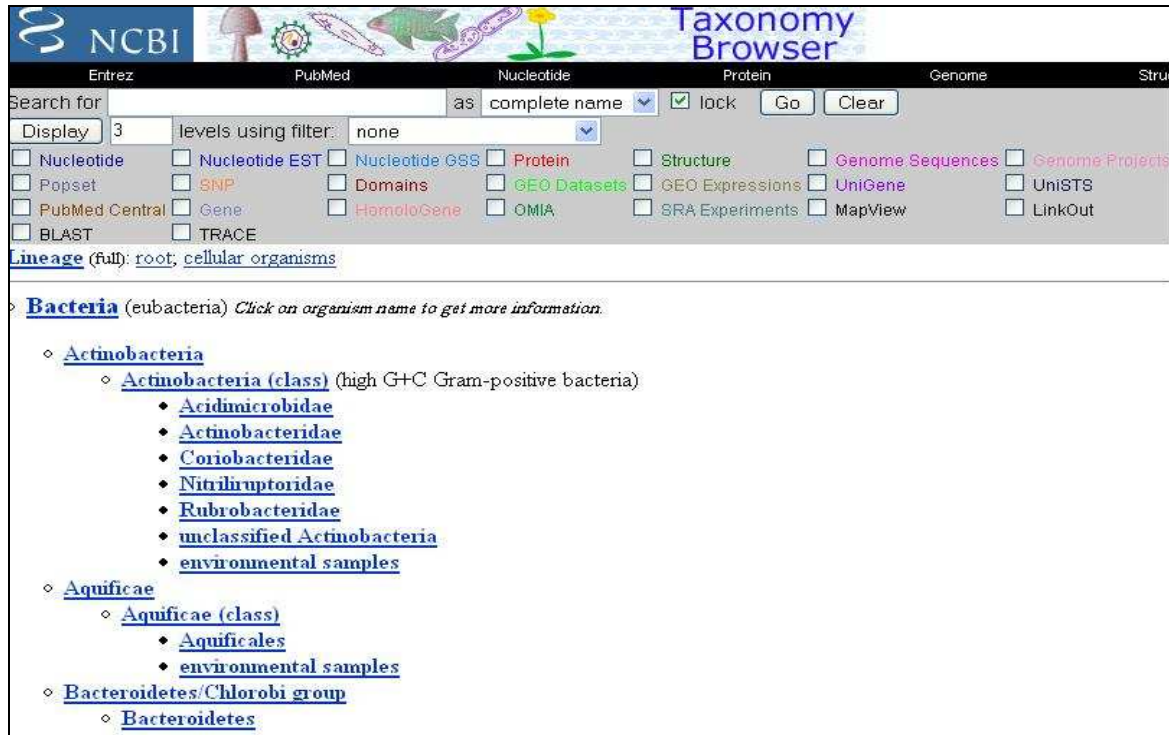
العلمي للكائن أو التعريف التصنيفي (Taxid (Taxonomy Identifier ويكون استعمال الأخير هو الأفضل نظرا لإعادة تسمية الأحياء ، فاستعمال Taxid يمكن ان يوصل الى الأسماء الحديثة مع ذكر التسميات القديمة للأحياء وعملية البحث موضحة في الشكل 16.2 :



شكل 16.2 : واجهة الموقع الخاص بالتصنيف في NCBI

يمكن الوصول الى المعلومات عن اي كائن مثل البكتيريا او الاركيا من استعمال الروابط الجانبية، فلبكتيريا تظهر قوائم التصنيف وكذلك الحال مع الاركيا ويمكن باختيار الكائن المطلوب وكذلك اختيار المعلومات المراد معرفتها من استعمال متصفح التصنيف Taxonomy browser كما موضح في الشكل (17.2) .

والبحث في التصنيف ليس هو الطريقة الأفضل لوضع الكائن مع المجاميع من الأحياء التي تتشارك في موقع بيئي معين مثل موقع بيئي متطرف الحرارة او متطرف في عامل بيئي آخر ولذلك يستعمل Geographical information system (GIS) خاصة في عالم الأحياء المجهرية لتحديد خارطة وجود الأحياء المجهرية في بيئة ما مثل البحار لذلك يمثل GIS أحد مظاهر عجز قواعد البيانات عن كشف العلاقات حيث لا توجد روابط بين GIS وقواعد البيانات، كما ان قواعد البيانات تعجز ايضا في الكشف عن أمراض الأحياء التي تم تحديد توالياتها وذلك لتداخل البيئة في إحداث الأمراض وان الأمر ليس معتمدا على التواليات بشكل مطلق. وفي هذا المجال يذكر ان أول توالي حدد للبكتيريا للـ *Haemophilus influenzae* لسلاسل حرة المعيشة غير ممرضة لذلك كانت الحاجة ملحة لإضافة بيانات إضافية مثل الجغرافية ونوع البيئة التي يقطنها الكائن الذي يتم تحديد تواليه.



شكل 17.2 : جزء من واجهة تصفح التصنيف في NCBI

قواعد بيانات الایض والإنزيمات

من أهم قواعد البيانات المميزة التي تتضمن تحت قواعد البيانات الحيوية هي قواعد البيانات المتعلقة بالایض اذ تمثل التفاعلات الحيوية الحقيقية التي تؤدي بواسطة الإنزيمات. ولهذا الغرض وجدت قواعد بيانات جديدة بالاهتمام منها:

- قواعد بيانات الإنزيمات منها القاعدة الألمانية **BRAunschweig ENzyme Database (BRENDA)** وقاعدة **ENZYME**. وكل من القاعدتين توفر إمكانيات البحث عن المعلومات الخاصة بالإنزيمات انطلاقاً من اسم الإنزيم أو رقم EC الخاص به وربما كانت معظم الروابط موضحة في الصفحة الرئيسة لقاعدة BRENDA المبينة في الشكل 18.2:

Nomenclature	Reaction & Specificity	Functional Parameters
Enzyme Names EC Number Common/ Recommended Name Systematic Name Synonyms CAS Registry Number	Pathway Catalysed Reaction Reaction Type Natural Substrates and Products Substrates and Products Substrates Natural Substrate Products Natural Product Inhibitors Cofactors Metals/Ions Activating Compounds Ligands Ligand Views	Km Value kcat/Km Value NEW Ki Value IC50 Value pI Value Turnover Number Specific Activity pH Optimum pH Range Temperature Optimum Temperature Range
Isolation & Preparation Purification Cloned Expression NEW		Organism-related information Organism Source Tissue

شكل 18.2 : واجهة قاعدة BRENDA الخاصة بالإنزيمات

• قواعد بيانات المسارات الأيضية Metabolic pathway databases

قواعد البيانات في هذا المجال تختلف من حيث الأسس التي استعملت في بناءها، أو الأحياء المنضوية تحتها.

* قواعد البيانات الخاصة بالنباتات: واغلب المسارات الأيضية في النباتات توجد ضمن شبكة خاصة بأبيض النباتات هي Plant metabolic network (PMN) وهذه الشبكة مغطاة بدعم رسمي حكومي تنفذ من قبل معهد كارنكي للعلوم Carnegie institution of science. والقواعد فيها مشتركة بين الكيمياء الحيوية وعلوم النبات. ومركز الشبكة هي قاعدة البيانات الخاصة بالنباتات PlantCyc وتضم العديد من المسارات الأيضية والمسارات الأيضية الثانوية المشتقة من المسارات الأساسية فمثلا لدورة الأحماض ثلاثية الكربوكسيل TCA يوجد أكثر من مسار مثل TCAII و TCAIII. وتتمتع الشبكة بنوعية جيدة إذ أن المعلومات فيها تم تنقيحها وتشذيبها يدويا والأساس في استقاء المعلومات هي المراجع العلمية بالدرجة الأولى وبدرجة ثانية تحليلات الحاسوب حول الجينات والإنزيمات والمركبات والتفاعلات التي تحصل في النباتات وكذلك مسارات الأيض الأولي والثانوي وهذه تكون لأكثر من 300 نبات. وتتمتع الشبكة بمصداقية عالية وتظهر فيها النتائج حاوية على إيقونة دورق مختبري أو صورة حاسوب (شفرة المصادقية Evidence code) الذي يمكن أن يوصل إلى التجارب التي اشتقت منها المعلومات. فضلا عن أن القاعدة توفر عددا من الوسائل والبرامج التي تسهل إجراء الدراسات والتحليل اللازمة سواء الدراسات التي تجري باستعمال الحاسوب أو التحاليل الإحصائية وكذلك يمكن أن تظهر نتائج مصورة باستعمال Omics view. وتضم الشبكة قواعد بيانات لأبيض

بعض النباتات المهمة مثل AraCyc الخاصة برشاد الصخر *Arabidopsis thaliana* الذي يمثل النموذج الدراسي للنباتات وبعض المسارات الايضية تكون خاصة بهذا النبات والأخرى تكون مشتركة مع نباتات أخرى وتوفر القاعدة وسائل لإظهار خرائط لمسارات الايض. فضلا عن توفر قواعد بيانات لمحاصيل نباتية وخضر كثيرة.

* قواعد البيانات الخاصة بالكائنات المختلفة، وهي تضم مسارات ايضية لمختلف الأحياء سواء كانت حقيقية النواة او بدائية النواة ومنها:

* قاعدة بيانات كيوتو Kyoto Encyclopedia of Genes and Genomes (KEGG) والإمكانات التي توفرها القاعدة ربما تختصر في الشكل التالي (شكل 19.2) الذي يوضح واجهة القاعدة.

ولعل مثالا يعطى في هذه الحالة حول الروابط والإمكانات هو الكم الهائل من المعلومات حول عملية ايضية محددة مثل الفسفرة التأكسدية الموضحة في الشكل (20.2).

KEGG PATHWAY Database
Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG GENES GENOME LIGAND DBGET

Select prefix: map Organism Enter keywords: Go Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps, change history, and last updates) representing our knowledge on the molecular interaction and reaction networks for:

- 0. Global Map
- 1. Metabolism
 - Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
 - Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
- 2. Genetic Information Processing
- 3. Environmental Information Processing
- 4. Cellular Processes
- 5. Organismal Systems
- 6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

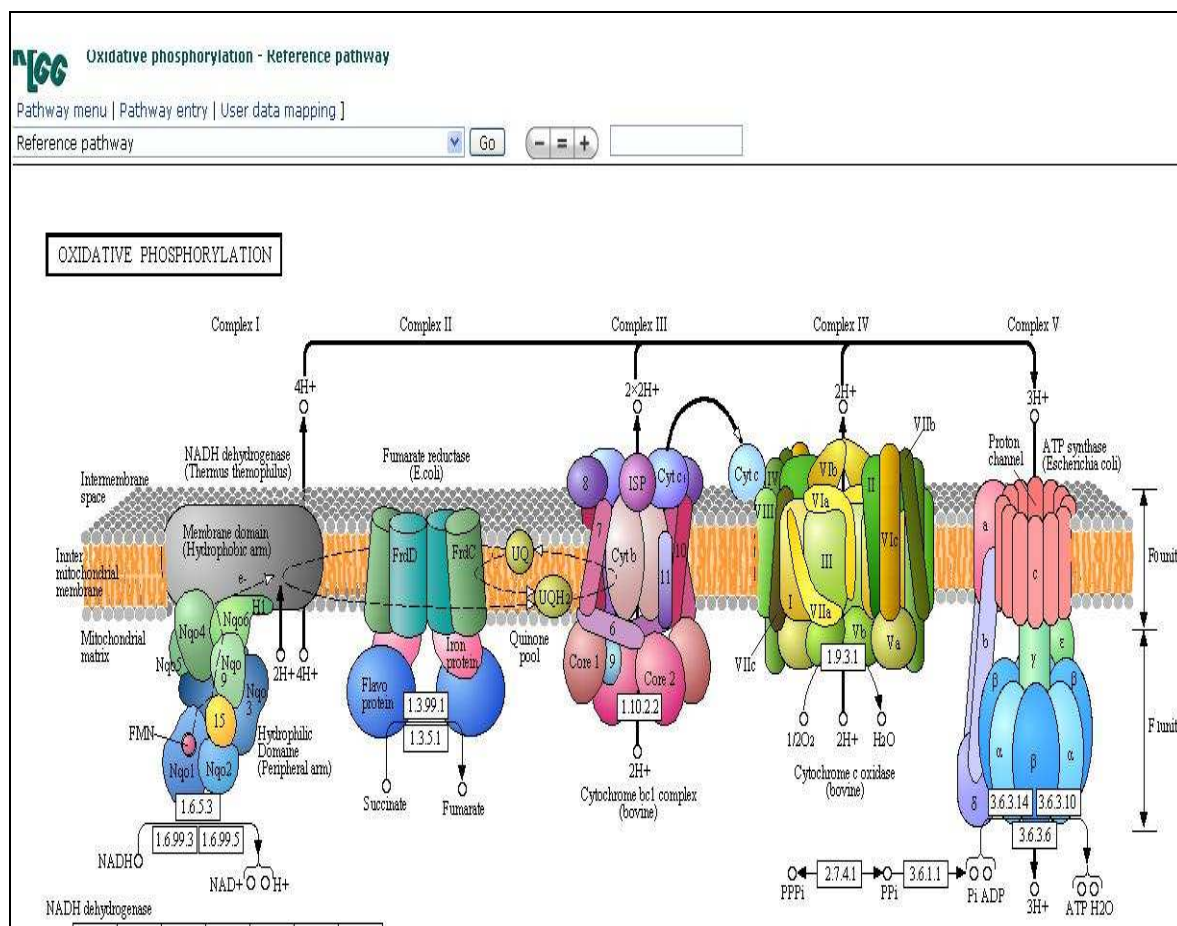
- 7. Drug Development

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

شكل 19.2 : واجهة قاعدة بيانات كيوتو للمسارات الايضية KEGG

والقاعدة فضلا عن ما ذكر أعلاه توفر أطالس للمسارات الايضية التي تضمها، وكذلك إمكانيات استعمال بياناتها لتطوير الأدوية. ولكن من المآخذ على هذه القاعدة هي ان اغلب بياناتها تم جمعها وتطويرها ببرامج حاسوب وليس بتجارب مختبرية فعلية، وتذكر القليل من المعلومات مثلا للإنزيمات يذكر فقط العوامل المرافقة ، لذلك توجد صعوبة في التحقق من صحة ودقة المعلومات فيها، خاصة وانه لا يوجد ذكر للمراجع العلمية فيها. ولا يذكر فيها النوع الذي تجري فيه المسارات الايضية ولا تستعمل نظام شفرة الدليل (المذكور سلفا). ولازالت القاعدة تحوي على أخطاء خاصة بالإنزيمات ناتجا عن عدم استعمال EC numbers الخاصة بالإنزيمات. وكان الإصدار الأخير لعام 2009 وهو الإصدار 50 يحوي مسارات ايضية اقل من بعض القواعد الأخرى المتخصصة في المجال نفسه، وهي في العادة غير متاحة للاستعمال الحر من قبل المستخدم أي انها تكون مقابل ثمن .



شكل 20.2 : الفسفرة التأكسدية كما موضحة في قاعدة البيانات كيو تو KEGG

● **قاعدة BioCyc:** أحد القواعد الكبيرة المهمة بالمسارات الايضية وتكون صيغ بياناتها منظمة وموحدة وبرامجها متطورة جدا اذ توجد مرشحات وبرامج تنبؤ وطرق للرسم وتضم قواعد

بيانات لمجاميع من الأحياء المهمة . وتوفر القاعدة إمكانيات للبحث عن الجينومات وكذلك مشاركة اي جين في المسار الأيضي الذي ينتمي اليه وغيرها من الوسائل لتحليل البيانات الموجودة فيها ضمن حقل Omics data ، اذ يمكن ان تظهر خرائط لعمليات الايض وخرائط للتنظيم فضلا عن توفر خرائط للجينومات ولذلك تعد هذه القاعدة من القواعد المتطورة على كافة الأصعدة سواء البيانات او البرامج المستخدمة لاستغلال المعلومات، والقواعد الفرعية التي تضمها تقسم الى ثلاثة درجات وهي:

1. الدرجة الأولى وتعد الأفضل لان المعلومات فيها نقحت وشذبت وعدلت يدويا ومعلوماتها مستقاة من المراجع العلمية.
2. الدرجة الثانية وتحوي مسارات ايضية تم بناؤها اعتمادا على التنبؤ الذي يجري بواسطة الحاسوب ولكنها عرضت لعمليات تنقيح وتشذيب معتدلة.
3. الدرجة الثالثة وهي قواعد أسست على دراسات مشتقة من تحليل الحاسوب ولم تتعرض الى اي تعديل او تنقيح وتضم مئات من قواعد البيانات الفرعية لذلك تعد مجالا للناشطين للإدلاء بدلوهم لتحسين أوضاعها.

ومن قواعد الدرجة الأولى **EcoCyc** التي تستقي معلوماتها من المراجع العلمية الخاصة بالبكتيريا *E. coli* K12 ، وتضم القاعدة شبكة لتنظيم الانتساخ وكذلك معقدات البروتينات فضلا عن الإنزيمات وبروتينات النقل مع ذكر وظائفها، وقد تعرضت القاعدة الى عمليات تنقيح وتشذيب كبيرة للإبقاء على الصحيح من البيانات . وتضم الدرجة الأولى ايضا القاعدة الخاصة بخميرة الخبز *Saccharomyces cerevisiae* وهي **YeastCyc** وتحوي على المئات من المسارات الايضية التي استقيت معلوماتها من المراجع العلمية ونقحت وشذبت من قبل مختصين.

وتحت القاعدة الرئيسة **BioCyc** تتضوي القاعدة المهمة جدا وهي **MetaCyc** وهي قاعدة تشمل مختلف الكائنات ولا تختص بكائن واحد ، الميزة المهمة لهذه القاعدة ان معلوماتها معتمدة على التجارب والبحوث المنشورة في قاعدة بيانات المراجع اي انها من القواعد ذات النوعية الجيدة وفيها مسارات وبرامج قابلة للتحميل على الحاسبات الشخصية. والميزة الأخرى المهمة توفيرها إمكانيات إجراء تحليلات المقارنة بين المسارات الايضية. ولها أربع إصدارات سنويا اثنين في المجال الضيق واثنين في المجال الواسع. والمعلومات المدرجة فيها تتعرض للتنقيح والتعديل من قبل أشخاص مختصين بعد استقائها من البحوث العملية ومقالات المراجعة وكذلك من القواعد الخارجية الأخرى مثل **EcoCyc** ، **AraCyc** ، **YeastCyc** و **ScoCyc** (الخاصة بـ *Streptomyces coelicolor*). اما التراكيب الكيماوية فيها فهي من **PubChem** والمراجع الأصلية، والتفاعلات الإنزيمية تستقى من قاعدة بيانات الإنزيمات **ENZYME**. وتتعامل القاعدة مع التغيرات في بعض المسارات الايضية على انها مسارات ايضية مميزة وتعطى تحت مداخل خاصة بها فمثلا دورة الأحماض الثلاثية الكربوكسيل يوجد لها في القاعدة إصدارات خاصة بها مثل **TCAII** ، **TCAIII** وكذلك مسارات ايض الحوامض الامينية وغيرها. وتظهر القاعدة مسارات الايض الخاصة بكائن واحد او المسارات التي تشترك فيها مئات من الأحياء ولذلك يمكن الاستعلام عن المسار الأيضي باستعمال النوع **Species**.

ترتبط القاعدة مع قواعد البيانات الأخرى ، وارتباطها قد يكون تقليديا كما في ارتباط القواعد الأخرى فيما بينها ولكن الأهم ان القاعدة توفر نوع من الروابط المتخصصة مثل رابط لتفاعل معين لمسار ايضي فيها يربط الى قاعدة بروتينات مثل PIR الذي يشارك البروتين فيها بالمسار الأيضي. وللتفاعلات الإنزيمية فيها طرق توضيح أكثر مما هو الحال مع قواعد البيانات الأخرى ، اذ ان كل تفاعل إنزيمي له طريقة خاصة وتعليقات وتوضيحات، فيعطي وصف للإنزيم من حيث عدد الوحدات الفرعية وتخصص مواد الأساس والحاجة الى العوامل المرافقة الإنزيمية وكذلك للمنشطات والمثبطات، فضلا عن توفير إمكانيات للمستخدم لإظهار التداخلات بين مختلف المسارات الأيضية، وكان آخر إصدار لها 13.5 عام 2009.

ومن القواعد المتخصصة في مجالات الايض القاعدة **Reactome** وتعد قاعدة بيانات منقحة ومشذبة خاصة بالفاعليات الحيوية للإنسان وتغطي معظم المسارات الأيضية الممتدة من العمليات الأساسية في الايض الى مستويات عالية منه مثل نقل الإشارات الهرمونية. معلوماتها مستقاة من المراجع العلمية وتذكر فيها التعليقات التوضيحية فضلا عن ذكر المراجع العلمية بشكل وافي . والمسارات الأيضية فيها أقل مما موجود في القواعد المناظرة مثل BioCyc وليس فيها وسائل تهتم بالتصنيف لأنها تهتم بكائن واحد وهو الإنسان.

وهناك قواعد بيانات تخص عمليات الايض الخاصة مثل قاعدة UM-BBD التابعة لجامعة Minnesota الأمريكية ، تحوي المسارات الأيضية الخاصة بالأحياء التي تقوم بتحليل المواد العصبية والداخلية Xenobiotics والمركبات الكيماوية الأخرى ، المعلومات فيها مشتقة من المراجع العلمية والتجارب وتحوي على تعليقات مهمة وموضحة وذكر كامل للمراجع المستعملة في كل حالة، والغاية منها توفير المعلومات اللازمة للتعامل مع البيئة وعمليات التقنيات الحيوية.

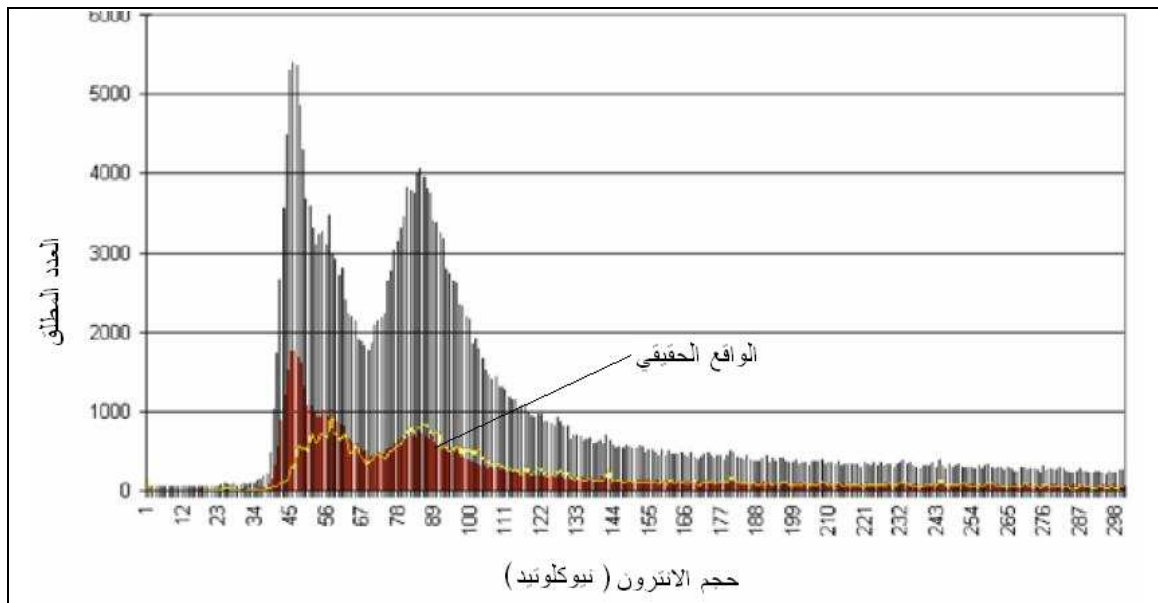
ومن القواعد التي لا بد من ذكرها في مجال المسارات الأيضية قاعدة aMAZE وتضم المسارات الأيضية الخاصة بالبكتريا *E. coli* وخميرة الخبز، معلوماتها من المراجع العلمية ولكنها قليلة ولا تحوي على تعليقات للإنزيمات والمسارات الأيضية التي تشارك فيها، فلا ذكر لمنشطات او مثبطات الإنزيمات او العوامل المرافقة للإنزيمات او الوحدات الفرعية المكونة للإنزيمات. اي ان البيانات فيها قليلة ولا تماشي متطلبات البحوث الحديثة ولذلك يلاحظ ان هذه القاعدة قد تقاعدت عن الخدمة وإنما ذكرت هنا لتوضيح ان قواعد البيانات هي ليست دائمة وإنما تكون ديمومتها بمدى البيانات والخدمات التي تقدمها.

مثالب ومصادقية قواعد البيانات

القاعدة العامة ان المعلومات المستقاة من التجارب العملية هي الأكثر وثوقا اذ ان هذه المعلومات تكون تحت إشراف بايولوجيين متخصصين وبذلك تكون أكثر مصداقية وقبولا من البيانات التي يتم التنبؤ بها بالحاسوب او الناتجة من برامج المحاكاة بالحاسوب وبذا لا يمكن ان تصل الى الوضع الذي تجري به تفاعلات الأنظمة الحيوية.

ومن جهة ثانية تكون قواعد البيانات عرضة للأخطاء مثل الأخطاء الطباعة وغيرها مهما كانت صغيرة فإنها عرضة للتفاقم في الخطوات اللاحقة عند إجراء عمليات المعالجة والإعداد للقواعد او عند استعمال القاعدة التي ستؤدي في النهاية الى استنتاجات خاطئة اذ تؤثر في الخرائط التي يمكن ان تنتج وكذلك أخطاء في عمليات انزلاق الانترنت.

فلو أخذت قاعدة البيانات بنك الجينات فهذه القاعدة تحوي على كميات كبيرة من المعلومات المهمة والمتكررة فهي قاعدة غير مشذبة Redundant database تصل الزيادات فيها الى أكثر من 80% فيما يخص جينات الخلايا حقيقية النواة الحاوية على الانترونات والاكسونات. وتصل الى 17% في الحالات العامة وقد تكون في هذه الحالة ناتجة من تحديد التواليات بشكل مستقل فباستعمال المرشحات المختلفة عام 2000 كان إصدار القاعدة يحوي على 52,022 من الجينات الحاوية على الانترونات وبعد عمليات التصفية انخفض الى 22,345. ومن جهة ثانية يلاحظ ان البرنامج المستعمل يؤثر في عدد الجينات فمثلا عند استعمال GeneFinder أشارت النتائج الى ان ثلث عدد الجينات الحاوية على الانترونات تعود الى الدودة *C. elegans* (17,138 جين) يليها رشاد الصخر 9041 جين ولكن باستعمال مرشح الكلمات المفتاحية انخفض العدد بالنسبة للدودة الى 316 والرشد الى 940 وذلك لان البرنامج يعتمد على تواليات DNA أي يأخذ بنظر الاعتبار وجود الانترونات والمكررات خارج منطقة الجين وعند الأخذ بنظر الاعتبار mRNA انخفض العدد الى 453 بالنسبة للدودة والى 860 بالنسبة للرشد. وفي الإنسان تكون الأعداد أكثر ثبوتاً في الحالة الأولى أي ان استعمال الكلمات المفتاحية يخفض الأعداد الى مستوى معين ، اما باستعمال mRNA فتتخفض أكثر وهذا يعود الى كثرة تواليات mRNA الخاصة بالإنسان المودعة في القاعدة . وتكون عمليات الترشيح على أكثر من مستوى، اذ تتم عملية الترشيح الأولى ثم تستعمل نتائجها لعملية ترشيح ثانية وهكذا ، وعليه أنشأت داخل هذه القاعدة قواعد ضمنية خاصة بالاكسونات والانترونات مثل EID database و ExInt database والتي توضح عدد الجينات المعنية وعدد اكسوناتها والانترونات التي فيها (التي مر ذكرها سلفاً) والشكل التالي (شكل 21.2) يوضح الفرق في عدد وحجم الاكسونات والانترونات قبل وبعد التنقيح.



شكل 22.2 : العلاقة بين العدد المطلق من الانترونات وحجم الانترون في قاعدة بنك الجينات

وتكون الأخطاء الناتجة من عمليات النسخ وكذلك عدم تحديد الحدود الفاصلة بين الانترونات والاكسونات 'GT...AG' بشكل صحيح. وفي عمليات التصفية يمكن للمستخدم ان يزيل كل الجينات التي انتروناتها لا تحوي على هذه القطيقات Motifs التركيبية ('GT...AG') ، ولكن هذه تكون ملائمة فقط للانترونات القابلة للإزالة ولكن بعض الانترونات لا تتعرض لعمليات الإزالة كما في المجموعة الأولى والثانية منها (الأنواع التي سيمر ذكرها لاحقاً) ، كما ان بعض الانترونات التي تكون قطيقات الانفلاق مكونة من "AT...AC" وهي شاذة عن القاعدة العامة وهذه غير شائعة وموجودة بمستوى حوالي 1% في رشاد الصخر .

اما الأخطاء الأخرى فهي موجودة في الجينات التي تم التنبؤ بها بالحاسوب فالعديد من الجينات في بنك الجينات لم تثبت عملياً ولذلك قد لا تكون حقيقية ومثل هذه يكون استعمالها خطر عند استعمالها لتطوير برامج للتعرف على مناطق الاكسونات /الانترونات، وهذه المشكلة يمكن ان تعالج الى حد ما بواسطة عمليات الترشيح للحصول على عدد أقل من الجينات التي تكون محددة بطرق التجربة، وكذلك بمقارنة جميع تواليات الجينات الموجودة في قاعدة البيانات مع تواليات mRNA ولذلك فان التواليات التي تتواءم مع تواليات mRNA تشير الى ان الجين صحيح وبالنتيجة ان 22% من الجينات بقيت في القاعدة وهذه النتيجة قد تكون غير آتية من الأخطاء وانما لقلة تواليات mRNA الموجودة في القاعدة.

المستويات الأخرى من الترشيح والتتقيق باستعمال الكلمات المفتاحية وهي الطريقة الأفضل اذ أقيمت على 44% من الجينات في القاعدة، وهذا المستوى ايضا غير مضبوط بشكل مطلق لان الكلمات المفتاحية تعتمد على ترويسة قاعدة بنك الجينات التي تكون عامة، ولذلك يكون أمام المستخدم اما استخدام قواعد بيانات كبيرة وأخطاء بمستويات معينة او باستعمال قواعد بيانات فرعية مضبوطة ومحددة بالتجارب. وقاعدة البيانات EID بعد التتقيق والتعديل فهي تحدد حجم الانترونات ونوعيتها هل هي من نوع Phase 0 او 1 او 2 وكذلك طول الاكسونات فضلا عن تحديد الطول الكلي للانترونات واذا لم يعرف طول اي من الانترونات فتأخذ هذه الفقرة القيمة (-1) أي ان الطول الكلي للانترونات غير معروف، وفي تقرير توالياتها توضع الاكسونات بحروف كبيرة والانترونات بحروف صغيرة وإذا كان توالي الانترون غير معروف توضع (...).

وبناء على ما ذكر أعلاه ظهرت فكرة شفرة المصادقية او شفرة الدليل التي تشير الى الدليل المعتمد في بناء القاعدة ومنها شفرات المصادقية للإنزيمات المشاركة وكذلك شفرات المصادقية للتجارب وهذه تظهر في مخرجات البحث في بعض القواعد وتكون على شكل إيقونة ورق مختبري Flask وصورة حاسوب عند استعماله وتظهر الشفرات المصادقية التجارب المعتمدة عند استخدامها . ويمكن ان تمثل بأي شكل آخر يحدد مصادقية المعلومات الموجودة في القاعدة على الأساس العملي والتجريبي وليس على أساس تحليلات الحاسوب .

الإبقاء على حداثة قواعد البيانات

يحتاج الباحث في كثير من الأحيان الحصول على أحدث المعلومات حول التواليات الجديدة المودعة او حول البيانات الإضافية التي تضاف الى ما موجود من التواليات ولهذه المهمة توجد بعض المواقع التي توفر مثل هذه الخدمات منها GOLD ويتعامل الموقع مع مستجدات تحديد التواليات ، المنحى الثاني يكون من استعمال الجداول المتوفرة حول جينومات الأحياء سواء

بدائية أو حقيقية النواة المخزونة في قواعدها مثل بنك الجينات التابع للـ (NCBI) وهذه الجداول تشير الى تاريخ أول إعلان عن التوالي ثم تذكر فيها بعض المعلومات الأخرى ومنها التحديث والتحويل الذي قد يكون على شكل إصدارات مختلفة للتوالي. أو استعمال صفحات المركز الأوروبي أو الرجوع الى صفحات الانترنت التي تعطي نظرة عامة عن المستجدات في هذا المجال ومنها التواريخ المتعلقة بها.

إرشادات للبحث عن المعلومات

ان الخوض في المعلوماتية الحيوية وقواعد بياناتها وبرامجها لا يكون مجدياً اذا كانت البدايات خاطئة، اذ تؤدي الى عدم وضوح الرؤيا وصرف الوقت والجهد وربما الحصول على نتائج خاطئة . لذلك كان على الباحث ان يضع خطة واضحة وقد لا تكون دقيقة التفاصيل ولكنها يمكن ان تكون مثمرة وفيما يلي بعض الإرشادات المساعدة:

- تحديد الهدف المطلوب من عملية البحث والدراسة ويعد هذا أهم أسباب النجاح في الوصول الى الهدف بأقصر الطرق الصحيحة وقد يظهر أثناء عملية الوصول توفر معلومات إضافية يمكن ان تكون مساعدة ولذلك كما ذكر أعلاه ان خطة البحث قد لا تكون دقيقة التفاصيل.
- بناء على الهدف المحدد ونوع المادة العلمية المراد الاستعلام عنها يتم اختيار قواعد البيانات التي يتم مسحها وكذلك البرامج الملائمة لاستخدامها، فمثلاً أفضل مكان للذهاب اليه للتعرف ودراسة tRNA هو tRNA databases وهذه القواعد يمكن ان تصنف هذه الجزيئات الخاصة وفقاً للحمض الأميني الذي تحمله، وتحتوي القواعد على إمكانية الحصول على التراكيب الثانوية لجزيئات tRNA وكذلك تستعمل في تحليل الاصطفافات Alignment analysis، وفضلاً عن ذلك توفر البيانات الوافية عن tmRNA كما في قاعدة عوائل جزيئات Rfam database RNA . ولذلك فليس من المعقول استعمال برنامج لإيجاد التراكيب الثانوية لبروتين ما وادخال تواليات الحوامض النووية حيث سيرفضها البرنامج ولذا نحتاج الى خطوة إضافية في هذه الحالة وهو ترجمة تواليات الحوامض النووية الى بروتينات مع أخذ الحيلة والحذر من تفضيل استعمال الشفرات لكل كائن.
- للبحث عن توالي معين في قواعد البيانات يفضل اختيار أكثر من قاعدة بيانات والركون الى النتائج التي تعطي أكثر تطابقاً وأكثر عقلانية، لان بعض قواعد البيانات تكون غير مرتبة وتحتوي تكرار وفائض من البيانات يمكن ان تؤثر في مخرجات النتائج. وفي العموم لابد من الاطلاع على بعض المعلومات عن قواعد البيانات قبل استخدامها لمعرفة مصادر بياناتها وطرق تركيبها ونوعيتها لاستخدام الوسيلة الملائمة لاسترداد البيانات منها. كما انه يجب الانتباه الى حداثة المعلومات المدرجة في قواعد البيانات من كل النواحي مثل التواليات المودعة والتهميشات عليها لان ذلك سيؤثر في مصداقية النتائج التي يتم الحصول عليها.
- اختيار البرامج الملائمة لأداء المهامات وفي هذا المجال يفضل استعمال أكثر من برنامج للتنفيذ. وقبل اختيار البرامج يجب الاطلاع على بعض الجوانب منها:
 1. اللغة التي كتب بها البرنامج ويفضل استعمال البرنامج نفسه مكتوب بأكثر من لغة.
 2. الاطلاع على الأسس التي كتب بها البرنامج والأخذ بنظر الاعتبار النواحي البيولوجية اذ ان بعض البرامج تعتمد على صفات ليس لها وزن ثقيل من النواحي الحيوية وبالتالي يكون البرنامج ضعيفاً ، وهذا ما يلاحظ من ظهور بعض البرامج ثم اختفاؤها نظراً لعدم إمكانياتها

في التوصيل الى نتائج مرضية من النواحي البايولوجية ، فضلا عن ملاحظة صدور إصدارات جديدة ومتكررة للبرنامج نفسه ، ويكون الإصدار الجديد قد عالج نقطة ضعف ما في البرنامج القديم.

3. اختبار القيم الإحصائية للبرنامج لأنها ستكون هي القيمة الترجيحية للنتائج التي يتم الحصول عليها، وملاحظة الطرق الإحصائية التي بني عليها البرنامج.

4. الانتباه الى طريقة تنفيذ البرنامج والمصفوفات الترجيحية التي يعتمد عليها فبعض البرامج قد تكون سريعة التنفيذ قليلة الدقة، اذ ان هذين المؤشرين لا يجتمعان في العادة، والسرعة العالية تكون على حساب الدقة.

5. يجب دراسة البرنامج، كي يتم التمكن من تفسير النتائج الظاهرة في مخرجاته وتفضل في هذه الحالة البرامج التي تعطي الأشكال والصور لأنها توضح النتائج بشكل أفضل، فضلا عن وجوب ان تكون المخرجات بملفات نصية لإجراء التطابق فيما بينها والتمكن من استخدام المؤشرات الإحصائية التي عادة ما تكون مدرجة في واجهة البرامج.

الفصل الثالث

الأنظمة الحيوية وفعاليتها

أقسام الأحياء الرئيسية
Biomolecules جزيئات الحياة
الجينات وأنواعها
أعداد الجينات
الطفرات
جينومات بدائية النواة
أنواع الجينومات
البلازميدات
توزيع وضرورة الجينات في الجينومات البكتريا
تضاعف جينومات بدائية النواة
جينومات الخلايا حقيقية النواة
العناصر القافزة Transposable elements
جينوم الإنسان
تضاعف المواد الوراثية
التعبير الجيني
RNA
الخيطة Splicing
البروتينات
عوامل الانتساخ Transcription factors
التنظيم Regulation
جزيئات RNA غير المشفرة
الوراثة اللاجينية Epigenetics
الجوانب العملية لدراسة الأنظمة الحيوية
الخرائط الوراثية

الأنظمة الحيوية وفعاليتها

أهم ميزات الأنظمة الحيوية انها أنظمة مختزلة جدا، وان العناصر التي تدخلها تكون — ربما — على أقصى درجات الاختزال ، فمثلا النتروجين في الأنظمة الحيوية يكون بتكافؤ (-3)، ولذلك يتحول من النتروجين الجوي ذو التكافؤ صفر الى التكافؤ (-3) الموجود في جذر الامونيوم او الامونيا بعد إضافة 3 من الالكترونات وكذلك الحال مع ذرة الكربون التي تتراوح نقطتي التكافؤ بين (4 -) و (4+) لذلك تكون ذرة إستراتيجية لعمل الأواصر التساهمية في المركبات التي تكون من صلب الأنظمة الحيوية. فذرة الكربون في أقصى درجات التأكسد (4+) عندما تكون في مركب مثل ثنائي اوكسيد كربون CO_2 ، وتتحول الى أقصى حالات الاختزال (4-) عندما تضاف لها ثمان من الالكترونات عند تثبيت ثنائي اوكسيد كربون سواء بعمليات التخليق الضوئي الباعث للأوكسجين الذي يحصل في النباتات والطحالب الخضر او غيره من التخليق الذي لا يكون معتمدا على الضوء كما في حالة الأحياء ذاتية التغذية Autotrophs. وكذا الحال مع باقي العناصر.

وتشكل ذرة الهيدروجين الوحيدة البروتون ذرة مثالية بالنسبة للأنظمة الحيوية لما لها من قابليات التداخل مع ذرات العناصر. وفي جميع الأحوال فان الفعاليات الحيوية او ما يعرف بالحياة يمكن ان تعرف على انها "هجرة الالكترونات" وترافق عمليات الاختزال وإضافة الالكترونات صرف كميات من الطاقة، في حين على الطرف الثاني يكون سحب الالكترونات مرافقا لإطلاق كميات مكافئة من الطاقة. ومن هذا المنطلق احتاجت الأنظمة الحيوية للطاقة للحفاظ على هويتها وقل ما يمكن هو توفر طاقة الإدامة Maintenance energy لتبقى في حالة اختزال، لان الانحراف في جهد الأكسدة سيكون مضرا بها. ولكن التذبذب بين حالة الأكسدة والاختزال هو أود العمليات الحيوية اذ تكون الجزيئات الحيوية في حالة ديناميكية بين الاختزال والتأكسد وفق الموقف المطلوب.

أقسام الأحياء الرئيسية

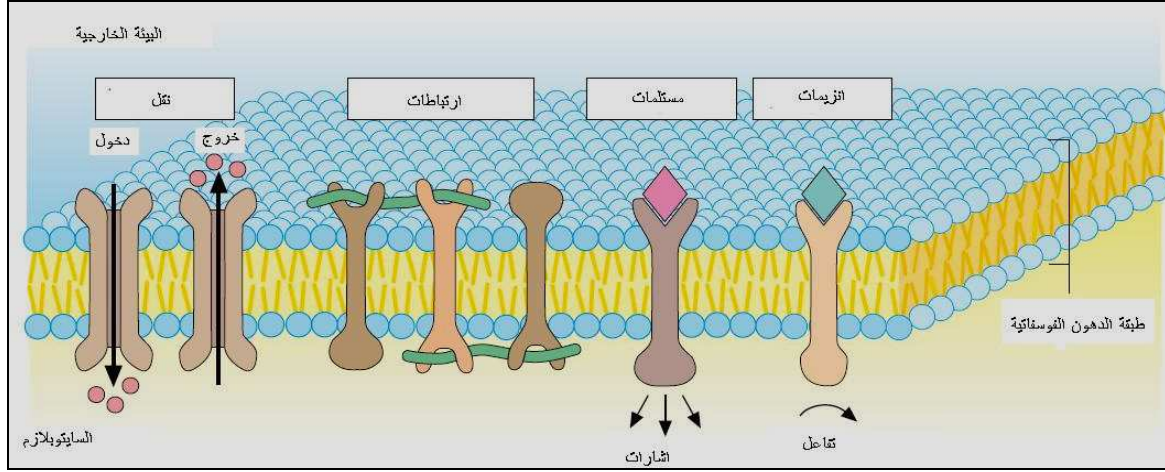
تقسم الكائنات الحية الى ثلاث أقسام او ممالك رئيسة اعتمادا على طبيعة موادها الوراثية DNA وتركيبها ووجودها ضمن الخلايا الحية. ومن هذه الأقسام هي الأحياء بدائية النواة Prokaryotes والتي تكون مادتها النووية غير مفصولة عن سايتوبلازم الخلية بغشاء او تركيب اي لا يوجد فيها تركيب النواة المميز في الأحياء الأخرى، وتضم البكتريا بمختلف اجناسها وأنواعها وتتبعها في ذلك من حيث هذه الصفة البكتريا القديمة Archaeobacteria التي فصلت في منتصف ثمانينات القرن الماضي لتوضع في دومين Domain خاص بها هو الاركيا Archaea. ويتكون جينوم هذه الأحياء من كروموسوم واحد او أكثر قد يكون حلقيا وهي الحالة الغالبة او يكون مع الكروموسوم الحلقي كروموسومات مفتوحة، فضلا عن ان الأغلبية العظمى منها تحوي على البلازميدات Plasmids وهي الأخرى قد تكون حلقية او مفتوحة مشكلة متضاعفات Replicons مستقلة في تضاعفها عن كروموسوم الخلية الأصلي بطريقة او أخرى.

والمجموعة الثانية من الأحياء هي حقيقية النواة Eukaryotes التي تكون موادها الوراثية معبئة في تركيب النواة Nucleus وتتصل بسايتوبلازم الخلية عبر قنوات خاصة في الغشاء النووي.

وتمتاز موادها الوراثية بأنها مرتبة في كروموسومات مفتوحة ذات نهايات محمية بتراكيب تعرف بنهايات أو أطراف الكروموسومات Telomeres. ويضم هذا القسم من الأحياء المملكة النباتية بمختلف شعبها وأجناسها وكذلك تضم المملكة الحيوانية بمختلف مستوياتها التطورية والتي على قمتها الإنسان . والخطوط العريضة للاختلافات بين الخلايا الحيوانية والنباتية ، هي ان الأخيرة تحوي على البلاستيدات لذا تقوم بتثبيت CO_2 وبذا تكون مفتاح الحياة على الأرض وتكون محاطة بجدار خلوي سميك من السليلوز وتحوي على الفجوات والتي تملأ منها الخلايا الحيوانية . وتضم حقيقة النواة أحياء مجهرية كثيرة تشكل الغالبية العظمى من حيث الكثرة العددية.

ويترتب على هذا التوزيع للمواد الوراثية اختلافات كبيرة بين الأحياء التابعة لمجموعة بدائية النواة والأحياء حقيقية النواة فبالإضافة الى اختلاف التركيب النووي تمتاز الأحياء حقيقية النواة بكون التراكيب الداخلية للخلايا مجزأة وكل جزء محاط بما يفصله عن باقي الأجزاء. ولعل أهم الأجزاء هي الماييتوكوندريا التي يكون جينومها في العادة حلقي وهي بيوت لإنتاج الطاقة وكذلك البلاستيدات التي يشكل جينومها الحلقي جزءاً من جينوم الكائن الكامل وهي مراكز لعمليات التخليق الضوئي. هذا فضلا عن وجود الأجسام البيروكسيمية Peroxisomes التي تقوم بمهمة تفكيك المواد في الغالب، والأجسام الحالة lysosomes التي تفكك العديد من البروتينات والحوامض النووية والدهون وغيرها من المواد. وتمتاز حقيقيات النواة بوجود الرايبوزومات التي تكون مركزاً لتخليق البروتينات مرتبطة للشبكة الاندوبلازمية لإنتاج البروتينات السكرية والدهون وتحوي الخلايا على المريكزات Centrioles بشكل جزيئات اسطوانية صغيرة مكونة من الأنابيب الصغيرة Microtubules تلعب دوراً أساسياً في انقسام الخلايا.

والخلايا تشابه الأحياء بدائية النواة من حيث احتوائها على الرايبوزومات التي تمثل مواقع تخليق البروتينات ولكن تختلف عنها في الحجم والتفاصيل. وتتشابه المجموعتان من حيث وجود الأغشية الخلوية اللازمة لانسياب الماء والمواد الأخرى من وإلى داخل الخلايا. وأوجه تشابه الأغشية يكون في انها مكونة من طبقتين من الدهون التي يكون طرفها المحب للماء الخارج والطرف الهيدروكربوني الكارهة للماء فيقع الى الداخل مكونة طبقة ثنائية. وفي الطبقة الثنائية تتطمر العديد من البروتينات، منها ما يستعمل كقنوات لنقل الجزيئات الى داخل الخلايا او طرح جزيئات أخرى الى الخارج، في حين توجد بروتينات أخرى ترتبط مع بعضها لتثبت تركيب الغشاء ويطلق عليها البروتينات التركيبية، وفضلا عن ذلك فإنها تحوي وخاصة الى الجهة الخارجية على المستلمات لنقل الإشارات الخارجية والتي تكون لها علاقة بجزيئات أخرى ذات طبيعة إنزيمية الوظيفة تقوم بالتفاعلات الكيماوية استجابة للإشارات الخارجية الواردة كما موضح في الشكل 1.3:



شكل 1.3 : تركيب الأغشية الخلوية

دورة حياة الخلايا حقيقية النواة

دورة حياة الخلايا حقيقية النواة تكون متشابهة بشكل أساسي في اغلب الأحياء لذلك فدراسة بعضها يمكن ان يؤدي الى فهم الدورة . وتمر الخلية أثناء دورتها Cell cycle بأربع أطوار رئيسية يكون نمو الخلايا فيها مستمرا ألا ان تخليق DNA يحصل في مدة محدودة ثم توزع المادة النووية قبل الانقسام النهائي للخلية لإعطاء خليتين. وتطور عمليات الانقسام وانتقال الخلايا من مرحلة او طور الى آخر يقع تحت سيطرة صارمة من أجهزة التنظيم التي تقوم بتنظيم الأحداث المختلفة إضافة الى ربط الخلايا بالإشارات الخارجية التي تؤثر في السيطرة وفي تكاثر الخلايا، وأي خلل في تنظيم دورة الخلية والإخلال بنقاط السيطرة Check points يؤدي الى عدم ثبوت جينوم الخلية وتتوفر عندها عوامل الاستعداد للسرطان ولذلك فان فهم دورة حياة الخلية يكون مهما من الناحية العلمية والطبية.

أطوار دورة الخلية: في مزارع الخلايا البشرية التي تتم دورة الخلية فيها في مدة 24 ساعة يلاحظ ان الدورة تنقسم الى قسمين رئيسية هي:

1- الانقسام الخيطي Mitosis

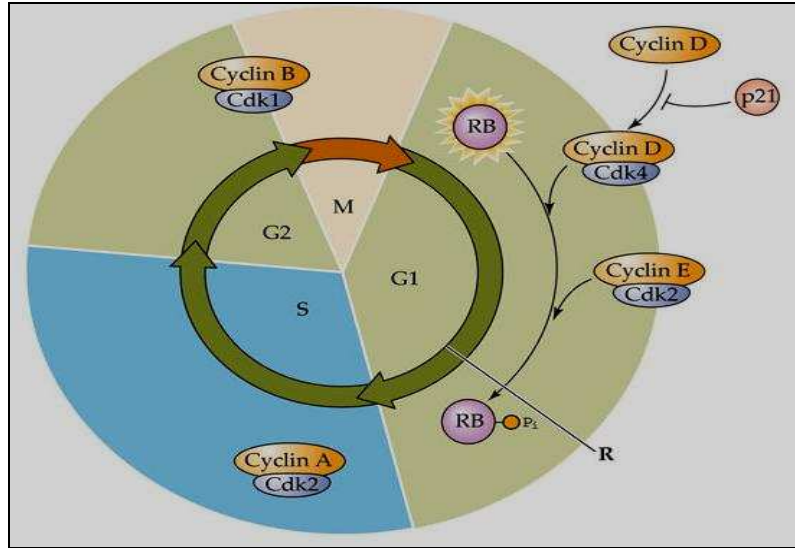
2- الطور البيني Interphase

ويمثل الانقسام الخيطي (انقسام النواة) الجزء الأهم من دورة الخلية وبعده تتوزع المادة الوراثية على نواتين يليها انقسام السايكوبلازم Cytokinesis وينتهي الانقسام الخيطي للنواة وانقسام السايكوبلازم في حوالي مدة ساعة اما الباقي الذي يمثل أكثر من 95% من دورة الانقسام فيشغلها الطور البيني الذي يمثل الطور بين انقسامين.

وأثناء الطور البيني تقل كثافة الكروموسومات وتوزع في النواة التي تبدو متجانسة، وعلى المستوى الجزيئي فانه أثناء هذا الطور يتضاعف DNA ومكونات السايكوبلازم بشكل مرتب ومنتالي استعدادا لدورة انقسام جديدة، اذ يكون النمو معتدلا الى ان يتضاعف حجم الخلية ولذلك فان توقيت تضاعف DNA يقسم الدورة الى أربعة ادوار او مراحل مميزة يختلف طولها بشكل كبير في الأنواع المختلفة من الخلايا وهي:

- **M phase (Mitosis):** ويعني الانقسام الخيطي الذي يعقبه انقسام السايوتوبلازم ويمتد لمدة حوالي ساعة في الخلايا البشرية سريعة الانقسام والتي يكون زمن دورتها 24 ساعة.
 - **G1 (Gap 1):** وتعني المرحلة الأولى وتمثل المدة بين الانقسام الخيطي وبدء تخليق DNA وتستمر حوالي 11 ساعة (لخلايا تنقسم في 24 ساعة) والخلايا في هذه المرحلة تكون نشطة من الناحية الأيضية وفي حالة نمو مستمر ولكن لا يوجد تضاعف للـ DNA.
 - **S phase (Synthesis):** وفيه يحدث تخليق وتضاعف DNA يلي المرحلة السابقة ويطول الطور أو المرحلة إلى حوالي 8 ساعات.
 - **G2 (Gap 2):** المرحلة التي تلي تخليق وتضاعف DNA وتطول حوالي 4 ساعات، وتستمر فيها الخلايا بالنمو وتخليق البروتينات استعدادا للانقسام الخيطي.
- والأطوار موضحة في الشكل (2.3).

وتختلف الخلايا في سرعة النمو وطول دورة الخلية فمثلا في الخمائر المتبرعمة تستمر دورة الخلية إلى حوالي 90 دقيقة تقريبا. وفي خلايا أخرى تصل الدورة إلى 30 دقيقة كما في الخلايا الجنينية بعد التخصيب مباشرة، إذ أن مثل هذه الخلايا لا تحتاج إلى نمو إذ ينقسم سايوتوبلازم البويضة إلى خلايا صغيرة ويلاحظ اختفاء كل من مرحلة G1، G2 ولكن يحدث تضاعف سريع للـ DNA وبذلك يكون S طور قصير ويتبادل مع M. وعلى النقيض فإن بعض خلايا الإنسان أو الحيوان البالغ مثل الخلايا العصبية تتوقف عن الانقسام، وأنواع أخرى تنقسم في مراحل معينة أو في حالات معينة عندما يراد تعويض الخلايا المتضررة أو الميتة أو عندما تحفز بعض الخلايا مثل الخلايا المناعية بإشارات من خارج الخلية.



شكل 2.3 : أطوار دورة الخلايا الحقيقية النواة

والخلايا عند توقفها عن الانقسام فإنها تخرج من مرحلة G1 لتدخل مرحلة G₀ إذ تكون الخلايا نشطة من الناحية الأيضية ولكن لا تتكاثر إلا إذا استدعت الضرورة بواسطة إشارات تأتي من

خارج الخلية . والخلايا في المراحل أعلاه يمكن ان تميز بطرق معينة، فالخلايا المنقسمة يمكن ان تميز بالمجهر، اما الأطوار الأخرى فتميز باستعمال مؤشرات كيميائية. ففي طور التخليق (S) يمكن متابعتها بتسجيل كمية ما يدمج من الثايميدين المعلم او المشع الذي يستعمل بشكل أساسي في تخليق DNA. كما يمكن تمييز الخلايا من محتواها من DNA:

• فالخلايا في G1 تكون مزدوجة الكروموسومات Diploid (ضعفاني) اي يوجد نسختين من كل كروموسوم (2n) (n تمثل عدد الكروموسومات ومحتوى DNA في الحالة الفردانية Haploid للجينوم).

• الخلايا في مرحلة S يزداد فيها DNA وتصل الى 4n اي ان المحتوى يتراوح بين 2n ← 4n.

• اما الخلايا في طور G2 فيبقى DNA بمستوى 4n.

• والخلايا في طور M يتناقص DNA الى 2n بعد انقسام السايكوبلازم. وتنظم دورة الخلية استجابة لإشارات من البيئة المحيطة او من إشارات داخلية باشتراك عدد كبير من البروتينات والإنزيمات منها كينازات البروتينات Protein kinases وتعرف بـ Cyclin- dependent kinases (CDKs) وذلك لأنها تنشط ببروتينات تدعى المدورات Cyclins التي تختلف مستوياتها أثناء دورة الانقسام ويمكن ان تفكك في مراحل معينة وتعود وتخلق من جديد بالتعبير عن جيناتها الخاصة لذلك سميت بالمدورات لان وجودها يكون بشكل دوري.

اما الفيروسات والتي تتمثل هويتها بالمواد الوراثية مثل DNA او RNA مع قليل من المستلزمات البروتينية فهي تكون فيصل بين الحياة ومظاهرها عندما تكون متطفلة داخل الأنظمة الحيوية وبين الالحياة عندما تكون خارج الأنظمة الحيوية، وذلك لنقصان العديد من مقومات الحياة لديها.

وفي هذا الفصل سيتم التطرق الى النواحي الحيوية للأحياء سواء كانت التركيبية منها او الوظيفية بقدر تعلق الموضوع بالمعلوماتية الحيوية.

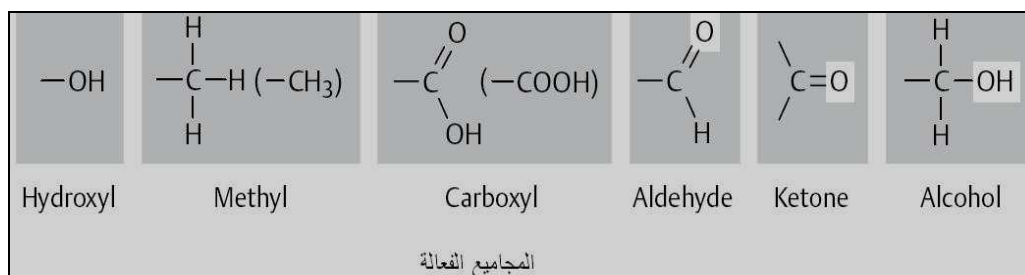
جزيئات الحياة Biomolecules

تتمثل في العادة بالجزيئات العملاقة مثل DNA او RNA والبروتينات بشكل رئيس، لان هذه الجزيئات الثلاث هي الجزيئات الرئيسة وتكون المواد الحيوية الأخرى هي مشتقات مباشرة او غير مباشرة عنها.

الأواصر الكيميائية في الأنظمة الحيوية

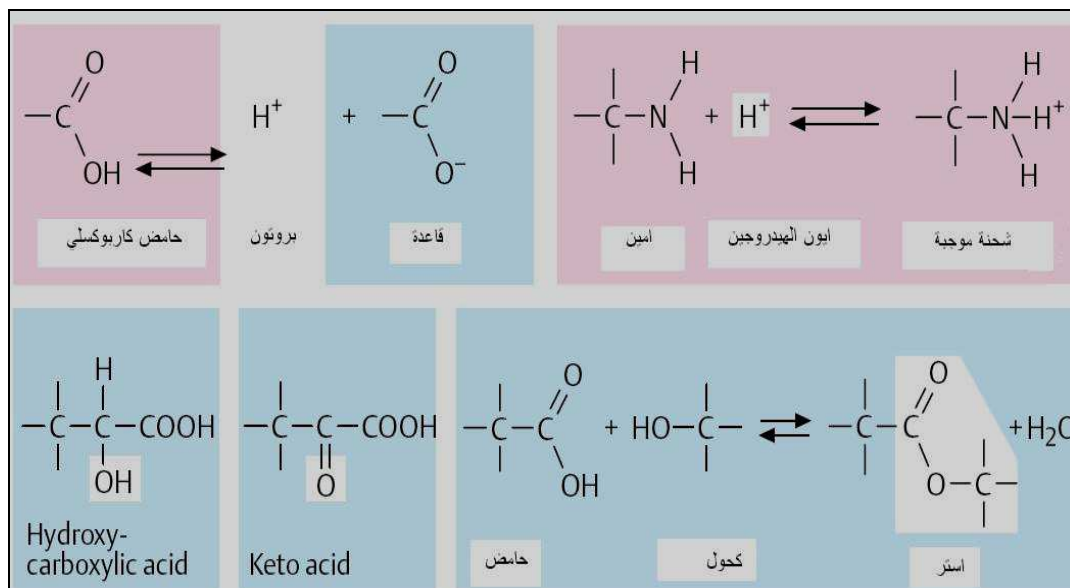
الخلايا الحية مكونة من حوالي 18% من كربون 10% الهيدروجين، 65% الأوكسجين وحوالي 3% نيتروجين. وبشكل الماء حوالي 70% من وزن الخلايا اما الباقي فهي مركبات للكربون التي تكون بالشكل الأكثر اختزالا (تكافؤها -4) وتكون هذه الذرة صغيرة وترتبط مع ذرات أخرى بأواصر تساهمية مكونة جزيئات ترتبط هي الأخرى مع بعضها مكونة سلاسل او تراكيب حلقية ذات أهمية حيوية .

وارتباط الكربون مع الهيدروجين أو الأوكسجين يمكن ان يؤدي الى تكوين بعض المجاميع الهامة مثل الهيدروكسيل Hydroxyl (-OH) او مجموعة ميثيل (-CH₃)، او مجموعة كربوكسيل Carboxyl (-COO⁻) او تكون بشكل كاربونيل Carbonyl (C=O) كما في الكيتونات والالديهيدات ، هذه المجاميع تضيف الصفات الكيماوية على الجزيئات التي ترتبط او توجد فيها والمجاميع موضحة في الشكل 3.3:



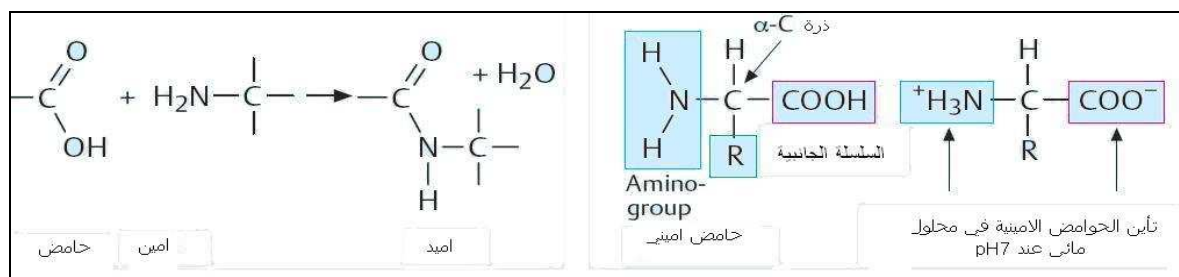
شكل 3.3 : المجاميع الفعالة في جزيئات الحياة

وارتباطات الكربون مع الأوكسجين والهيدروجين يمكن ان يكون ضمن مركبات حامضية او قاعدية. والتعبير العادي عن الحموضة يكون بقياس الرقم الهيدروجيني pH الذي يمثل تركيز ايون الهيدروجين [H⁺] الممتد على مدى فالتركيز 10⁻¹ مول\ لتر يعني ان الرقم الهيدروجيني حامض جدا (pH=1) والتركيز على النقيض الآخر 10⁻¹⁴ مول\ لتر يكون قاعدي جدا (pH=14). والماء النقي يمثل حالة الاعتدال حيث يكون حاويا على 10⁻⁷ مول\ لتر (pH=7). وفضلا عن هذا فان الذرات والمجاميع البسيطة تتداخل وتتفاعل مع بعضها مكونة تراكيب تعطي تفاعلات وقيم للرقم الهيدروجيني مختلفة فضلا عن تكوين مركبات مهمة تدخل في تركيب الجزيئات العملاقة والشكل التالي (4.3) يوضح جزءا مهما منها:



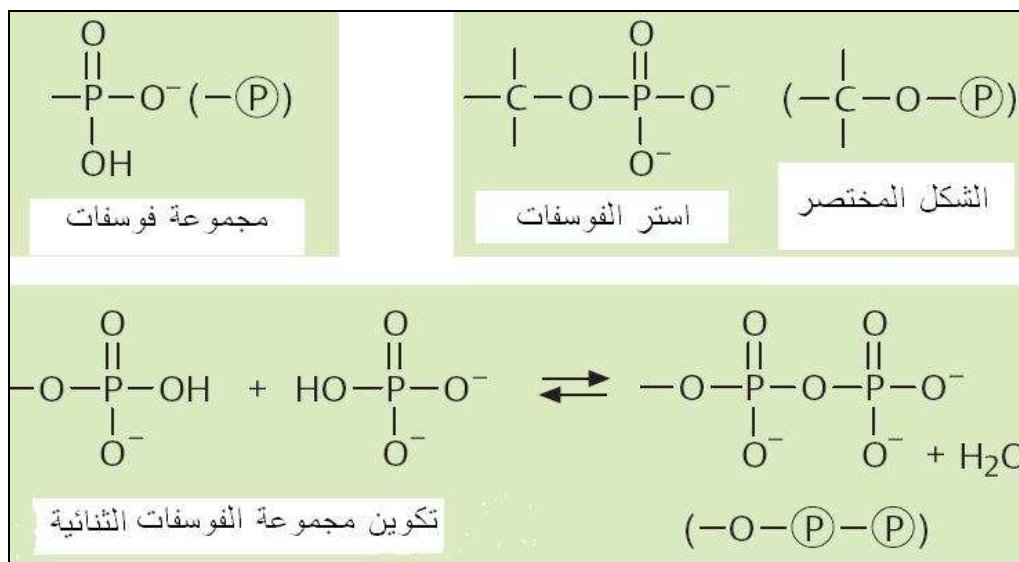
شكل 4.3 : التراكيب المهمة لعنصر الكربون التي تدخل في تكوين جزيئات الحياة

اما اشتراك الكربون مع النتروجين وتفاعلهما فانه يؤدي الى تكوين مجاميع ومركبات مختلفة عما ذكر آنفاً. فالأواصر C-N توجد في العديد من الجزيئات الحيوية منها مجاميع الأمين Amino groups، والأمينات Amines والأميدات Amides الخاصة بالبروتينات، وأكثرها أهمية هي الحوامض الأمينية ذات الأهمية الكبيرة سواء في بناء الكائنات او وظائفها. والشكل التالي (5.3) يوضح بعض ارتباطات الكربون مع النتروجين وتفاعلاتها.



شكل 5.3 : التراكيب التي يشترك فيها النتروجين وتفاعلاتها

وما ذكر أعلاه يخص العناصر الرئيسة في الأنظمة الحيوية اي التي تكون بكميات كبيرة، ولكن هناك بعض العناصر التي تكون نسبها قليلة ولكن لها الأثر الكبير في الفعاليات الحيوية ومنها: مركبات الفسفور : وأهمها مجموعة الفوسفات المتأينة التي تلعب دورا حيويا في تحولات الطاقة الحيوية Bioenergetics في الأنظمة الحيوية، وذلك بالتأرجح بين مركبات الفوسفات المختلفة فضلا عن ان مجموعة الفوسفات تتفاعل مع مجاميع الهيدروكسيل لتكون استر الفوسفات Phosphate ester لتشكل العمود الفقري للحوامض النووية سواء DNA او RNA كما موضح في الشكل 6.3:



شكل 6.3 : تفاعلات مجموعة الفوسفات

مركبات الكبريت : تلعب هي الأخرى بعض الأدوار المهمة وذلك بارتباطها للجزيئات الحيوية مع بعضها خاصة مجموعة sulfhydryl (-SH) التي تتحد لتكوين الجسور أو الأواصر الكبريتيدية الثنائية Disulphide bridges (-S=S-) لتساهم في تكوين التراكيب الثلاثية للبروتينات (كما سيأتي ذكره) ، ويوجد الكبريت في الحامض الأميني الميثاينين والسستئين Cysteine وكذلك يوجد في بعض السكريات والسكريات المكثرة Polysaccharides.

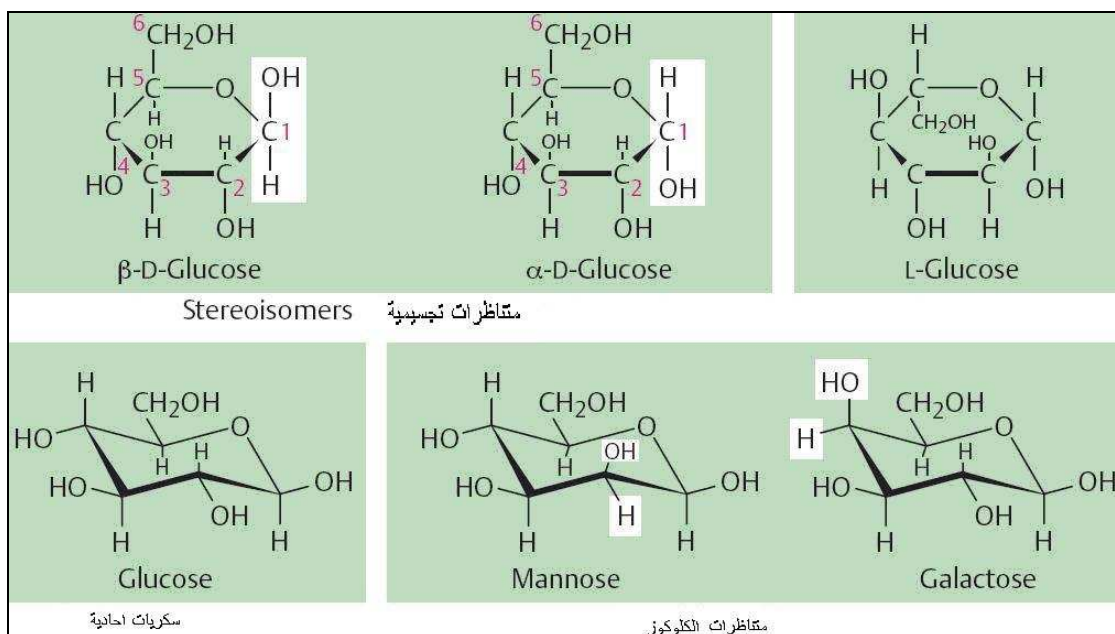
الكربوهيدرات

مركبات كيميائية مختلفة ولها مشتقات كثيرة تكون ضرورية لبناء المواد الوراثية فسكر الرايبوز Ribose يشكل العمود الفقري للـ RNA وسكر الرايبوز منقوص الأوكسجين Deoxyribose يشكل العمود الفقري للـ DNA.

ونظرا لمرونة هذه الجزيئات فهي يمكن ان تعمل في مجالات شتى مثل نقل المعلومات الوراثية أي تساعد في العمليات التي يقوم بها DNA وكذلك تساعد في تحولات الطاقة الحيوية، فضلا عن مشاركتها في النواحي التركيبية مثل الجدران الخلوية في البكتريا والنباتات وغيرها. كما ان ارتباطها مع الدهون والبروتينات يؤدي الى تكوين جزيئات او تراكيب قد تكون على سطوح الخلايا بمثابة مستضدات Antigens او مستلمات لنقل الإشارات من البيئة المحيطة بالخلايا الى داخل الخلايا. وتقسم الى عدة مجاميع ومنها السكريات التي تقسم بدورها الى مجاميع ثانوية أخرى.

السكريات الأحادية Monosaccharides

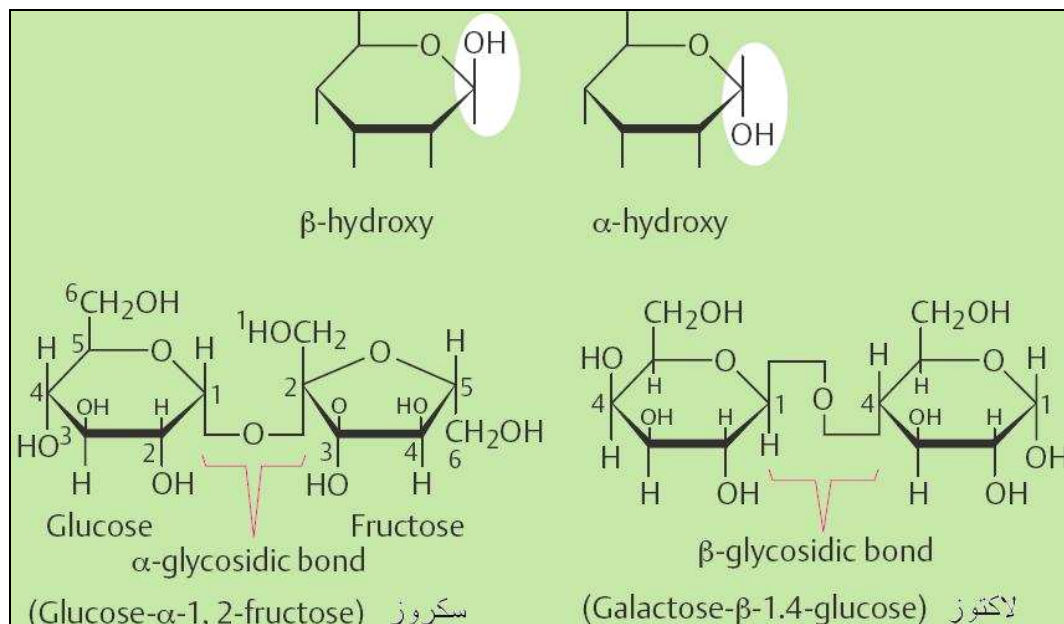
جزيئات بسيطة بشكل الديهايدات ($-C=O$) أو تكون بشكل كيتونات وترتبط باثنين أو أكثر من مجاميع الهيدروكسيل ، ولها الصيغة الكيميائية العامة $(CH_2O)_n$. ومجاميع الكيتونات أو الالديهايدات يمكن ان تتفاعل مع احد مجاميع الهيدروكسيل مكونة تركيب حلقي وهو التركيب الطبيعي للسكريات الخماسية والسداسية. ويتم ترقيم ذرات الكربون باتجاه عقرب الساعة. ولها متناظرات ضوئية هي D (Dextro-) الذي يحرف الضوء الى اليمين والنظير اليساري L (Levo-) الذي يحرف الضوء الى اليسار، وكذلك يمكن ان تكون الجزيئات اما β - أو α - التي تمثل النظائر المجسمة للجزيئات. وفي الأشكال الحلقية لا تكون ذرات كربون على مستوى واحد وإنما تتخذ أشكالاً مجسمة تشبه الكرسي أو القارب ، فمثلا التوزيع β -D-glucopyranose, لجزيئة الكلوكوز يكون هو الأفضل من ناحية الطاقة، وفضلا عن ذاك فان توزيع ذرات الهيدروجين يمكن ان يختلف مكونا نظائر أخرى للكلوكوز مثلا ليعطي المانوز أو الكلاكتوز كما موضح في الشكل 7.3:



شكل 7.3 : تركيب السكريات الأحادية

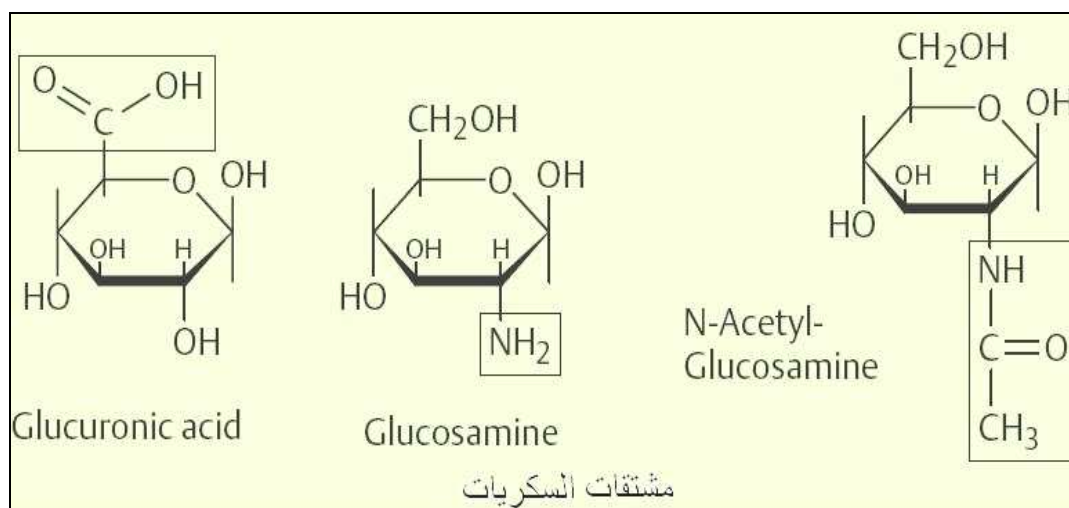
السكريات الثنائية

مركبات تنتج من ارتباط اثنين من السكريات الأحادية، اذ ترتبط مجموعة الالديهايد أو الكيتون من السكر الى مجموعة α -hydroxyl أو β -hydroxyl من السكر الثاني، وأكثر السكريات شيوعا هي سكر المائدة (السكروز Sucrose) وسكر الحليب Lactose الموضحة تراكيبها في 8.3:



شكل 8.3 : تراكيب السكريات الثنائية المهمة

وللسكريات مشتقات يمكن ان تنتج من استبدال مجموعة هيدروكسيل بمجموعة أخرى، مثل استبدال (OH) مع مجموعة أمين، او يستبدل بمجموعة N-acetyl group كما موضح في الشكل (9.3)، وهذه الاستبدالات تحدث في بعض السكريات المكثرة واضطرابها يؤدي الى بعض الأمراض الوراثية.



شكل 9.3 : بعض مشتقات السكريات

كما ان السكريات البسيطة يمكن ان تتجمع مع بعضها لتكون السكريات المكثرة التي تختلف في أطوال سلاسلها وتمثل التراكيب الأساسية للخلية، فضلا عن ان بعضها يرتبط بالدهون او

البروتينات لتكون تراكيب سطحية مثل المستضدات مثل مستضدات مجاميع الدم وغيرها من التراكيب.

الدهون

جزيئات حيوية مهمة تنتج في العادة من ارتباط الكليسرول مع الحوامض الدهنية وتكون الأساس في تركيب الأغشية الخلوية التي تكون بمثابة الحاجز الفاصل بين البيئة الداخلية الحيوية والبيئة الخارجية وتكون الصفة المميزة لها هي كراهية الماء Hydrophobic. تشارك الدهون في تكوين مركبات حيوية مهمة مثل الستيرويدات وبعض الهرمونات وجزيئات أخرى تعمل في نقل الإشارات داخل الخلايا. وهناك أصناف عدة منها مثل الدهون السكرية Glycolipids والدهون الفوسفاتية Phospholipids التي تساهم في بناء الأغشية الخلوية أنفة الذكر، إذ تكون لها صفات تؤهلها لبناء الأغشية الخلوية والعمل كحواجز، فهي تمتلك طرف محب للماء والجزء الآخر كاره للماء، وبذا تصطف بشكل طبقة ثنائية تكون مناطقها الخارجية محبة للماء Hydrophilic وتطمر في داخلها المناطق الهيدروكربونية الكارهة للماء.

الحوامض الدهنية Fatty acids

سلسلة هيدروكربونية تنتهي بمجموعة كربوكسيل (COOH -) أي يكون هذا الطرف محبا للماء، والنهاية الثانية الحاوية على (CH_3 -) كارهها للماء، وتختلف أطوال سلاسلها ودرجة التشبع للأواصر بين ذرات الكربون فهي إما أن تكون غير مشبعة (وجود الأواصر المزدوجة) وهذه تضيف على الجزيئة صلادة. وتمثل جزءا مهما في تركيب العديد من الجزيئات العملاقة. وتتأين الحوامض الدهنية وتتذبذب بين COOH إلى COO^- .

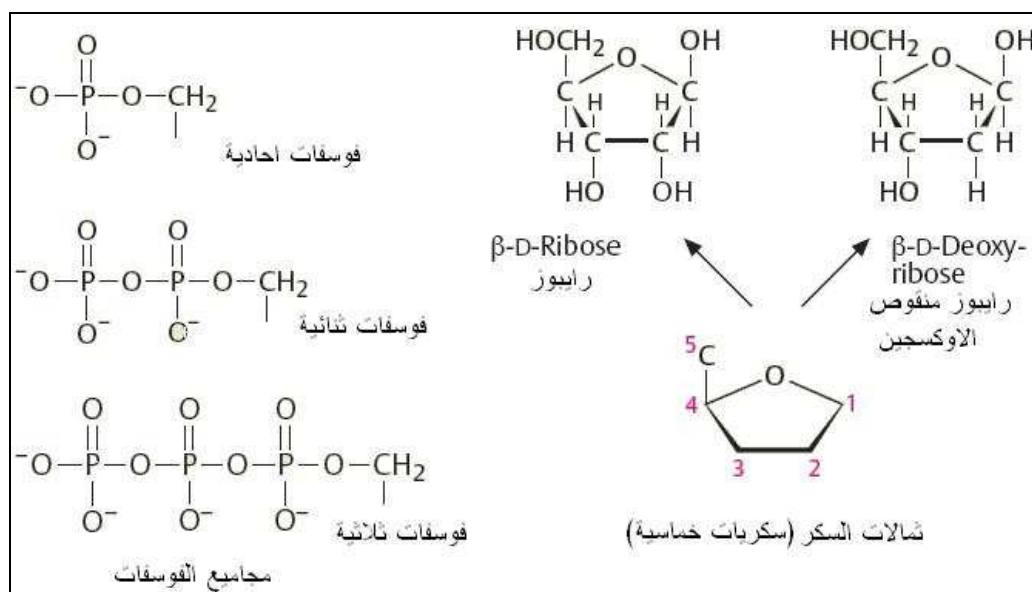
مشتقات الدهون

ترتبط الحوامض الدهنية مع غيرها من الجزيئات مكونة تراكيب مثل الدهون الكارهة للماء ولا تذوب فيه وإنما تذوب في المذيبات العضوية. ويمكن لمجموعة الكربوكسيل أن تكون أواصر استرية أو اميدية. ويمكن أن تختلف الجزيئية التي يرتبط بها الحامض الدهني فبدلا من الكليسرول في الكليسيريدات الثلاثية Triglycerides الشائعة الوجود يمكن أن ترتبط إلى Sphingosine لتكون Sphingolipids المهمة في الأغشية الخلوية. فضلا عن ذلك فإن الحوامض الدهنية والدهون الناتجة منها قد تكون ضمن تشكيلات كيميائية مختلفة لتكون الكوليسترول والهرمونات وغيرها .

النيوكليوتيدات والحوامض النووية

هذه الجزيئات أكثر أهمية من ناحية الدراسات في المعلوماتية الحيوية إضافة إلى البروتينات، لذلك سيتم التركيز على دراساتها أكثر من الجزيئات الحيوية المارة الذكر. النيوكليوتيدات تشارك في معظم الفعاليات الحيوية ولعل أهمها مشاركتها في تركيب DNA و RNA أي الجزيئات الحاملة للمعلومات الوراثية ، فضلا عن مشاركتها في تحولات الطاقة الحيوية مثل ATP و GTP، وكذلك تكون في بعض الأحيان جزءا مهما في الإنزيمات (Cofactor) وتشارك في عمليات نقل الإشارات وعمليات تنظيم الفعاليات الحيوية. يتكون كل نيوكليوتيد من ثلاثة جزيئات :

- ثمالة السكر وهذه في الغالب قد تكون رايبوز كما في RNA او تكون رايبوز منقوص الأوكسجين في DNA كما في الشكل (10.3)
- مجموعة الفوسفات وهي اما ان تكون بمثابة مجموعة واحدة او اثنين او ثلاثة وترتبط الى ذرة الكربون رقم 5 في السكر الخماسي وتراكيبها موضحة في الشكل (6.3) المذكور سابقا .

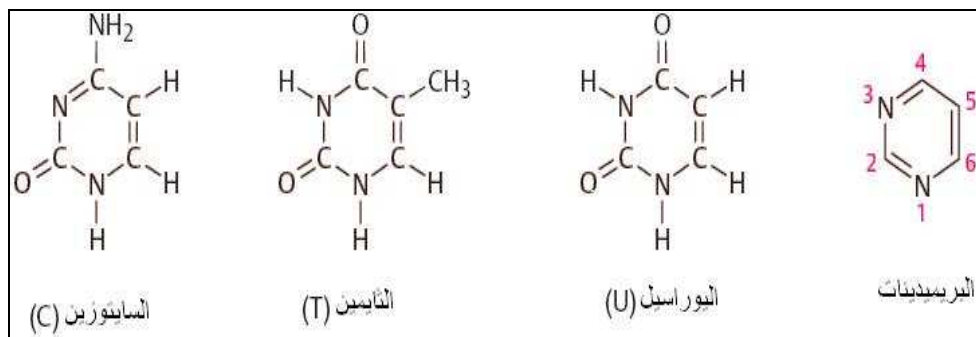


شكل 10.3 : تراكيب سكريات الحوامض النووية

- الجزء الثالث هو القواعد النيتروجينية وهذه تقسم الى قسمين:

القواعد البريميدينية Pyrimidines

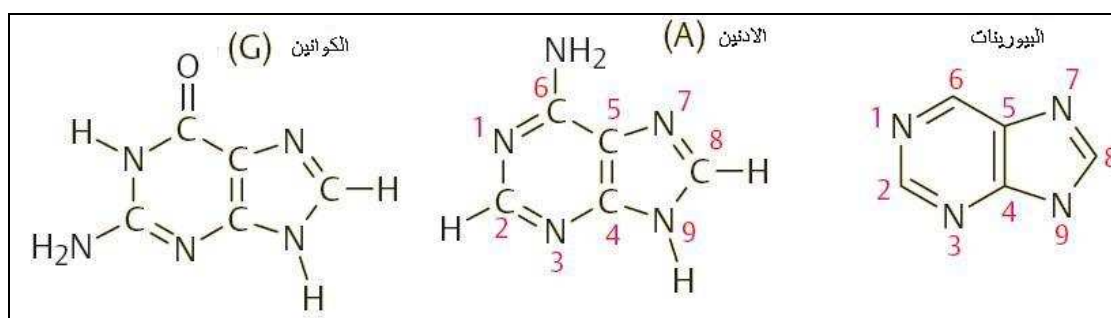
قواعد أحادية الحلقة، تركيبها موضح في أدناه وتضم ثلاثة قواعد مهمة في المواد الوراثية، وتختلف القواعد فيما بينها في السلاسل الجانبية التي ترتبط الى ذرات كربون حلقة البريميدين. الساييتوزين Cytosine (C) يحوي على مجموعة أمين ($-NH_2$) على الذرة الرابعة وهذه تشكل مناطق ساخنة لحدوث الطفرات نظرا لإزالة مجموعة الأمين ، وقاعدة الثايمين Thymine (T) تحوي على مجموعة مثيل على الذرة الخامسة. وقاعدة اليوراسيل Uracil (U) التي تحوي على الأوكسجين على الذرة الرابعة من الحلقة فضلا عن وجود اختلافات في توزيع الأواصر المزدوجة في الحلقة وتحل في العادة محل الثايمين والتراكيب موضحة في الشكل 11.3:



شكل 11.3 : القواعد النيتروجينية البريميدينية

القواعد البيورينية Purines

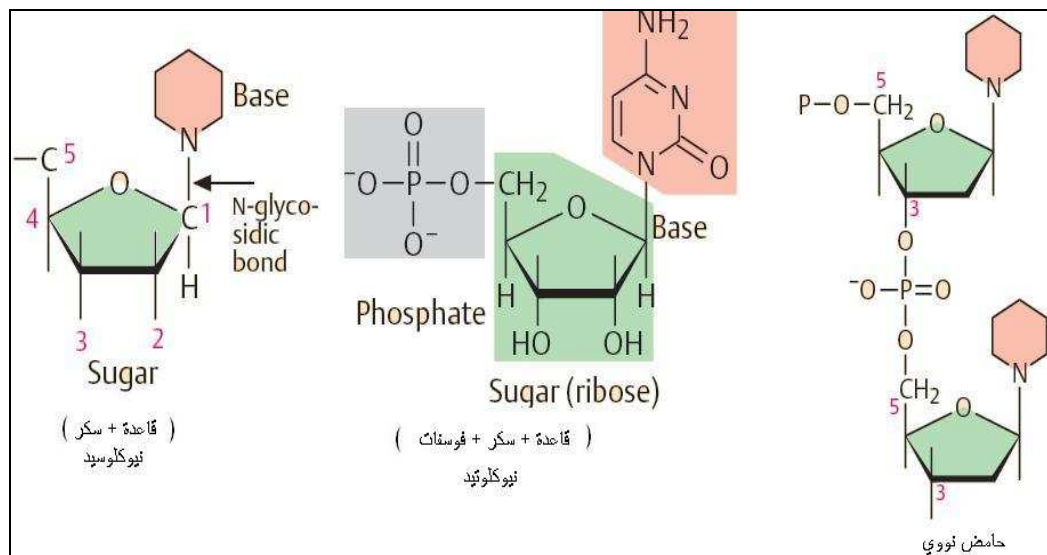
قواعد نيتروجينية ثنائية الحلقة تشمل الأدينين (A) Adenine والكوانين (G) Guanine التي تدخل في تركيب الحوامض النووية، تختلف القاعدتان في ترتيب الأواصر المزدوجة وكذلك المجاميع المرتبطة بذرات كربون الحلقة كما موضح في الشكل (12.3) :



شكل 12.3 : القواعد النيتروجينية البيورينية

النيوكليوسيدات Nucleosides

المركبات الناتجة من اتحاد السكريات (سواء سكر الرايبوز او الرايبوز منقوص الأوكسجين) مع احد القواعد النيتروجينية، والارتباط يكون بين ذرة كربون الأولى لجزيئة السكر والنيتروجين من القاعدة النيتروجينية بأصرة N-glycosidic bond وتسمى اعتمادا على نوع السكر الداخل في تركيبها وهي موضحة في الشكل 13.3:



شكل 13.3 : تركيب النيوكلوسيدات و النيوكلوتيدات

جزيئات مكونة من ثمانية سكر خماسي (وفي مجال المواد الوراثية يكون بطبيعة الحال اما الريبوز او الريبوز منقوص الأوكسجين)، ترتبط ثمانية السكر بأحد القواعد النتروجينية المذكورة أعلاه وكذلك ترتبط الى مجموعة فوسفات كما موضح في الشكل أعلاه :

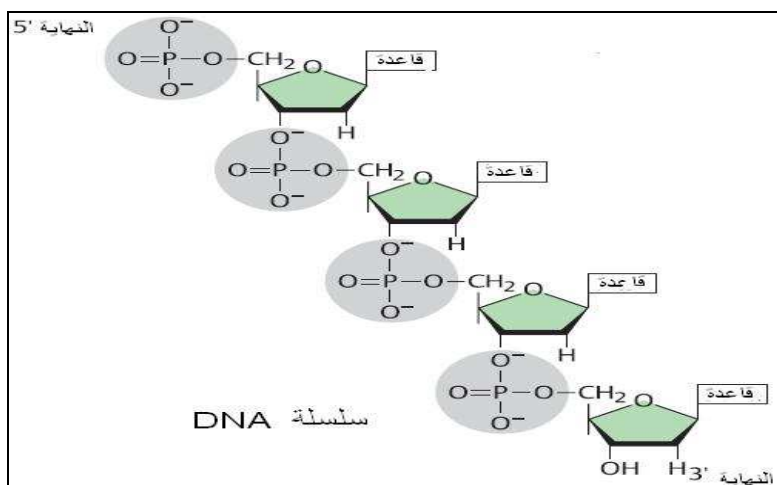
والنيوكليوتيدات هي الوحدات المكونة للحوامض النووية. وتختلف النيوكليوتيدات من حيث عدد ذرات الفوسفات المرتبطة فقد تكون مجموعة واحدة كما في Adenosine (AMP) monophosphate ويطلق في العموم أسماء مختصرة للنيوكليوتيدات مثل Adenylate وكذلك تشترك السكريات في تسميتها:

فالمجموعة الحاوية على سكر الريبوز تسمى Ribonucleotides مثل:

GMP	Guanylate monophosphate
UMP	Uridylate monophosphate
CMP	Cytidylate monophosphate
اما المجموعة الحاوية على سكر الرايبوز منقوص الأوكسجين مثل:	
dAMP	deoxyadenylate monophosphate
dGMP	deoxyguanylate monophosphate
dTMP	deoxythymidylate monophosphate
dCMP	deoxycytidylate monophosphate

الحوامض النووية Nucleic acids

مكوّنات حيوية تتكون عندما ترتبط النيوكليوتيدات مع بعضها بواسطة جسور فوسفاتية ثنائية الاستر Phosphodiester bridges بين ذرة كربون 3' للنيوكليوتيد وذرة 5' للنيوكليوتيد آخر كما موضح في الشكل 14.3:



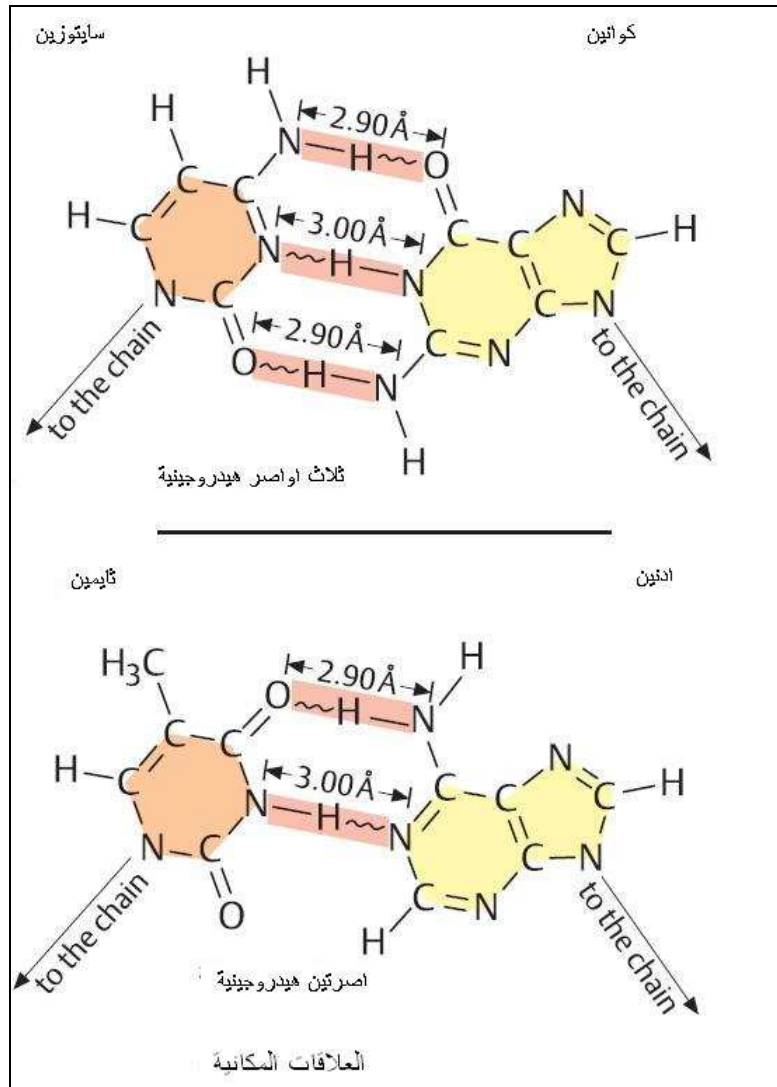
شكل 14.3 : توالي النيوكليوتيدات المكونة للحوامض النووية

والتوالي للنيوكليوتيد يعطي أو يكتب عادة بالاتجاه من 5' إلى 3' فمثلا التوالي ACCT يعني ان A في الموقع 5' وان T في الموقع 3'.

DNA

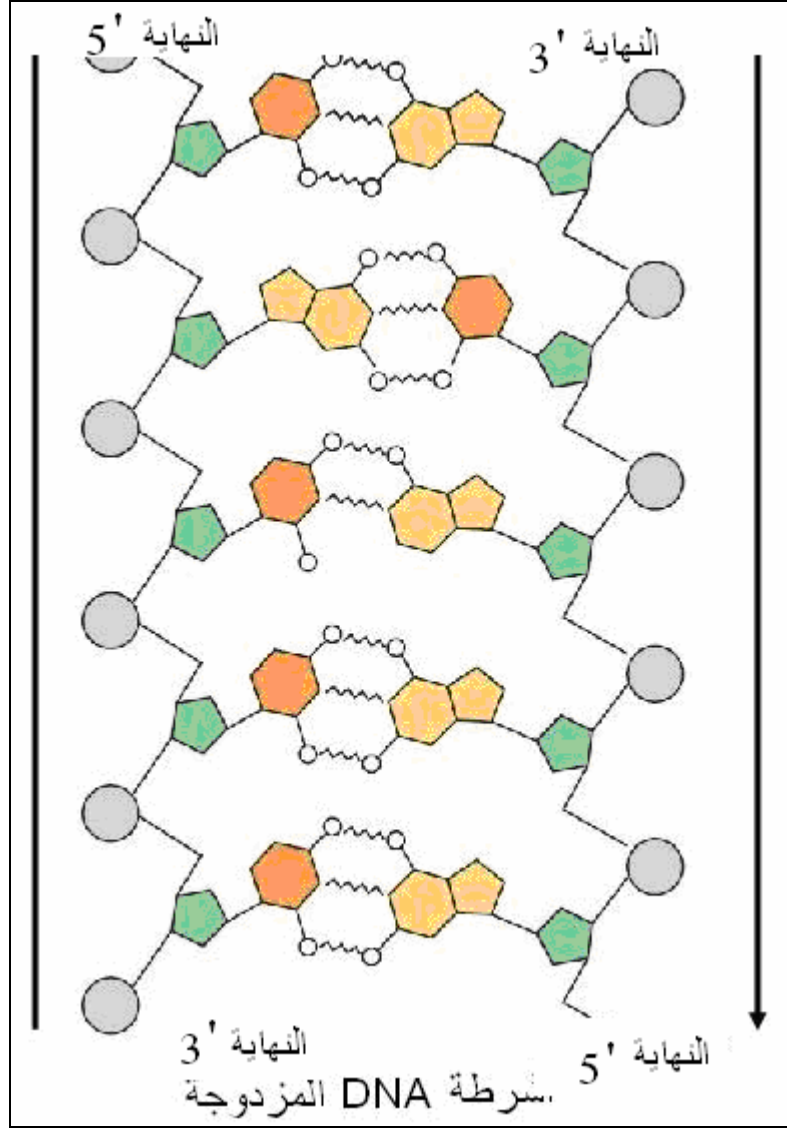
يمثل المكوّن الحيوي الأهم في الخلايا وكان يسمى النويين Nuclein ، ويمثل وديعة الصفات الوراثية. وهو سلسلة من النيوكليوتيدات فيها سكر الرايبوز منقوص الأوكسجين، ترتبط مجموعة الفوسفات لأحد النيوكليوتيدات مع جزيئة السكر للنيوكليوتيد الآخر، وبهذا الارتباط يتكون الجزء الثابت من تركيب DNA ، والجزء المتغير فيه هو القواعد النتروجينية. وتكون جزيئة DNA قطبية وذلك نتيجة لعملية الارتباط كما موضح في الشكل أعلاه .

لذلك تكون إحدى نهايتي السلسلة حاملة لمجموعة الفوسفات الحرة 5'-triphosphate والنهاية الثانية حاوية على مجموعة الهيدروكسيل الحرة 3'-OH. ويرتبط شريطين من جزيئات DNA لتكوين الحلزونات المزدوجة Duplexes حيث تتقابل البيورينات G,A مع البريميدينات C,T وتتكون ثلاث أواصر هيدروجينية بين C و G وأصرتين بين A و T أي تكون متكاملة ولا يمكن ان تتقابل بشكل آخر، أي ان كل قاعدة بيورينية يجب ان تقابل قاعدة بريميدينية ولا يمكن ان تتقابل البيورينات او البريميدينات مع بعضها وبمعنى آخر ان كل قاعدة نتروجينية أحادية الحلقة يجب ان تقابل قاعدة ثنائية الحلقة أي تكون هناك علاقة مكانية كما موضح في الشكل 15.3:



شكل 15.3 : تكامل أشرطة DNA

وبذا فتوالي القواعد النيتروجينية في احد الأشرطة في الاتجاه 5' الى 3' يكون مكملًا لتوالي القواعد النيتروجينية على الشريط الثاني والذي يكون بالاتجاه المعاكس اي من 3' الى 5'، لذلك كان تخصص ازدواج القواعد في أشرطة DNA من أهم الخصائص التركيبية لجزيئة DNA العملاقة، والشكل (16.3) يوضح هذه العلاقة.

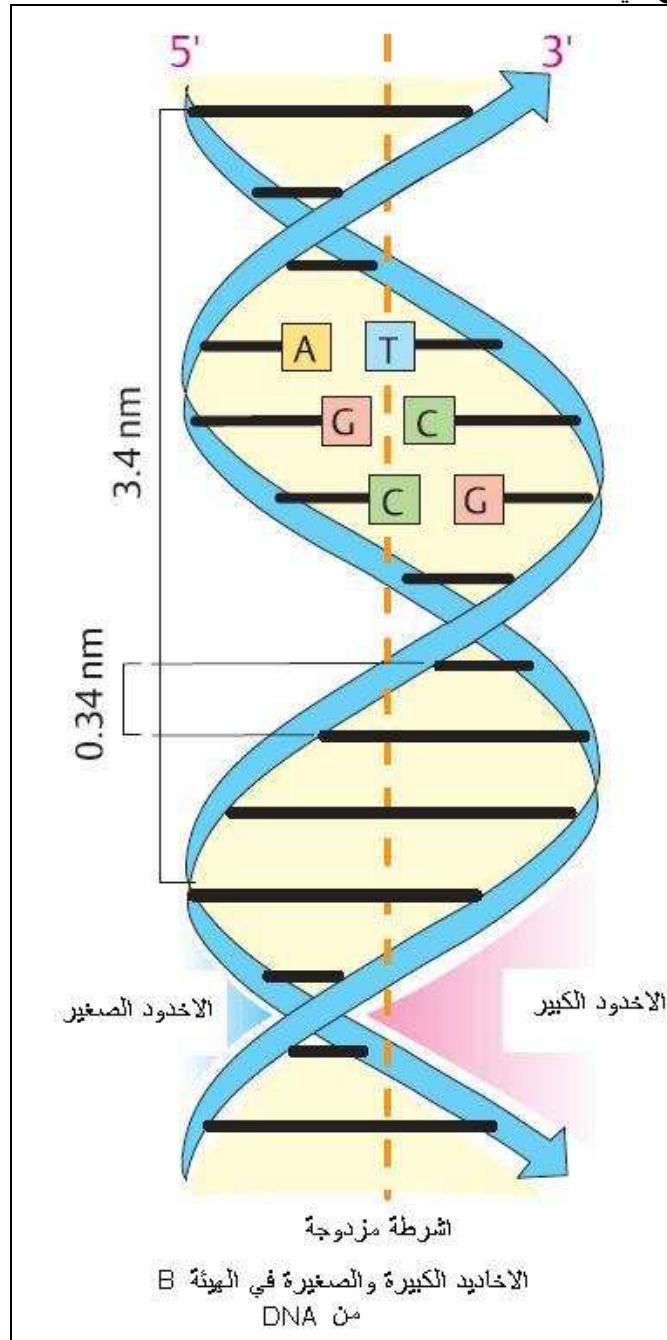


وبذا يكون التركيب العام لجزيئة DNA هو الشريط المزدوج ولكن في حالات خاصة او شاذة قد توجد جزيئات DNA بشكل أشرطة ثلاثية نتيجة لتكون أواصر هيدروجينية مع بعض قواعد الكوانين.

هياكل DNA

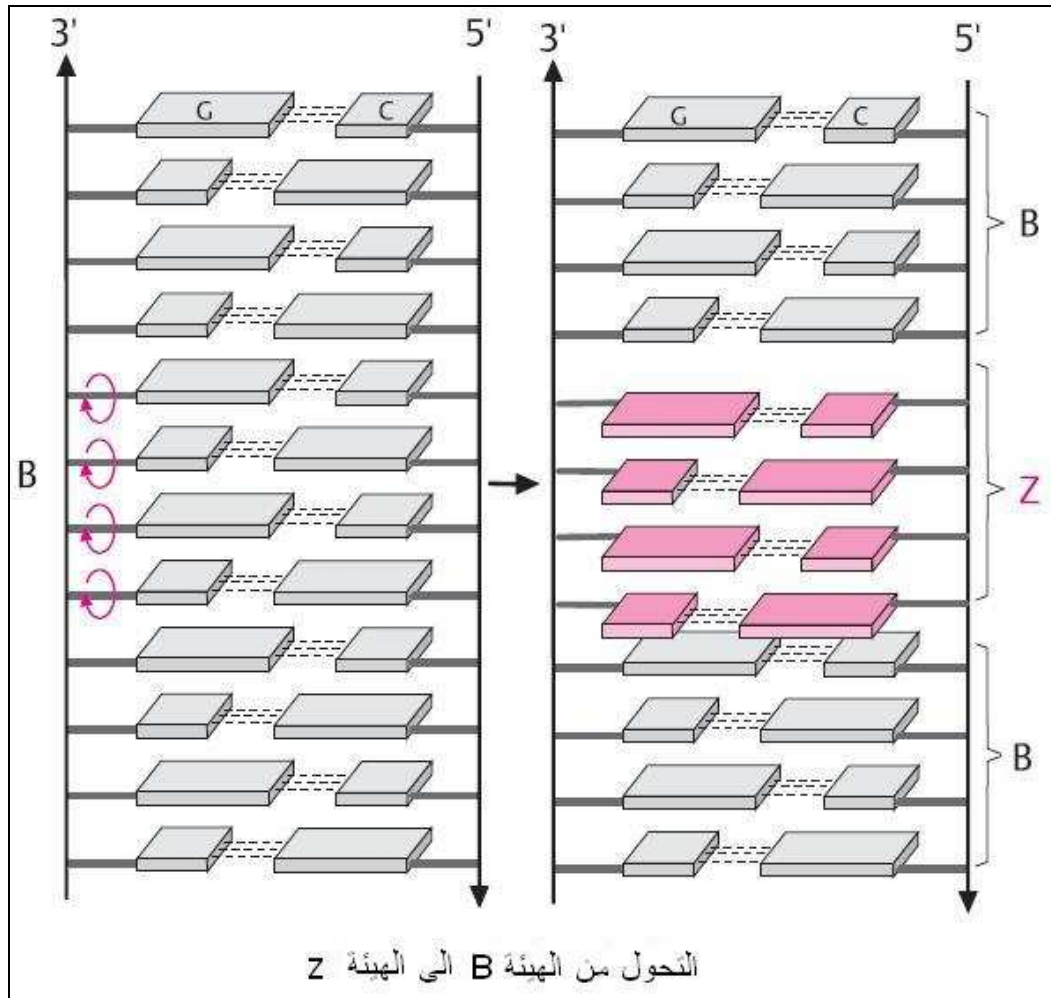
اتضح مما ذكر أعلاه ان التركيب الحلزوني المزدوج هو السمة المميزة للـ DNA ، ويلتف الشريطان على محور مشترك، ويكون قطر الحلزون 20 انكستروم (2×10^{-7} ملليمتر). وتركيب الحلزون المزدوج هو التركيب الثانوي لأشرطة DNA. وتكون هياكل DNA الناتجة عن تكامل القواعد على ثلاثة أنواع رئيسية هي:

- **النوع B :** وهو الحلزون الطبيعي ويكون ملتفا الى جهة اليمين وهناك 10-10.5 قاعدة نروجينية لكل لفة التي تلتوي بدرجة 360 °م ويكون البعد بينهما 3.4 نانومتر، ونتيجة الالتفاف يتكون أخدود صغير مفلطح وآخر كبير وعميق والأخير هو الذي يسهل دخول البروتينات التي ترتبط بالـ DNA ، اما الأخدود الصغير فيكون في اغلب الأحيان مأوى للقواعد الثايمين والأدينين كما موضح في الشكل 17.3 :



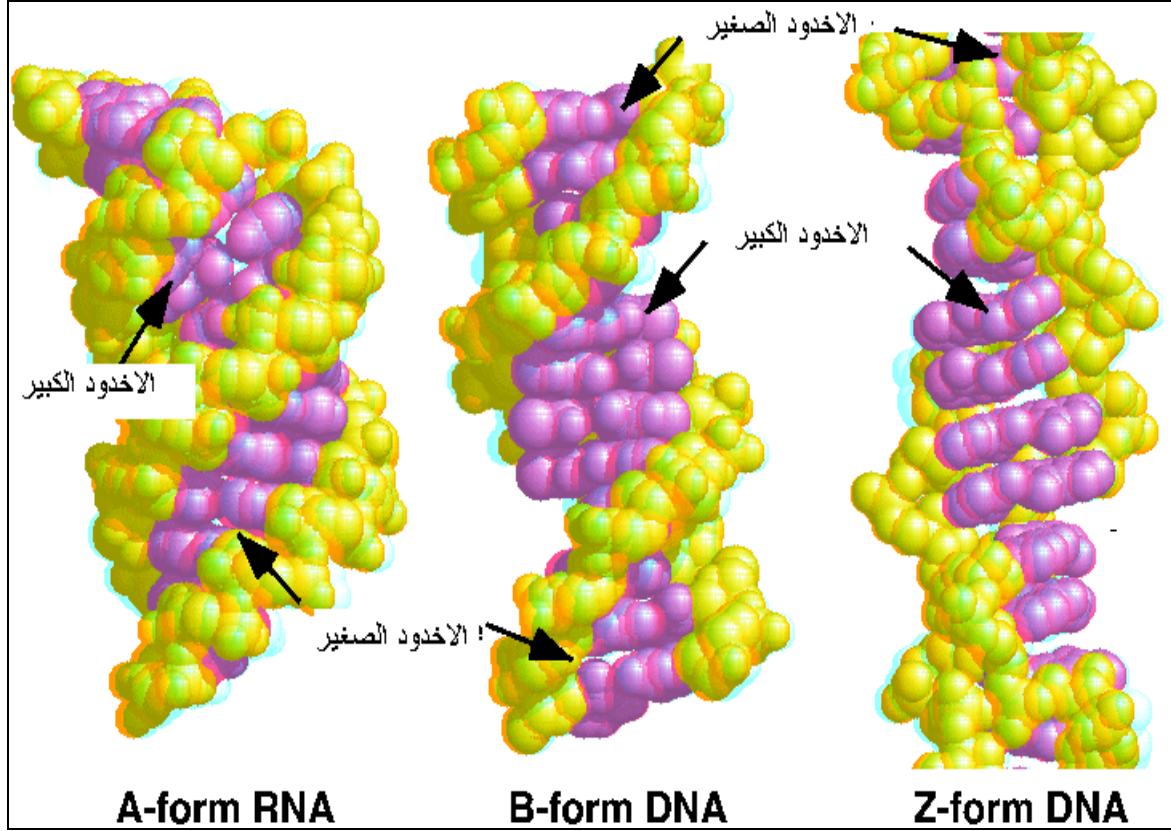
شكل 17.3 : هيئة B للـ DNA

- **الهيئة A :** إحدى هيئات DNA الطبيعية (نادرة الحدوث) التي يمكن ان توجد فيها هذه الجزيئة العملاقة، وتختلف في الأبعاد الثلاثية المجسمة عن النوع B وتظهر في الغالب عند حصول حالة الجفاف اي سحب الماء مثلاً عند تحضير النماذج للدراسة وكذلك يمكن ان تظهر عند تهجين أشربة DNA مع أشربة RNA. ويختلف عن الهيئة B بدرجة دورانه على المحور العمودي للحلزون.
 - **الهيئة Z-DNA** تمثل احد الهيئات او الأشكال التركيبية للـ DNA غير الثابتة من ناحية الترموديناميك Thermodynamic ويكون مبروم من اليسار الى اليمين بشكل خاص بحيث يؤدي الى ظهور الجزيئة بشكل متعرج ويؤدي الى توسيع المسافة الى 0.77 نانومتر. وتوجد هذه الهيئة مع هيئات DNA الأخرى وتكون القطع المتعرجة متصلة بالنوع B - DNA الموجودة في الجينوم.
- تنشأ هذه الهيئة في المناطق التي تكثر فيها تكررات مكونة من 2 من أزواج القواعد. وتختلف الأخاديد الصغيرة والكبيرة التي يكونها عن أخاديد الهيئات الأخرى.
- في العموم فان تكوين الهيئة المتعرجة تكون غير مفضلة في الخلايا ولكن تمليه بعض الظروف مثل وجود تكررات من أزواج القواعد ، او التقاف جزيئات DNA بشكل مخالف Negative DNA supercoiling فضلا عن ظروف أخرى مثل انخفاض تراكيز الأملاح في البيئة المحيطة بالمادة الوراثية او وجود بعض الايونات او بتأثير درجات الحرارة او تغير الأرقام الهيدروجينية في البيئة المباشرة للحوامض النووية. وعلى العموم فان هذه الهيئة كما تبين هي حالة عابرة وتنتج من التذبذب مع الهيئة B كما موضح في الشكل (18.3) . ومن الشكل يلاحظ ان الهيئة B هي الشكل المنضبط مقارنة بباقي الهيئات ما عدا ان قواعد الأشربة لا تقع على مستوى واحد ويمكن ان تلتوي بنمط يشابه Propeller-like ، ولذا فان جزيئة DNA يمكن ان تتحني دون ان يؤدي ذلك الى تغيرات أساسية في التراكيب الموضعية. اما في حالة الهيئة المتعرجة (Z) فان الأخدود المفرد يحوي على كثافة عالية من الجزيئات المشحونة بشحنة سالبة. وكما ذكر أنفا ان الهيئة المتعرجة يمكن ان تحدث في بعض مناطق الجسم من الهيئة B بعد حدوث استدارة للمناطق الغنية بـ GC بزواوية 180°م.
- وأهميته الحيوية يمكن ان تستنبط من الحالات التي يتكون فيها، فهو يتكون بعد بدء عملية بدء الاستنساخ مما يوفر نوعاً من الاسترخاء في المناطق المراد انتساخها والملتقة بشكل فائق وهذا ما أثبتته بعض التجارب التي أجريت على بعض كروموسومات الإنسان ، اذ ان له علاقة بعمليات تحويل الإنزيمات المحورة لإنزيمات كثر RNAPs . وبذا تكون له علاقة بمناطق الانتساخ الفعالة، كما انه يرتبط ببعض البروتينات عند الإصابة بالفيروسات لتوفير الحماية ضدها كما في حالة الإصابة بفيروسات الجدري (Poxviruses) وكذلك يظهر التركيب عند مثيلة بعض المناطق لغرض التنظيم. فعند حدوث مثيلة للسايتوزين عند الذرة C5 فتحول الهيئة الى الشكل المتعرج.



شكل 18.3 : التذبذب بين هيتي B و Z للـ DNA

وفي العموم ان لأشرطة DNA بأي كانت من الهيئات القابلة على الانحناء والالتواء والانضغاط وكل من هذه القابليات تؤدي دورا وظيفيا معيناً مثل حصول حالة تكوين الجزيئات الحلقية او ترتيب ارتباط البروتينات، وكل هذه تعتمد على قيم محددة معروفة في مجال كيمياء الحوامض النووية وهي ترتبط بشكل مباشر او غير مباشر بمفهوم الطاقة. والهيئات الثلاثة B، A و Z موضحة في الشكل 19.3:



شكل 19.3 : هياكل DNA

كمية DNA في الخلايا

وضعت بعض المؤشرات للتعريف بكمية DNA بالخلية ومنها **C-value**: وهي مقياس للـ DNA في الخلايا الفردانية Haploid في الأحياء مزدوجة المجاميع الكروموسومية Diploid وتقاس بالببيكوغرام Picogram/cell وتصل في اللبائن الى حوالي 2-3 بيكوغراما خلية ولكن هذه القيمة غير صالحة كمؤشر في هذه الحالة نظرا لاختلافها في الأنواع وتأثرها بالتعقيد التطوري للأحياء، وبذلك فهي لا تعطي اي معلومات يمكن الاستفادة منها على الأقل في الجوانب التطورية. فحجم الجينوم ليس له علاقة بتعقيد وتطور الكائن الحي (انظر موضوع الجينومات) فمثلا في بعض الابدائيات يكون جينومها اكبر من جينوم الإنسان وكذلك كائنات Salamanders تحوي على DNA يصل الى 40 ضعف من جينوم الإنسان ولكن حلت هذه المشكلة بكشف المناطق غير المشفرة في DNA ، ولذا فان حجم الجينوم لا يعكس عدد الجينات (كما سيأتي ذكره لاحقا).

انتظام وتراكيب DNA

يؤدي DNA وظيفتين مهمتين وهما:

- استمرار المحافظة على المعلومات الوراثية في الأحياء الناتجة وهذا يعني ان كل المعلومات اللازمة للحياة ووظائف الخلية مخزونة فيه.

• التفسير لإعطاء البروتينات اللازمة لإدامة الحياة. وينقل DNA المعلومات الوراثية ولكن بتأثير ظروف أخرى تقع ضمن مجال الوراثة اللاجينية. ويحوي DNA على مناطق متخصصة هي الجينات تمثل مناطق تحوي المعلومات الوراثية لإنتاج البروتينات وجينات أخرى لتنظيم الفعاليات على كافة المستويات وكل هذه المواد الوراثية تتضمن في تركيب الكروموسوم الذي يكاد يكون وحيدا في الخلايا بدائية النواة ويكون تراكيب متعددة تمثل الكروموسومات في الخلايا حقيقية النواة. ومنها جاءت فكرة التواليات المشفرة (CDs) اي التي تترجم محتوياتها الى بروتينات وتواليات او مناطق أخرى غير مشفرة . وتوجد تغايرات كبيرة في محتوى DNA من القواعد النتروجينية التي تؤثر في تركيب DNA فضلا عن تأثيرها في انتظام تراكيبه وبالتالي التأثير في وظيفته ، فهناك حقائق تشير الى وجود تراكيب عامة وأخرى خاصة تعتمد على التواليات وتغايرها فمحتوى DNA من AT تعتمد على الظروف البيئية مثل وجود الأملاح (كما مر ذكره مع Z-DNA) وكذلك يؤثر المحتوى في حالة الالتفاف الفائق وبالتالي التأثير في قابلية الانتساخ والتضاعف.

الشفرات الوراثية

توليفة من ثلاث قواعد نتروجينية تقع في المناطق المشفرة يمكن ان تترجم الى حامض اميني وتوجد 64 شفرة (³4) لذا تسمى الشفرة الثلاثية لتعطي 20 حامض أميني وثلاث شفرات وقف (UGA, UAG, UAA) وهذا يعني ان هناك بعض الحوامض الامينية التي يشفر لها بأكثر من شفرة وراثية، وهذه حقيقة ما عدا الحامض الاميني الميثايونين والتربتوفان يشفر لهم بشفرة واحدة اما باقي الشفرات التي تشفر لحامض أميني فتسمى بالشفرات المترادفة Synonymous codons ، في حين يطلق على الشفرات التي تشفر لحوامض مختلفة بالشفرات غير المترادفة non-synonymous. ومعظم الشفرات المترادفة تختلف في الموقع الثالث للقاعدة النتروجينية اي عند النهاية 3' كما موضح في الجدول (1.3) .

ووفقا للشفرات القياسية هناك شفرة بدء واحدة التي تبدأ بها تواليات المناطق المشفرة Start codon وهي شفرة الميثايونين (AUG) وهذه تمثل أحد المؤشرات المستعملة في تصميم البرامج للبحث عن الجينات المشفرة للتواليات غير المعروفة.

جدول 1.3 : الشفرات الوراثية والحوامض الامينية التي تشفر لها

القاعدة الثانية في الشفرة الوراثية

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

القاعدة الاولى في الشفرة الوراثية

القاعدة الثالثة في الشفرة الوراثية

الشفرات الوراثية

استعمال الشفرات Codon usage

كما ذكر آنفاً ان هناك وفرة في الشفرات الوراثية لذلك هناك عدداً من الحوامض الامينية التي يشفر لها بشفرتين او ثلاثة او أربعة فضلاً عن وجود حوامض يشفر لها بست شفرات وراثية كما موضح في الجدول أعلاه ويلاحظ بالنسبة للحوامض التي يشفر لها بأكثر من شفرة انها تشترك في القاعدتين الأوليتين وتختلف في القاعدة الثالثة.

وتقرأ الشفرات من قبل tRNA الحاوي على مقابل الشفرة Anticodon او الشفرة المضادة (ضمن أساس تكامل القواعد النيتروجينية المار الذكر) وهناك جزيئة واحدة من tRNA هي التي تقرأ الشفرة الأصلية المقابلة لها وتسمى الجزيئات الملائمة Cognate tRNA ، اما جزيئات الأخير التي تقرأ الشفرات المترادفة فتسمى النظائر Isoacceptor tRNA والتي يقع عملها ضمن مبدأ المراوغة Wobbling hypothesis ، وهذا يعني ان الجزيئة الواحدة من tRNA يمكن ان تميز شفرات مختلفة.

ان الشفرات الوراثية القياسية المذكورة في الجدول أعلاه تستعمل من قبل معظم البكتريا والاراكيا وفي أنوية الخلايا حقيقية النواة. وعلى العموم فان صفة القياسية لا تكون ثابتة في كل أشكال الحياة المذكورة فمثلا في *Mycoplasma* و *Spiroplasma* تستعمل UGA للتربوفان في حين انها شفرة وقف بالنسبة للقياس العام، وكذلك الحال بالنسبة للهدبيات Ciliata وبعض

الخمائر فهي تستعمل بعض شفرات الوقف للتشفير للكلوتامين. اما في مايتوكوندريا الخلايا حقيقية النواة التي لها جينومها المستقل فهي تحوي على بدائل للشفرات الوراثية وتخلق tRNA الخاصة بها. ومن الجدير بالذكر ان الحامض السستئين السليبي Selenocysteine يعد الحامض الاميني الواحد والعشرون ويشفر بـ UGA ويوجد في بعض البكتريا وعدد من الأحياء حقيقية النواة.

وكان يعتقد سابقا ان الشفرات المترادفة للحامض الاميني تحدث بشكل عشوائي في الجينوم نظرا لانها في النهاية تعطي الحامض الاميني نفسه وهذا يعني ان كل الجينومات لها الفرصة نفسها في تكرار استخدام الشفرات المترادفة، ولكن بعد زيادة الدراسات وكثرة التواليات المسجلة من نهاية سبعينات القرن الماضي الى بداية الثمانينات تبين ان استعمال الشفرات ليس عشوائيا وان بعض الجينومات تفضل استعمال بعض الشفرات للحوامض الامينية وسمي هذا بانحياز استعمال الشفرات Codon usage bias وهذا بطبيعة الحال يؤثر في تصميم البودائ فيما اذا استعملت تواليات البروتين، لذلك فعند اختيار بروتين لتصميم البودائ يفضل اختيار بروتين يكون بأقل ما يمكن من تشتت الشفرات Degeneracy.

اما تفضيل استعمال شفرات معينة للحوامض الامينية فله عدة أسباب منها إتباع محتوى الجينوم العام، فمثلا *E. coli* محتوى جينومها من GC يصل الى 50% التي تعد النموذج الدراسي والموضحة بعض الحقائق عن استعمالها للشفرات في الجدول 2.3:

جدول 2.3 استعمال الشفرات في جينات <i>E. coli</i>															
	الشفرة	الحامض الاميني	%	النسبة	الشفرة	الحامض الاميني	%	النسبة	Codon	الحامض الاميني	%	النسبة	الشفرة	الحامض الاميني	%
U	UUU	Phe (F)	1.9	0.51	UCU	Ser (S)	1.1	0.19	UAU	Tyr (Y)	1.6	0.53	UGU	Cys (C)	0.4
	UUC	Phe (F)	1.8	0.49	UCC	Ser (S)	1.0	0.17	UAC	Tyr (Y)	1.4	0.47	UGC	Cys (C)	0.6
	UUA	Leu (L)	1.0	0.11	UCA	Ser (S)	0.7	0.12	UAA	STOP	0.2	0.62	UGA	STOP	0.1
	UUG	Leu (L)	1.1	0.11	UCG	Ser (S)	0.8	0.13	UAG	STOP	0.03	0.09	UGG	Trp (W)	1.4
C	CUU	Leu (L)	1.0	0.10	CCU	Pro (P)	0.7	0.16	CAU	His (H)	1.2	0.52	CGU	Arg (R)	2.4
	CUC	Leu (L)	0.9	0.10	CCC	Pro (P)	0.4	0.10	CAC	His (H)	1.1	0.48	CGC	Arg (R)	2.2
	CUA	Leu (L)	0.3	0.03	CCA	Pro (P)	0.8	0.20	CAA	Gln (Q)	1.3	0.31	CGA	Arg (R)	0.3
	CUG	Leu (L)	5.2	0.55	CCG	Pro (P)	2.4	0.55	CAG	Gln (Q)	2.9	0.69	CGG	Arg (R)	0.5
A	AUU	Ile (I)	2.7	0.47	ACU	Thr (T)	1.2	0.21	AAU	Asn (N)	1.6	0.39	AGU	Ser (S)	0.7
	AUC	Ile (I)	2.7	0.46	ACC	Thr (T)	2.4	0.43	AAC	Asn (N)	2.6	0.61	AGC	Ser (S)	1.5
	AUA	Ile (I)	0.4	0.07	ACA	Thr (T)	0.1	0.30	AAA	Lys (K)	3.8	0.76	AGA	Arg (R)	0.2
	AUG	Met (M)	2.6	1.00	ACG	Thr (T)	1.3	0.23	AAG	Lys (K)	1.2	0.24	AGG	Arg (R)	0.2
G	GUU	Val (V)	2.0	0.29	GCU	Ala (A)	1.8	0.19	GAU	Asp (D)	3.3	0.59	GGU	Gly (G)	2.8
	GUC	Val (V)	1.4	0.20	GCC	Ala (A)	2.3	0.25	GAC	Asp (D)	2.3	0.41	GGC	Gly (G)	3.0
	GUA	Val (V)	1.2	0.17	GCA	Ala (A)	2.1	0.22	GAA	Glu (E)	4.4	0.70	GGA	Gly (G)	0.7
	GUG	Val (V)	2.4	0.34	GCG	Ala (A)	3.2	0.34	GAG	Glu (E)	1.9	0.30	GGG	Gly (G)	0.9
	U				C				A				G		

% * تمثل معدل التردد لكل 100 شفرة
 ** وفرة الشفرة نسبة لكل الشفرات الخاصة بالحامض الاميني ، فهي بالنسبة للحامض
 الاميني التريوفان او السيتاوين تكون 1

وهذا صحيح بالنسبة لمعظم البكتريات وعليه فاتباعا للمعدل فإنها تميل الى استعمال الشفرات التي تكون فيها القاعدة الثالثة من G او C ولكن هذا لا يكون عاما وشاملا نظرا لوجود التباير بين الجينات حتى في الجينوم الواحد، فضلا عن ان العديد من الجينومات الأخرى تحوي على مناطق غير مشفرة تحوي على GC. ومن جهة ثانية قد يكون للطفرات أثر في انحياز استعمال الطفرات.

وقد تكون عملية كفاءة الترجمة هي المؤثر الأكبر اذ ان الكفاءة العالية تعني المعدل العالي من تخليق البروتينات والمعدل الواطئ من الأخطاء في الوقت نفسه، والكفاءة هذه تتأثر بشكل كبير بوجود جزيئات tRNA الملائمة والجزيئات المناظرة فبعض الأحياء تحوي على كميات كبيرة من جزيئات tRNA المناسبة الخاصة بحامض اميني محدد وهذا يعني وجود الشفرات الملائمة لها ويؤدي الى تخلي الخلايا عن القوة الدافعة للطفرات المؤدية الى التغيير، وهذا ما يحصل في الفقريات الراقية او البكتريا التي تعيش بطريقة متطفلة داخل خلايا المضيف وكذلك بالنسبة للعضيات الخلوية مثل المايكوبلازما والبلاستيدات الخضراء وغيرها.

الجينات وأنواعها

يعد الجين الوحدة الوراثية في الكائنات الحية ويمثل امتداد من DNA (او RNA في بعض الفيروسات) التي تشفر للبروتين او جزيئات من RNA التي لها وظائف خاصة ، وهي الوحدات التي تمرر الصفات الوراثية من الآباء الى الأبناء. والجينات الموصوفة بالكلام العام هي نسخ او صور Alleles للجين والتي تؤدي الى ظهور صفات متفاوتة في الأحياء وتختلف في توالي القواعد النيتروجينية فيها (كما سيأتي ذكره).

تصنف الجينات الى عدة أصناف اعتمادا على المؤشرات المستعملة، فهناك جينات تركيبية تقوم بإنتاج البروتينات التي تدخل في بناء التراكيب الخلوية وأخرى تنظيمية تنتج البروتينات التي تقوم بتنظيم الفعالية الحيوية. وتصنف آخر يعتمد على نوع التوالي فالحاوية على الانترونات يطلق عليها المتقطعة Interpreted genes او الجينات الموزائكية ، اما الخالية فتسمى الجينات المستمرة او المعالجة Processed genes اي انها لا تحوي أصلا على الانترونات كما في معظم جينات الأحياء بدائية النواة او تكون قد أزيلت منها الانترونات. ويمكن ان تقسم الى عوائل كما هو الحال مع العوائل البروتينية ، فالعائلة الجينية هي التي تضم مجموعة من الجينات التي فيها تشابه كبير في توالي القواعد او البروتينات، وتظهر فعاليات حيوية متشابهة تقريبا وهذه يمكن ان تنشأ من تضاعف الجين ويمكن الكشف عنها بالتقنيات الخاصة بالعلاقات التطورية وبرامج حاسوب خاصة مقرونة بطرق وموديلات إحصائية.

وفضلا عن ذلك فيمكن ان تقسم وفقا لأهميتها مثل الجينات الأساسية وهي التي تحتاجها الخلايا للوظائف الرئيسية ولذلك يكون موقعها في الغالب قرب مناطق بدء التضاعف في البكتريا لضمان زيادة التعبير عنها وثبوتها، والجينات غير الأساسية وتحتاجها الخلايا تحت ظروف خاصة مثل ظروف الإجهاد او غيرها.

وفضلا عما ذكر أعلاه فهناك الجينات الكاذبة Pseudogenes وتعني الجينات المشابهة للجينات العادية في العديد من الصفات الا انها فقدت وظيفتها لسبب ما مثل حصول طفرة أدت الى تحويل شفرة معينة الى شفرة وقف او فقد أجزاء خاصة بعملية الانتساخ وهذه تسمى الجينات المعطلة Disabled genes. او انها تنشأ من الانتساخ العكسي لبعض أجزاء نسخ mRNA

وتعود لتستقر في الجينوم مرة أخرى وتسمى الجينات المعالجة وهذا يعني ان لها صفات مماثلة للـ cDNA ، وتكون قد فقد مميزاتا او غيرها من العناصر التي تؤهلها للعمل كجينات طبيعية ، او يمكن ان تنشأ من تضاعف الجين. وعلى العموم يميل البعض لتسميتها بالمادة الوراثية السقط Junk DNA ولكن كشفت الدراسات الحديثة ان الجينات الكاذبة تقوم بعدد من الوظائف لعل أهمها عمليات التنظيم والتعبير الجيني مثل تشفيرها لبعض جزيئات RNA المنظمة siRNA (التي سيأتي ذكرها في موقع آخر). فضلا عن مشاركتها في توليد الأمراض.

تغاير DNA

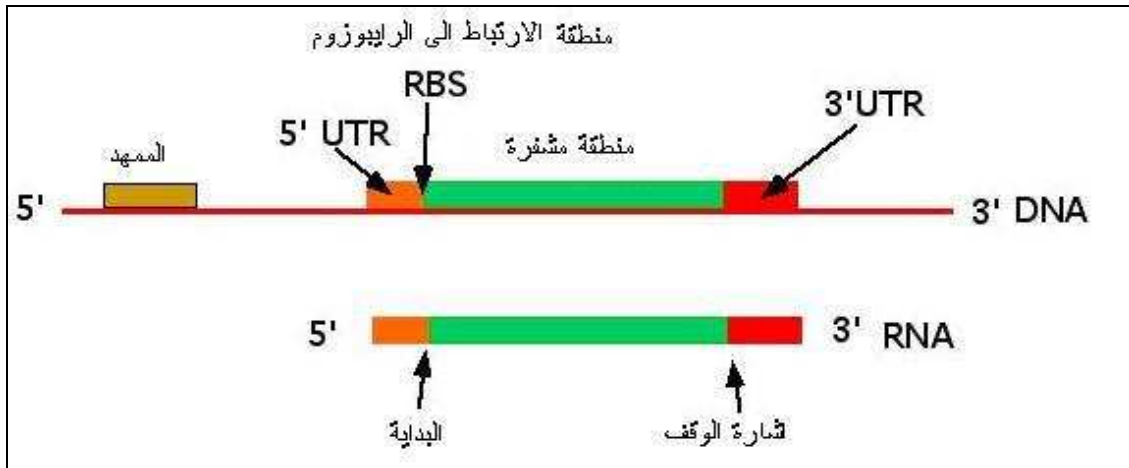
يشمل تغاير DNA (DNA polymorphism) مستويات مختلفة منها تغاير في الجينات والتي عبر عنها بالمتشابهات (الآليلات Alleles) وكذلك مستوى تركيب الكروموسوم مثل الحجم وموقع المريكزات ، ووجود الكروماتين المتباين Heterochromatin والكروماتين الحقيقي Euchromatin. ويشمل التغاير ايضا منتجات الجين يعني التغير في البروتينات الناتجة من حيث الألفة والفعالية وغيرها، وكذلك النمط المظهري. وكل هذه جواهرها التغيرات الحاصلة في توالي القواعد النتروجينية، وتعد من الواسمات المهمة للتعرف والتمييز للعديد من المتعلقات الوراثية، ومن أفضل الأمثلة عليها تغيرات النيوكلوไทيد الواحد SNPs التي سيأتي ذكرها لاحقاً.

تشريح الجين

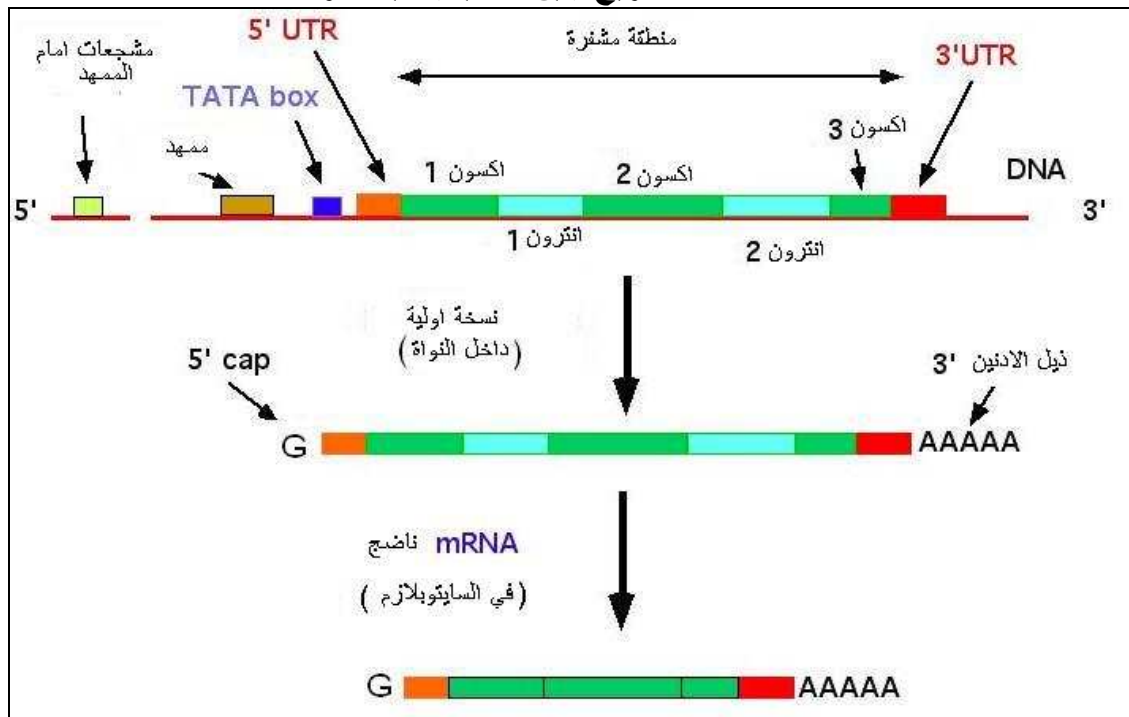
مثلما ذكر في موقع آخر ان الجين المشفر هو المسئول عن إنتاج بروتين فعال وهذا يعني انه يحوي على عدد من الشفرات التي تعطي الحوامض الامينية، ويشمل تواليات من DNA تسيطر وتنظم التعبير عنه وقد تكون هذه التواليات المنظمة تبعد بالآلاف القواعد النتروجينية عن المنطقة المشفرة في خلايا بدائية النواة وقد يصل بعد هذه المناطق مثل المشجعات Enhancers الى 50 كيلو قاعدة في بعض جينات الخلايا حقيقية النواة.

وهذه السيطرات قد تكون موجبة (وهو ارتباط بروتين او معقد بروتيني الى الممهد حتى يفتحه والا يكون مغلقا) او تكون سيطرات سالبة (وهي كون الجين مفتوح التعبير ويغلق عند ارتباط بروتين او مادة أخرى) . فضلا عن بعض مناطق السيطرة المذكورة آنفاً هناك تواليات أخرى تحيط بالجين وهي مناطق لا تترجم (Untranslated regions (UTRs وتكون موجودة عند الطرف 5' و 3' وهذه المناطق تؤثر في التعبير الجيني مثل التحكم بالعمر النصفى Half-life لجزيئات mRNA الناتجة وتكون مجاورة للاكسون الأول (5') والاكسون الأخير (3'). والشكل (20.3) يوضح تشريح بسيط لجين في الخلايا بدائية النواة .

جينات الخلايا حقيقية النواة هي أكثر تعقيدا اذ تحوي على الانترونات (مناطق غير مشفرة) والتي عند حدوث بعض الطفرات فيها يمكن ان يؤدي الى الأمراض ومناطق الاكسونات وهي المناطق المشفرة وحدث الطفرات فيها يمكن ان يؤدي الى الأمراض ايضا، فضلا عن وجود مناطق منظمة أخرى والشكل التالي (21.3) يوضح تشريح الجين وبعض العمليات التي تجري عليه أثناء التعبير الجيني.



شكل 20:3 : تشرح جين لخلايا بدائية النواة



شكل 21.3 : تشرح جين في خلايا حقيقية النواة وبعض المعالجات التي تجري عليها

وتمثل المناطق غير المشفرة UTRs مناطق مهمة جدا، وهي تفوق في ثبوتها المناطق المشفرة، وتساهم في عملية التنظيم اذ تضم بعض المناطق التي تكون مواقعاً لارتباط بعض البروتينات وجزئيات RNA النقيضة Antisense RNA ، وتساهم في تنظيم نقل RNA بين النواة والسايكوبلازم فضلا عن تحديد مواقع mRNA والمساهمة في الترجمة وثبوت mRNA. والمناطق 5' UTR تحوي على العديد من مواقع الارتباط لعدد من معقدات الانتساخ وتساهم في جلب الوحدة 40S من وحدات الرايبوزوم وربطها بتركيب الرايبوزوم لبدء عملية الترجمة.

ان لطول منطقة 5' UTR ووجود عناصر أخرى الى يسار شفرة البدء تكون مهمة في تنظيم انتساخ وتخليق mRNA. وقد وجد ان إحاطة شفرة البدء بتواليات ملائمة وقصر منطقة 5' UTRs يشجع عمليات الانتساخ في حين ان طول المنطقة يؤدي الى ضعف الانتساخ. ومن التواليات الأخرى الموجودة في الطرف 5' هي تواليات Internal ribosome entry sites (IRESs) وجدت في جزيئات mRNA التي تشفر للبروتينات المنظمة.

اما المنطقة 3' UTRs فهي تلعب دورا مهما في فلق النسخ أثناء عمليات الخياطة وكذلك في إضافة ذيل الأدنين والتصدير من النواة وتنظيم الانتساخ وثبوت النسخ. ويمكن ان تحوي على تواليات تعمل في التنظيم السالب الذي يحصل بعد الانتساخ، كما ان إحباط ترجمة mRNA بواسطة الجزيئات الصغيرة من RNA (sRNA) يكون بازدواج الجزيئات الأخيرة مع قطيقات في 3'-UTRs. فضلا عن ان عمليات التفكك والانقلاب لجزيئات mRNA تتم بعناصر Cis-acting AU-rich توجد في المناطق الثابتة من 3'-UTRs لعدد من جزيئات mRNA التي تشفر للبروتينات المنظمة. وبالإضافة الى القطيقات ذات التأثير السلبي فان المناطق تحوي على مواقع لارتباط عدد من العوامل المشاركة في إنهاء الترجمة وإطلاق الببتيدات المخلفة.

وتعد التراكيب الثانوية في 3'-UTRs مهمة في حالات خاصة مثل في mRNA الخاص بعائلة البروتينات السيلينية، فكل البروتينات السيلينية المعروفة لحد الآن تحوي على ثمالة السستئين الحاوية على السلينيوم التي يشفر لها بشفرة الوقف UGA وان اندماج الحامض الاميني الحاوي على السلينيوم في الببتيد النامي يعتمد على تركيب ثابت Stem-loop في mRNA الذي يتكون من اندماج هذه الثمالة مع التوالي Selenocysteine insertion sequence (SECIS) التي تكون ضرورية لفك شفرة UGA لتشفر للسستئين السيليني بدلا من كونها شفرة وقف.

والمناطق الجينومية الخاصة بـ UTRs للمRNA يمكن ان تحوي على الانترونات مما يساعد في تكوين UTRs بديلة. والانترونات تميل ان تكون أكثر في 5'-UTRs بالرغم من ان 3'-UTRs تكون أطول من 5'-UTRs في بعض الحالات، وهذه التغيرات وغيرها تختلف في الأنسجة ومرحلة التطور وفي النهاية تؤثر في التعبير الجيني.

طبيعة الجينات في الأحياء

المراد هنا احتواء الجينات على المناطق المشفرة وغير المشفرة. ففي معظم جينات الخلايا بدائية النواة تكون كل مناطق الجين مشفرة اي بدون وجود الانترونات كمناطق اعتراضية وأكثر من ذلك ان بعض الجينات المشتركة (في الغالب) في مسار ابيض واحد تتجمع في عنقود يضم معه ممد واحد وعناصر تنظيم بنسخ واحدة لتكوين الاوبرونات وتكون النسخ الناتجة من انتساخها ذات رسائل متعددة تجري معالجتها فيما بعد.

اما الجينات في الخلايا حقيقية النواة فهي تكون حاوية على الاكسونات والانترونات وتكون هناك علاقة واضحة (طردية) بين عدد الانترونات الكلي مع درجة التعقيد في الكائن. وترقم كل من الاكسونات والانترونات بالاتجاه من 5' الى 3' في الشريط المشفر او الحساس .

الانترونات: وهذه لها مواقع مختلفة.

- اما ان تكون بين الشفرات الوراثية ويطلق عليها النوع الصفري Phase 0
- او تكون داخل الشفرة بعد القاعدة النتروجينية الأولى Phase 1

• أو تكون داخل الشفرة بعد القاعدة الثانية Phase 2

ويختلف نمط توزيع هذه الانترونات فالنوع الصفري يزداد بالاتجاه من 5' الى 3' في الجينات وفي الوقت نفسه يقل النوع الأول وهذا ما مسجل في الإنسان وفي رشاد الصخر (النموذج الدراسي النباتي) وأحياء أخرى، ووفق النمط المذكور يمكن ان تنتج تسع من تشكيلات الانترونات ثلاث منها متناظرة (0,0; 1,1; 2,2) وستة غير متناظرة (1,2; 2,1; 2,0; 0,2; 0,1; 1,0) ويكون النوع الصفري هو الشائع والنوع 2,2 نادر جدا.

ويختلف عدد الانترونات في الجين الواحد وكذلك توزيعها، وقد وجد ان توزيعها العام 2(phase 2): 3(phase 1): 5(phase 0). مع وجود بعض الشواذ. وقد وجدت انه ربما كانت عدة أسباب وراء هذا التوزيع منها إدلاف الاكسونات Exon shuffling وثبوت إشارة فلق الاكسونات عن الانترونات او انزلاق الانترونات وربما خاصية البروتين الناتج، كما ان تفضيل استعمال الشفرات له دخل في توزيع الانترونات في الجين والذي يظهر على شكل متشابه في بعض الأنواع وهذه توضح الانحياز في توزيع الانترونات والتي تؤثر في تناظر الانترونات. فالانترونات المتناظرة هي الأكثر في جينات الخلايا حقيقية النواة والانترونات غير المتناظرة هي الأقل. وأشارت الدراسات الى ان النوع الأول يكون أكثر في النهاية 5' في حين يكون النوع الصفري أكثر في النهاية 3' الانترونات غير المتناظرة خاصة في جينوم الإنسان ورشاد الصخر، فالنهاية 5' من الجين في الإنسان تكون مشبعة بالنوع الأول من الانترونات وتقل بالاتجاه من 5' الى 3' ومقابله يزداد النوع الصفري بشكل متدرج وبالتالي يؤثر في توزيع الاكسونات المتناظرة وغير المتناظرة.

ومن جهة ثانية فان وجود النوع الأول من الانترونات في النهاية 5' للمناطق المشفرة CDS يظهر في الجينات المشفرة للبروتينات الحاوية على ببتيدات الإشارة. وهناك قبول لفكرة ان الانترونات القديمة تكون في النهاية 5' والانترونات الجديدة تكون في النهاية 3' في اغلب الجينومات المدروسة، وهذا آتياً من فقدان او اكتساب الانترونات التي تكون عملية مميزة لحالة التطور في الجينات.

ويمكن ان تقسم الانترونات الى عدة مجاميع وفقاً لمواقعها ومنها:

- انترونات خاصة في جينات tRNA وتكون هذه خاصة لأنها تزال بإنزيمات غير التي يقطع بها mRNA اذ تجرى عملية فسفرة للبروتينات المشتركة ثم يلحم طرف جزيئة tRNA بإنزيمات أخرى.
- انترونات المجموعة الأولى شائعة الانتشار في tRNA و rRNA وتوجد في الماييتوكوندريا والبلاستيدات وكذلك في الجينوم النووي للأحياء اللاقارية وبعض العاثيات البكتيرية وبعض جينومات البكتيريا الحقيقية Eubacteria وهذه المجموعة يمكن لبعضها ان يستأصل دون مشاركة البروتينات مما يشير الى الفعالية الإنزيمية لبعض جزيئات RNA.
- انترونات المجموعة الثانية توجد في الماييتوكوندريا والبلاستيدات الخضر وكذلك بعض أنواع البكتيريا المزرقة Cyanobacteria وبعض Proteobacteria. يمكن لبعضها الانفلاق ذاتياً وتشارك بعض البروتينات في عمليات خياطها كما هو الحال مع

المجموعة الأولى اذ تكون خياطها مشابه للمجموعة الأولى، لذلك يعتقد ان انترونات المجموعة الأولى وانترونات المجموعة الثانية تكون ذات علاقة تطورية، ويحتمل ان snRNA (small nuclear RNA) قد نشأت من هذه المجموعة.

- المجموعة الأخرى هي انترونات النسخ الأولية للـ RNA (pre-mRNA) ويطلق عليها جزيئات الخياطة Spliceosomal RNA الموجودة في النواة.

ومن جهة ثانية يمكن ملاحظة ان الانترونات الموجودة في الجينات العالية التعبير تكون أقصر من تلك الموجودة في الجينات الواطئة التعبير، وهذا الاختلاف يكون كبيرا في الإنسان ، حيث يكون طول الانترونات (كمعدل) 14 مرة أقصر في الجينات عالية التعبير مقارنة بتلك التي يعبر عنها ببطيء.

وظيفة الانترونات: تقوم الانترونات بأكثر من وظيفة:

- 1- تنظيم التعبير الجيني للجين الذي توجد فيه.
 - 2- إنشاء جينات جديدة وذلك بجمع الاكسونات بتشكيلات مختلفة أثناء عمليات إدلاف الاكسونات والتي يوجد العديد منها في جينومات اللبائن.
 - 3- مشاركتها في تكوين الجسيمات النووية وتنظيم الكروماتين وهي أكثر فعالية في هذا المجال من الاكسونات وبعض التواليات المكررة مثل Alu sequence
 - 4- تكوين تراكيب الأرضية او المنصات Scaffold/ matrix attachment regions (S/MARS) التي يعتقد انها ترسي عروات الكروماتين الى الأرضية النووية.
- أصل الانترونات:** هناك فرضيتان حول وجود الانترونات في الجينات

- نظرية او فرضية الاكسون: وتشير الى ان الانترونات موجودة في جينات الأوليات او الأسلاف التي هي أصل كل أنواع الحياة. وهذه الانترونات قد تكون من أنواع التي تفلق ذاتيا ، وبعد تشعب الأحياء الى ممالك الأحياء الأساسية فقدت هذه الانترونات من البكتريا الحقيقية والبكتريا القديمة (الاركيا).
- النظرية الثانية تشير الى ان الانترونات هي نوع من أنواع العناصر القافزة التي تستطيع القفز من وإلى الجينات الموجودة التي ليس من الضروري ان تحوي على الانترونات أصلا. ولذلك يعتقد ان انترونات المجموعة الثانية غزت الجينات النووية بعد حصول عمليات التعايش الحيوي والتي أسفرت في النهاية الى وجود المايكوبلازما والبلاستيدات الخضراء في الخلايا حقيقية النواة، وهذا الافتراض جاء من ان انترونات المجموعة الثانية يمكن ان تفقر الى الجينات في مواقع محددة.

الاكسونات Exons

تعد المناطق المشفرة في الجينات. وفي العادة تكون شفرة البدء في بداية الاكسون الأول وشفرة الوقف في نهاية الاكسون الأخير. وتختلف نسب الاكسونات في الجين الواحد فمثلا في:

الإنسان هناك حوالي 9.08 اجين

ذبابة الفاكهة حوالي 4.45 اجين

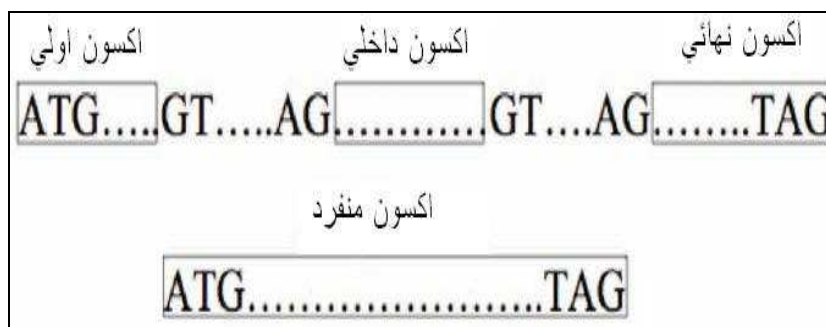
الدودة *C. elegans* حوالي 6.71 اجين

نبات رشاد الصخر حوالي 6.63 اجين وفي الإنسان تكون الاكسونات مفرقة ويتراوح عددها من 9 ويمكن ان يصل الى 43.

أنواع الاكسونات

تقسم الاكسونات الى أنواع اعتمادا على مواقعها في الجين:

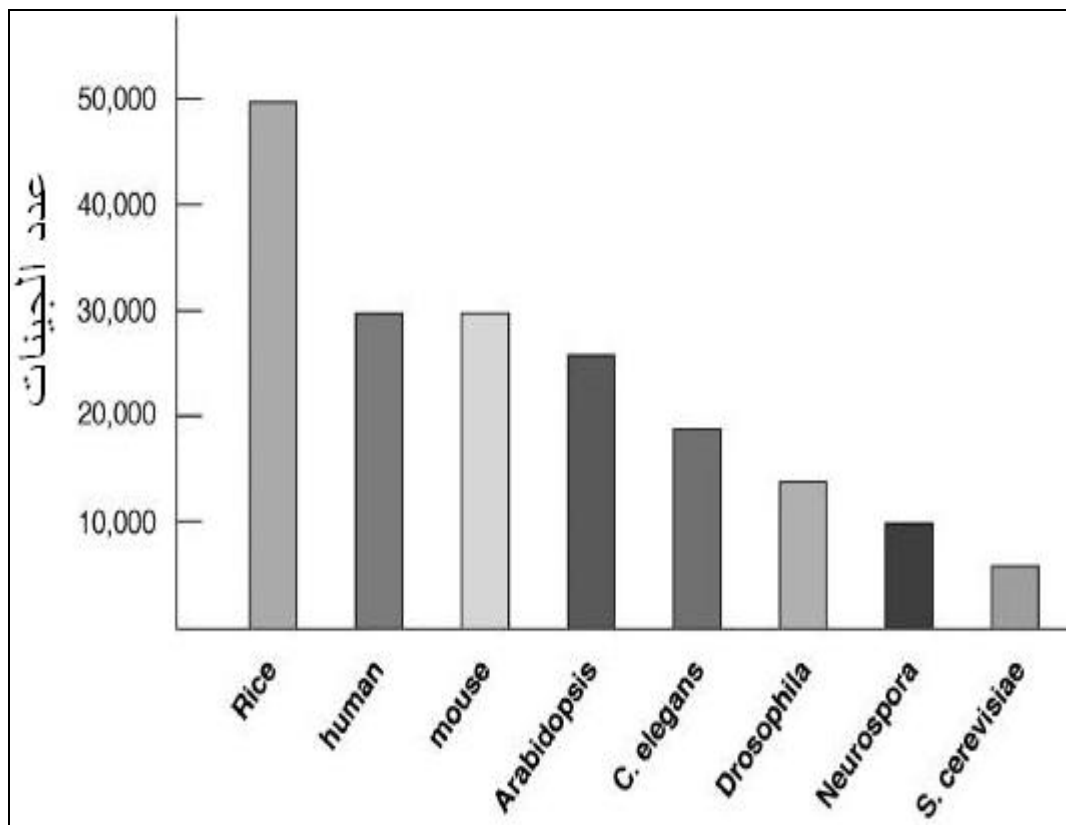
- اكسونات البداية Initial exons وتمتد هذه من شفرة البدء الى أول موقع لفلق وخياطة الانترون الذي يليه والمسمى بالموقع الواهب .
- الاكسونات الداخلية Internal exons ويبدأ من موقع استقبال الى موقع الواهب الذي يليه.
- الاكسونات النهائية Final exons ويمتد من آخر موقع استقبال الى شفرة الوقف للجين.
- الاكسونات المفردة Single exons وهذه توجد في الجينات الخالية من الانترونات وتبدأ من شفرة البدء الى شفرة الوقف وهذه الأنواع موضحة في الشكل 22.3:



شكل 22.3 : أنواع الاكسونات

أعداد الجينات

من أهم مهام تهميش الجينوم هو محاولة إيجاد العدد الدقيق لعدد الجينات الكلي في الجينوم، وهذه قد تكون سهلة وواضحة بالنسبة للأحياء بدائية النواة نظرا لبساطة تراكيبها الجينية ولكن في الأحياء حقيقية النواة وخاصة الإنسان فان المسألة تكون موضوع نقاش وذلك نظرا للتركيب المعقد لجينوماتها التي تشوش على عمليات التنبؤ بالجينات. وقبل اكتمال تحديد تواليات الجينوم البشري كان يعتقد ان هناك بين 20,000 – 120,000 جين ولكن بعد تحديد التوالي وباستعمال برامج معقدة لإيجاد الجينات هبط العدد الى 25,000 – 30,000 وان كان الرقم لا يزال غير متفق عليه بين جميع الباحثين ولكن الاتفاق العام ان العدد الكلي لجينات الجينوم البشري لا يزيد عن 30,000 والشكل التالي (23.3) يوضح أعداد الجينات لعدد من الكائنات.



شكل 23.3 : أعداد الجينات لأحياء مختلفة

وبهذه المستويات وغيرها تظهر درجة التعقيد على الكائن. وفي العموم فإن أعداد الجينات في الجينومات يفضل أن يعبر عنه بالكثافة الجينية والتي موضح بعضها في الجدول 3.3

جدول 3.3 : الكثافة الجينية لبعض الأحياء

الكثافة الجينية	عدد الجينات	حجم الجينوم (زوج قاعدة)	مجموعة الأحياء/البكتيريا
831	479	5.8×10^5	<i>Mycoplasma genitalium</i>
826	677	8.2×10^5	<i>Mycoplasma pneumoniae</i>
937	853	9.1×10^5	<i>Borrelia burgdorferi</i>
946	1041	1.1×10^6	<i>Treponema pallidum</i>
945	1512	1.6×10^6	<i>Aquifex aeolicus</i>
935	1590	1.7×10^6	<i>Helicobacter pylori</i>
959	1727	1.8×10^6	<i>Haemophilus influenzae</i>
976	4100	4.2×10^6	<i>Bacillus subtilis</i>
892	3924	4.4×10^6	<i>Mycobacterium tuberculosis</i>
932	4288	4.6×10^6	<i>Escherichia coli</i>

880	3168	3.6×10^6	البكتريا الم <i>Synechocystis</i> PCC6803
1022	1738	1.7×10^6	الارکيا <i>Methanococcus jannaschii</i>
1039	1871	1.8×10^6	<i>Methanobacterium thermoautotrophicum</i>
1107	2436	2.2×10^6	<i>Archaeoglobus fulgidus</i>
446	5800	1.3×10^7	الفطريات <i>Saccharomyces cerevisiae</i>
~ 175	~ 21 000	1.2×10^8	النباتات <i>Arabidopsis thaliana</i>
~ 140	~ 14 000	1.0×10^8	لا فقریات <i>Caenorhabditis elegans</i>
~ 71	~ 12 000	1.7×10^8	<i>Drosophila melanogaster</i>
~ 97	~ 15 500	1.6×10^8	الحبليات <i>Ciona intestinalis</i>
~ 150	~ 60 000	4.0×10^8	الفقریات <i>Fugu rubripes</i>
~ 20	~ 60 000	3.0×10^9	الفيران <i>Mus musculus</i>
~ 20	~ 60 000	3.0×10^9	<i>Homo sapiens</i>

وهي تمثل علاقة ضعيفة في حقیقات النواة ، في حين تكون العلاقة وثيقة في البكتريا نظرا لغياب الانترونات وغيرها من المواد الوراثية غير المستعملة ، والملاحظ ان الكثافة الجينية للإنسان هي الأقل وتقرّب من 20 والعدد القليل لجينات الإنسان الذي يقدر بخمسة أضعاف جينات خميرة الخبز وأقل من جينات نبات الرز والمقاربة للفيران لا يتفق مع درجة التعقيد للأحياء، وهنا يبدو ان التعقيد لا يتعلق بعدد الجينات وانما هناك جوانب أخرى لظهور التعقيد وضمن مستويات أخرى منها:

- الخياطة البديلة
- التعبير الجيني
- التنظيم
- التعبير البروتيني
- التحويرات
- التداخل بين الجزيئات

الطفرات

تغير في تواليات DNA على مستوى ضيق أو مدى واسع وذلك لأن الجينات وحدات قابلة للتغيير وقد أصبحت الحجر الأساس في الدراسات الوراثية المبكرة وهناك بعض الحقائق العامة حول الطفرات ومنها:

- الطفرات في Germ cells في الأحياء الراقية تؤدي إلى التطور.
- بعض الطفرات تؤثر في جين واحد كما في طفرات الخلايا الحقيقية النواة المؤثرة في وحدة انتساخ واحدة وأخرى تؤثر في اثنين من الجينات ذات العلاقة وأخرى تؤثر في عدة جينات Pleiotropic خاصة إذا كانت في التواليات المنظمة كما في الخلايا بدائية النواة عندما تقع الجينات في اوبرون واحد.
- الطفرات في الاكسونات أو الانترونات تؤدي في بعض الأحيان إلى الأمراض الوراثية.
- حدوث الطفرات يمكن أن يؤثر في عمليات التضاعف والانتساخ .
- الطفرات الواقعة في المناطق غير المشفرة يمكن أن يكون لها تأثيرا "كبيرا" فمثلا في الخلايا حقيقية النواة تسيطر بعض هذه المناطق في تخصيص مواقع فلق 3' أو إضافة ذيل الأدينين Poly (A) site أو عمليات خياطة النسخ الأولية من mRNA وحصول الطفرات في هذه المناطق تمنع معالجة hn RNA (Heterogeneous nuclear RNA) وإنتاج mRNA الناضج الفعال وبالتالي تمنع التشفير للبروتينات.
- تراكم الطفرات في الجينوم على مرور الزمن يغير من الجينوم وعليه فإن الجينوم ليس وحدة ثابتة وإنما يتغير عبر الزمن أثناء عملية التطور
- الطفرات تسبب الأمراض كما في حدوث السرطانات
- بعض الطفرات قد تكون حميدة وتزيد من تلاؤم الكائن مع بيئته وأخرى قد تكون كارثية تؤدي إلى موت الكائن، والأخيرة تحدث عندما يتغير بروتين مهم للخلايا ويصبح غير فعال مما يؤدي إلى موت الخلايا وفي أحيان أخرى قد تكون بدون تأثير محسوس. وعندما تكون الطفرة مفيدة وتساعد الكائن لابد أن تتوارث في الأجيال الناتجة خلال عملية التكاثر وإلا سيذهب تأثيرها بانتهاء الفرد كما ذكر أعلاه.

أنواع الطفرات

هناك عدة أسس لتقسيم الطفرات منها:

❖ النمط المظهري الذي ينتج عن حدوث الطفرة وفي هذا المجال يمكن ذكر

- **الطفرات الصامتة Silent mutants** وهذه يمكن أن تحدث في المناطق الموجودة بين الجينات أو في المناطق غير المشفرة في الجين، أو تحدث في المناطق المشفرة دون التأثير في البروتينات الناتجة وهذا يحصل بالنسبة للحوامض الامينية التي لها أكثر من شفرة مترادفة، إذ أن التغير سيؤدي إلى إنتاج الحامض الاميني نفسه كما موضح في الشكل (24.3) .

وهذا يكون وراء ظهور السلالات المتطابقة من حيث المواصفات وتوالي الحوامض الامينية في بروتيناتها ولكنها تظهر الاختلاف في توالي جينوماتها.

طفرات صامتة			
ATG	GAA	GCA	CGT
Met	Glu	Ala	Gly
↓			
ATG	GAG	GCA	CGT
Met	Glu	Ala	Gly

شكل 24.3 : الطفرات الصامتة

- **الطفرات الخاطئة Missense mutants**، طفرات تكون موضعية او نقطية Point mutation اي يحصل تغير في قاعدة نروجينية واحدة يؤدي الى تغير الحامض الاميني المشفر له وبالتالي يدمج حامض اميني غير صحيح ويؤدي الى تغير البروتين كما موضح في الشكل 25.3

طفرات خاطئة			
ATG	GAA	GCA	CGT
Met	Glu	Ala	Gly
↓			
ATG	GAC	GCA	CGT
Met	Asp	Ala	Gly

شكل 25.3 : الطفرات الخاطئة

- وفي اغلب الأحيان تكون هذه الطفرات يمكن تحملها اذ لا تؤدي الى تغيير كبير في وظيفة البروتين وفي بعض الأحيان يكون لها تأثير كبير.
- **الطفرات التافهة Nonsense mutants** طفرات تحدث بحيث تغير الشفرة الوراثية للحوامض الامينية الى شفرة وقف تؤدي الى إنهاء عملية الترجمة مما يؤدي الى تقصير طول السلسلة الببتيدية الناتجة من الجين كما موضح في الشكل (26.3) . وتأثيرها يعتمد على الموقع الذي حصلت فيه ولكن في العادة تكون كارثية وتنتج بروتين مبتور غير فعال.



شكل 26.3 : الطفرات التافهة

- **طفرات القراءة المستمرة Readthrough mutants:** طفرات نقطية أيضا تؤدي الى تغير شفرة الوقف الى شفرة لأحد الحوامض الامينية مما يؤدي الى امتداد البروتين الى ان يصل الى شفرة وقف قادمة خاصة في رسائل mRNA المتعددة. ومعظم البروتينات تتحمل الاستطالة البسيطة دون التأثير في وظيفتها ولكن الاستطالات الكبيرة تؤثر في طوي البروتين وفي العادة تؤدي الى تخفيض فعاليته.
- **طفرات إزاحة أطر القراءة الصحيحة Frameshift reading mutants:** طفرات نقطية بطبيعتها وتكون اما إضافة او إزالة قاعدة نتروجينية واحدة في التواليات المشفرة مما يؤدي الى تغير كل الشفرات الوراثة التي تقع الى يمينها كما موضح في الشكل 27.3 :



شكل 27.3 : طفرات إزاحة الأثر القراءة الصحيحة

اذ ان إزالة قاعدة واحدة سيزحف تأثيرها في الشفرات التي بعدها وتتغير الحوامض الامينية في السلاسل الببتيدية الناتجة.

- طفرات في مواقع الخياطة: طفرات تحدث في المنطقة الحدودية بين الانترونات والاكسونات في جينومات الخلايا حقيقية النواة تؤدي الى اضطراب خياطة الجين (التي سيأتي ذكرها لاحقا) والطفرات يمكن ان تؤدي الى عدم إمكانية إزالة الانترونات من النسخ الأولية للـ RNA. ولكن الخلايا يمكن ان تستعمل مواقع بديلة وهي المواقع الخفية وتتمثل بتواليات تشبه التواليات الأصلية وربما نتجت من حدوث طفرات مثل طفرات إزاحة أطر القراءة وبعض الأحيان يكون وجود التواليات الخفية أفضل من الأصلية .

ان حدوث اضطرابات في مواقع الخياطة يؤدي اما الى إضافة جزء من البروتين او حذف جزء منه (للجين تحت التعبير) وفي العادة ترتبط ببعض الأمراض مثل الثلاسيميا Thalassemia.

العوامل المطفرة

هناك عدداً من المواد الكيماوية او العوامل الفيزيائية الموجودة في البيئة المحيطة بالخلايا تحت الطفرات وبذا فهي تمثل الرابط بين الوراثة والبيئة، إضافة الى ان الطفرات يمكن ان تحصل بشكل تلقائي. وتختلف العوامل المطفرة في تأثيرها أثناء عملية التطهير.

وفي العموم فان المطفرات يمكن ان تؤدي الى الأكسدة او حث المثيلة غير المسيطر عليها او التحلل او تحويل القواعد النتروجينية ، وفي التالي بعض الحالات :

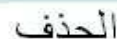
- بعض المواد الفعالة تغير تركيب القواعد النتروجينية الى تراكيب كيماوية مختلفة او تؤدي الى إزالة القواعد النتروجينية.

فكل من القواعد النتروجينية مثل السايترين والادين والكوانين تحوي على مجموعة أمين التي يمكن ان تزال مؤدية الى عدم انتظام تكامل القواعد في أشرطة DNA ومن العوامل التي لها هذا التأثير حامض النتروز Nitrous acid . ومن جهة ثانية فان إزالة مجموعة الأمين يمكن ان يحدث بشكل تلقائي بمعدل 100 تغيير قاعدة 1 الجينوم 1 يوم ، وأكثر المناطق التي تحدث فيها هذه العملية هي قاعدة السايترين في الموقع 4 لذا تعد هذه بمثابة مناطق ساخنة، وينتج من عملية إزالة الأمين تكوين القاعدة النتروجينية اليوراسيل التي ترتبط مع الأدينين بدلا من الكوانين، ووجود اليوراسيل في أشرطة DNA غير مسموح به، لذا تقوم الخلايا بإصلاح هذا العطب بكفاءة بواسطة الإنزيم Uracil-DNA glycosylase وربما بآليات أخرى.

اما التفاعل الآخر الذي يحصل في البقع الساخنة فهي إضافة مجموعة مثيل على الذرة الخامسة للسايترين وإنتاج Methyl cytosine وإزالة مجموعة الأمين من القاعدة الممثلة ينتج الثايمين الحاوي على الأوكسجين في الموقع 4 بدلا من مجموعة الأمين. وهذه قد لا يتم إصلاحها لان الثايمين يمثل احد المكونات الطبيعية في DNA . وهناك تحويلات أخرى يمكن ان تطول القواعد النتروجينية الأخرى .

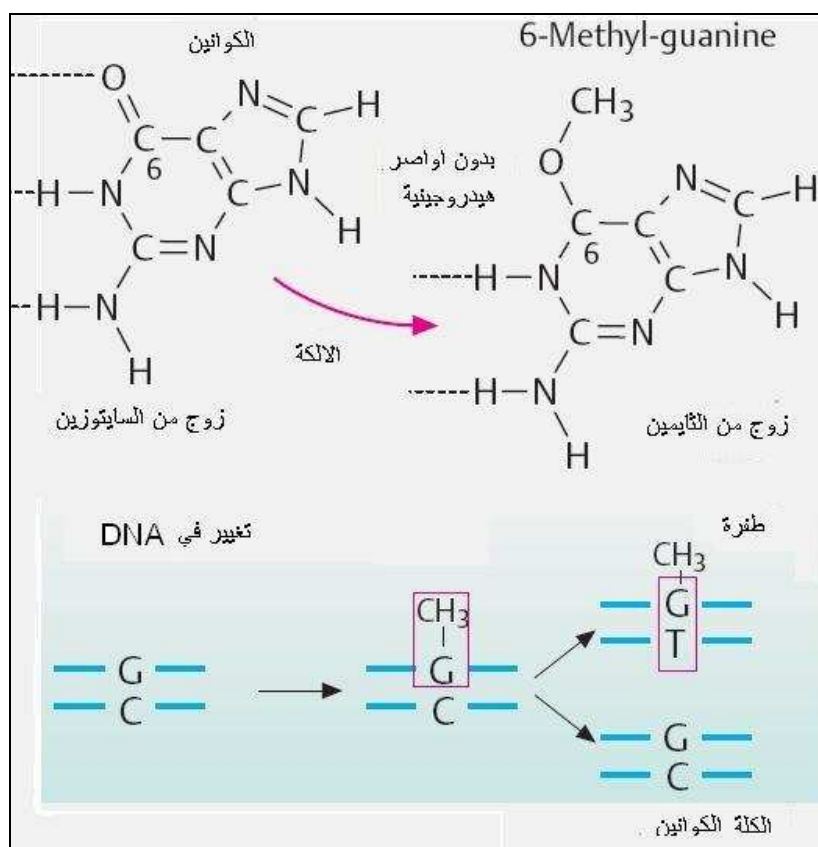
- بعض المواد يمكن ان تؤدي الى إزالة قاعدة نتروجينية او امتداد من النيوكليوتيدات وهذا ما يسمى بطفرات الشطب او الحذف ، والامتدادات المحذوفة قد تكون جين واحد او عدد من الجينات وعندما تكون الامتدادات كبيرة تسمى العوامل Clastogenic agent وفي حالة

:28.3



شكل 28.3 : حدوث طفرات الحذف

29.3 بالتايمين في الجزيئة الجديدة كما موضح في الشكل 29.3



شكل 29.3 : ألكلة القواعد النيتروجينية

ومن العوامل المسببة لهذه الحالة

Ethyl nitrosourea (ENU)

Ethyl methanesulfonate (EMS)

Dimethylnitrosamine

N-methyl-N-nitro-N' Nitrosoguanidine (NTG)

مضاهيات قواعد النيوكليوتيدات Nucleotide base analogue

مركبات تؤدي الى تطفير DNA نظرا لتشابهها مع قواعد وتستطيع ان تندمج بالأشرطة أثناء تضاعفها. فالمادة 5-Bromodeoxyuridine (5BrdU) تشابه الثايمين وتحتوي على ذرة بروم بدلا من مجموعة الميثيل في الموقع 5 وبذا تستطيع ان تندمج أثناء التضاعف ، ووجود ذرة البروم يكون غير ملائما في DNA ويؤدي الى اضطراب تكامل القواعد وبالتالي الخطأ في ازدواجها.

الأشعة فوق البنفسجية

الأشعة عند الأطوال الموجية القصيرة منها مثل 260 نانومتر وبالتحديد 254 نانومتر التي تمثل ذروة امتصاص DNA للأشعة تؤدي الى تكوين أواصر تساهمية بين ثمالات الثايمين المتجاورة عند الموقع 5 ، 6 مكونة مزدوجات الثايمين ويؤدي تكوينها الى التداخل السلبي مع

عملية التضاعف والانتساخ اذا لم تتم معالجتها. وتؤثر الأشعة فوق البنفسجية بطريق غير مباشر ايضا وذلك بإنتاج نواتج التحلل الضوئي مؤدية الى تكوين أواصر تساهمية بين ذرات الكربون في المواقع 4 و 6 من النيوكليوتيدات المتجاورة وتسمى 4-6 photoproducts .

الطفرات التلقائية

الطفرات التي تحدث دون الحث بالمواد المطفرة ويمكن ان تنتج عن:

- أخطاء أثناء عملية تضاعف DNA ، فالمعروف ان تضاعف الأشرطة يكون بشكل شبه محافظ ويعتمد على أساس تكامل القواعد في الأشرطة ، ويمكن ان يحدث الخطأ أثناء التضاعف بمعدل 1 في 10^5 قاعدة وينخفض التردد الى 1 في 10^7 - 10^8 بألية تصحيح القراءة للإنزيمات العاملة في التضاعف وكذلك آليات تصحيح عدم التلائم Mismatch repair ، وإذا لم تصحح الطفرة فإنها تعبر الى الخلايا الناتجة ويكون تأثيرها معتمدا على مكان وجودها مثلا في التواليات المشفرة فانه سيؤثر في وظيفة البروتين الناتج.

- طفرات تحدث نتيجة لانزلاق او تفاوت عملية انزلاق التضاعف Replication slippage وهذه بطبيعة الحال لا تتدرج تحت المجاميع المدرجة أعلاه ، وتنتج من عدم الاصطفاف المضبوط بين تواليات DNA الاليلية وغير الاليلية أثناء عمليات التضاعف فمثلا عند احتواء أشرطة القالب على مكررات من القواعد مثل CA repeats كما في التتابع الدقيقة التتابع الدقيقة (التي سيأتي ذكرها لاحقا)، فان الأشرطة الجديدة والقوالب يمكن ان يحصل زحف في مواقعها واحدا بالنسبة للأخرى وبالتضاعف يحصل انزلاق لأنزيم الكوثر Polymerase slippage مما يؤدي الى أخطاء في ازدواج المكررات حيث تنتسخ بعض المكررات مرتين في حين لا تنتسخ الأخرى اعتمادا على اتجاه الزحف الحاصل. ويمكن تمييز اتجاه الزحف بالنسبة للأشرطة الحديثة، فاذا كان انزلاق الشريط الجديد للأمام فهذا يؤدي الى بقاء منطقة من الشريط الأبوي او القالب غير مزدوجة مما يؤدي الى عمليات إقحام فيها . اما اذا كان انزلاق الأشرطة الجديدة الى الوراء فانها ستعاني من عمليات حذف.

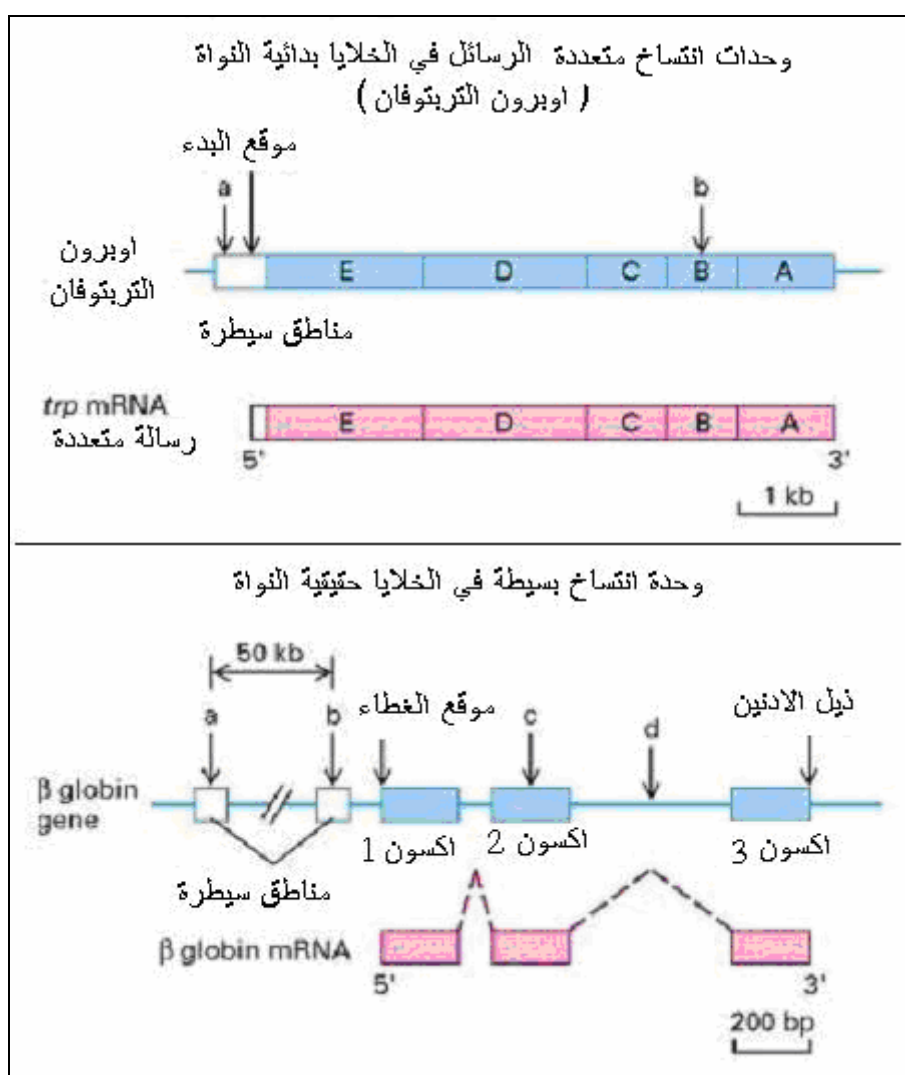
جينومات بدائية النواة

تمتاز الأحياء التي تضم البكتريا والاراكيا بان موادها الوراثية لا تكون محددة ومحاطة بغشاء نووي مميز كما في الأحياء حقيقية النواة، ومعظم الأحياء المدروسة يكون بشكل كروموسوم حلقي، والبعض منها يحوي على كروموسومات مفتوحة، او تكون حالة تذبذب بين الاثنين وفق الحالة التي تمر بها الخلايا.

تتجمع موادها النووي في منطقة تسمى الجسم المضغوط Compactosome وتكون المنطقة خالية من الرايبوزوم والتراكيب الأخرى ، وتتوزع الجينات المراد انتساخها الى المناطق الخارجية. وتختلف أعداد الكروموسومات والمتضاعفات مثل البلازميدات بين الأنواع المختلفة، فمثلا الاراكيا *Haloarcula marismortui* التي تعيش في البحر الميت ذو تركيز ملحي يصل الى 4.5 مولر تحوي على تسع متضاعفات منها 2 من الكروموسومات 7 من البلازميدات وهي تمثل اكبر جينوم للاراكيا تم تحديد تواليه، وفي الغالب تميل الأحياء المحبة للملوحة الى امتلاك أكثر من متضاعف.

وتختلف حجوم جينومات البكتيريا وتمتد من 150 كيلوقاعدة طولا الى 15 ميكاقاعدة، والمتطفلات منها تكون جينوماتها صغيرة نوعا ما نظرا لفقدان الجينات المشفرة للبروتينات التي تكون ضرورية للبيئات التي توجد فيها، وفي العموم تكون جينومات البكتيريا المرضية والمتطفلة اصغر من أقرانها حرة المعيشة.

وفي البكتيريا على وجه الخصوص تميل الجينات الخاصة بمسار ايضي معين الى التجمع في مكان محدد بشكل متجاور كما في اوبرون اللاكتوز او غيره مكونة وحدة انتساخية واحدة الموضح احدها في الشكل 30.3 .



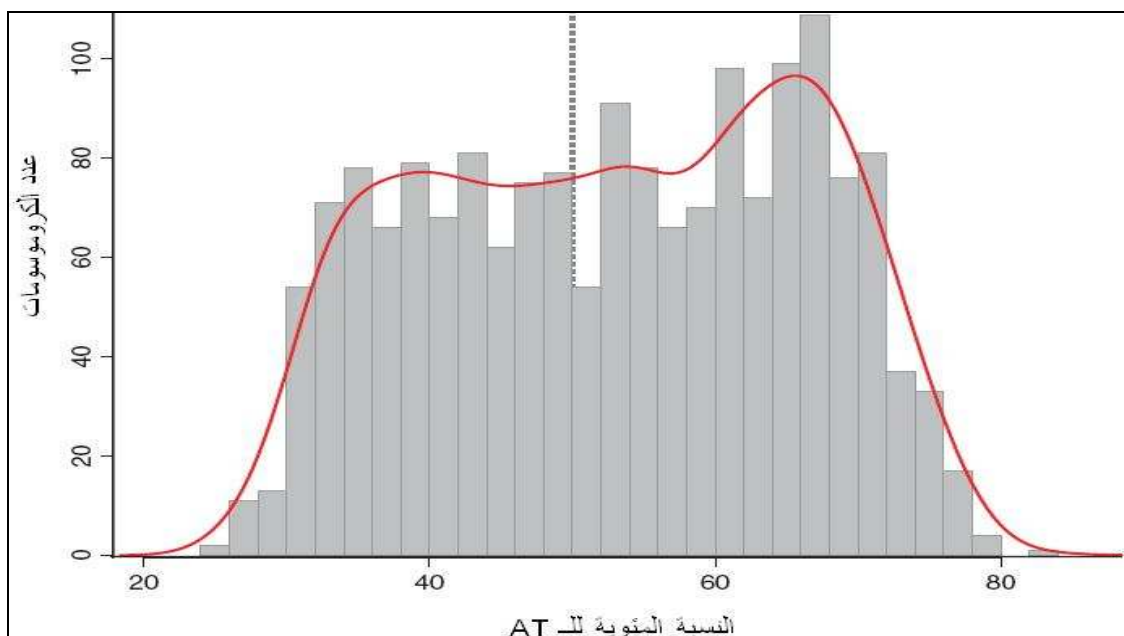
شكل 30.3 : وحدات انتساخ مختلفة

والملاحظ انه في البكتيريا *E. coli* ينتج الاوبرون نسخة من mRNA يصل حجمها الى حوالي 7 كيلو قاعدة متعددة الرسائل وإذا حصلت طفرة في منطقة السيطرة على الانتساخ فان هذا

يؤدي الى توقف الاوبرون كله. وفي مثل هذه الاوبرونات يمكن ان يبدأ الانتساخ من المنطقة المجاورة لمنطقة الارتباط الى الرايبوزوم .rbs.

محتوى الجينومات من AT

يمكن وصف الجينومات من تحديد محتواها من AT و GC ($\%AT + \%GC = 100\%$) وذلك لان طغيان زوج على اخر يؤثر في الموصفات الفيزيائية نظرا للاختلافات بين AT و GC ، والمحتوى الافتراضي والنظري ان يكون بحدود 50% لكل منهما وهذا ما يحصل في جينوم بعض سلالات *E. coli* التي تحوي على 49.5% من AT، اما العموم فيتراوح بين 25-75% ويكون توزيعها غير منتظم بين مناطق التضاعف والمناطق حول منطقة أصل التضاعف. اي يكون التوزيع والمحتوى غير منتظم كما في الشكل 31.3:

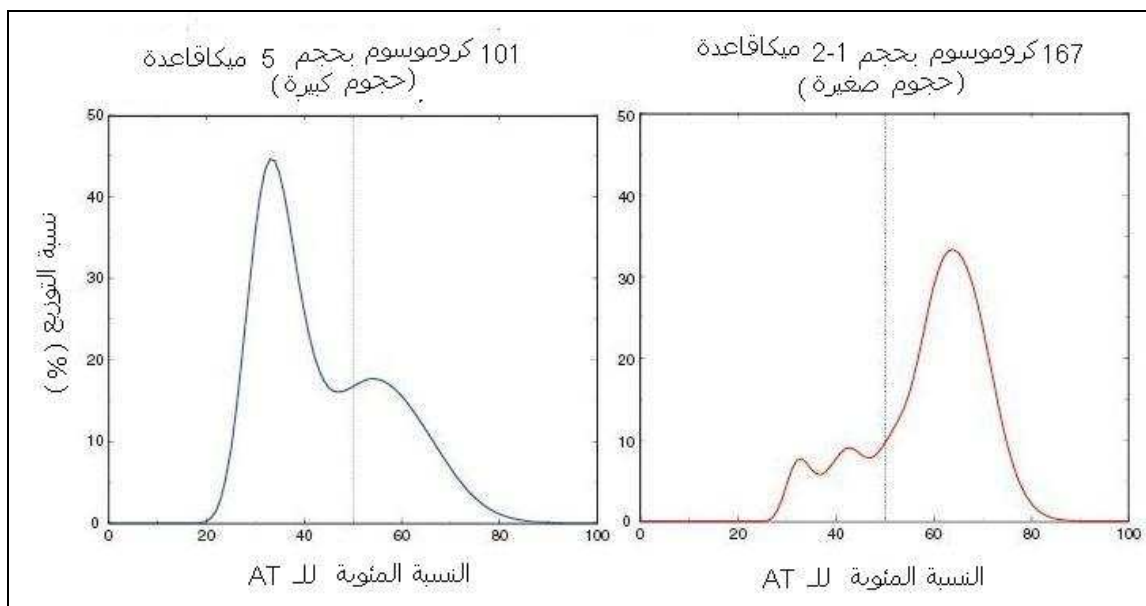


شكل 31.3 : توزيع النسب المئوية للـ AT على جينومات أكثر من ألف من خلايا بدائية النواة والبلازميدات المحددة التوالي ويتضح زيادة عدد الكروموسومات التي فيها مستوى عالي من AT

ويظهر التوزيع في منطقة الوسط 50% هو الأقل ومناطق محيطة بها تحوي على نسبة 35% و 65% وهذا التوزيع يعود للأحياء التي تم تحديد تواليات جينوماتها ، وقد يعود السبب الى ان الجينومات الصغيرة تكون غنية بـ AT كما يظهر عند فصل الجينومات الأكثر 5 ميكافادة والأخرى الأقل من 1-2 ميكافادة كما موضح في الشكل (32.3) .

وهذه النتائج لا تشير الى ارتباط قوي بين محتوى AT% وحجم الجينوم ولكن الملاحظة العامة ان البكتيريا المتطفلة تكون جينوماتها غنية بـ AT في حين تكون جينومات بكتيريا التربة غنية بـ GC. فضلا عن حجم الجينوم فان تفضيل استعمال الشفرات يلعب دورا أساسيا في التأثير

في محتوى الجينوم من AT وان كان استعمال الشفرات له علاقة ضعيفة بحجم الجينوم. والمعتقد الآخر ان الجينومات الغنية بـ AT يميل فيها GC الى التطفر والتحول الى AT نظرا لان إزالة مجموعة الأمين من السايتوزين تحدث أكثر من باقي العمليات خاصة في الجينومات الصغيرة التي تقل فيها إنزيمات تصليح DNA لذا تزداد فيها نسبة AT.



شكل 32.3 : توزيع النسب المئوية للـ AT على الجينومات الصغيرة والكبيرة

أنواع الجينومات

هناك بعض الخصائص المميزة لجينومات الأحياء بدائية النواة، فيمكن للجينوم ان يكون:

جينوم مركزي Core genome:

ويمثل مجمع من الجينات التي تشترك فيها كل سلالات النوع الواحد وتكون مسئولة عن النواحي الأساسية من بايولوجية الأنواع معطية الصفات العامة للنمط المظهري اي يكون من النوع الثابت والجينات التي لا تعود الى الجينوم المركزي هي المسئولة عن التباين الكبير في حجم الجينوم الخاص بالكائن داخل النوع. ويمثل الجينوم المركزي الجينات الأساسية مثل تلك المسئولة عن الانتساخ والترجمة وتضاعف الجينوم والبروتينات الأساسية في العمليات الايضية، ولذلك يعد المجموع الأدنى من الجينات Minimal gene set وعند إجراء عمليات المقارنة فان الحد الفاصل Cutoff المستعمل من قبل بعض الباحثين لاعتبار الجينات ضمن الجينوم المركزي او الجينوم الشامل هو التماثل بنسبة 50% ، ويمثل الجينوم المركزي الوحدة التصنيفية Taxa للأنواع .

الجينوم غير الضروري Dispensable genome:

يمثل مجمع الجينات التي توجد في بعض سلالات النوع الواحد ويقسم بدوره الى:

- جينات توجد في اثنين أو أكثر من السلالات
- جينات خاصة بسلالة واحدة

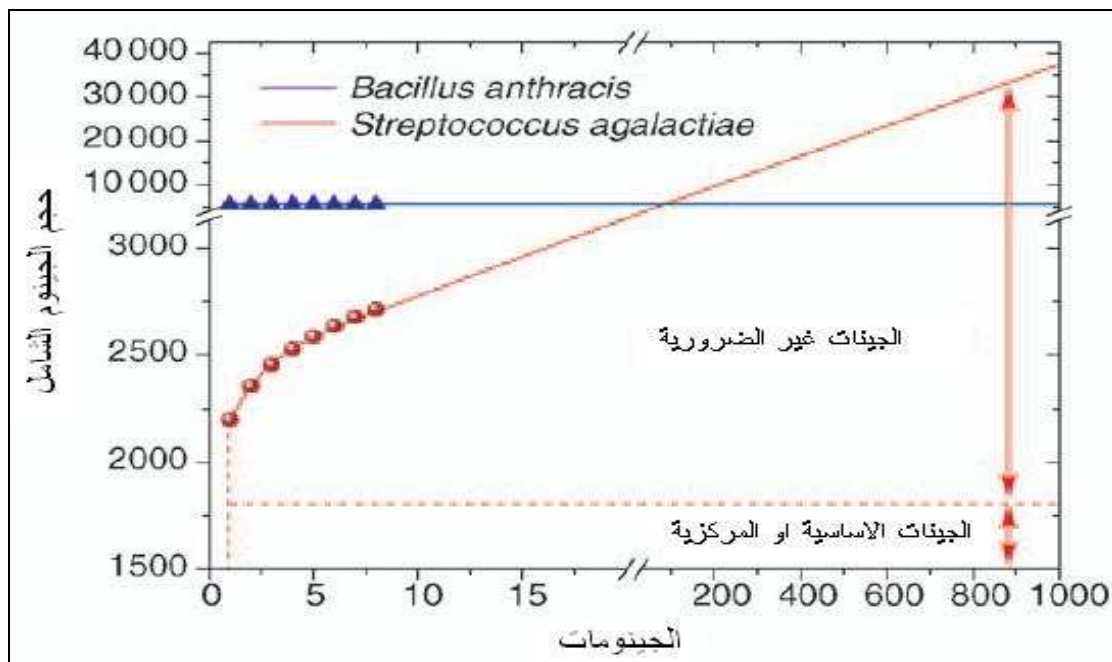
وتساهم في التنوع الذي يظهر على أفراد النوع الواحد ويمكن ان تشفر لمسارات ايضية مساندة او وظائف ليست أساسية لحياة الخلية ولكنها تضيف مزايا انتخابية جيدة للكائن مثل التطبع للبيئات المختلفة والمقاومة للمضادات الحيوية واستعمار مضاييف جديدة. وتتجمع هذه المجموعة في جزر من الجينوم وتكون في الغالب محاطة بتواليات متكررة من النيوكلووتيدات تمتاز بمحتواها غير الطبيعي من G+C. واغلب عمليات التهميش الوظيفي تضعها ضمن المجموعة الافتراضية ، وتكون ذات علاقة بالجينات الآتية من العاثيات او العناصر القافزة. وعلى العموم في الأحياء المرضية يكون فقدان هذه العناصر المتحركة مرافقا لضعف الضراوة.

ويمثل الجينوم غير الضروري الفرق بين الجينوم الشامل والجينوم المركزي او الأساسي ويمكن ان يمثل من 4%-50% من الجينات الموجودة في جينوم النوع. وتكون الجينات غير الأساسية شبكات معقدة زائدة، وفي الأغلب فان وجود جين او جينات معينة يمكن ان تؤدي الى غياب جينات أخرى.

الجينوم الشامل Pan genome

يمثل مجمع لكل الجينات في النوع اي انه يضم الجينوم المركزي مع الجينوم غير الضروري. ومن هنا يتضح ان الاستمرار في تحديد تواليات سلالات مختلفة للنوع تساعد في اكتشاف جينات جديدة او عدم ظهور جينات كانت موجودة في سلالات سبق تحديد تواليها، لذلك كان من الأفضل استعمال الجينوم الشامل لوصف النوع كما موضح في الشكل التالي (33.3) لبعض البكتريا.

والملاحظ انه عند تحديد تواليات جينومات بعض الأحياء ان هناك زيادة في أعداد الجينات وهذه تدعى الجينوم الشامل المفتوح وتمثل حالة إضافة تواليات جديدة عند تحديد تواليات سلالات جديدة اي ان الإضافة ستكون للجينوم الشامل والمركزي. وتوجد في الأنواع التي تستعمر بيئات متعددة ومختلفة ولها طرق متعددة لتبادل المواد الوراثية ومنها على سبيل المثال لا الحصر المسبقيات الكروية Streptococci، E. coli، Helicobacter pylori، Salmonella ومكورات السحاية Meningococci. في حين انه في بعض الأنواع المحددة مثل Bacillus anthracis لا تزداد الأعداد وانما توقفت عند عدد محدد بعد تحديد توالي 4 سلالات ومثل هذا يطلق عليه الجينوم الشامل المغلق، ويعد الجينوم مغلقا اذا كانت ثبوته اكبر من 20-35%. وتوجد هذه في الأنواع التي تعيش في بيئات معزولة ولها قابليات محدودة لاستقبال الجينات من الخارج كما في بكتريا الجمرة الخبيثة المذكورة توأ وكذلك بكتريا السل Mycobacterium tuberculosis و Chlamydia trachomatis ولعل أفضل الأمثلة عليها البكتريا Buchnera aphidicola التي تكون في حالة تعايشية مع المن Aphids وهذه لم يحصل تغيير او إعادة ترتيب لجينوماتها ولم تفقد او تكتسب مواد وراثية على مدى 50 مليون سنة الماضية كما تشير الدراسات الخاصة بها لذلك يعد جينومها من النوع الثابت بشكل كبير.



شكل 33.3 : الجينوم الشامل وعدد الجينات الأساسية وغير الضرورية للبكتيريا.

Strep. agalactiae و *B. anthracis*

وتتبع الأحياء استراتيجيات مختلفة في توزيع جيناتها بين الجينوم المركزي والعام، فمثلا الجنس *Streptomyces* ذا المحتوى العالي من GC والمهم من عدة نواحي تمتلك خلاياه جينوما عاما يتراوح بين 8-10 ميكافاعدة، ويمثل الجينوم المركزي 5-6 ميكافاعدة حاوية على الجينات الأساسية اما الباقي فهي الأطراف التي تعاني من تغييرات كبيرة ، هذه المناطق تمتلك حدود 10⁵ جين لإنتاج مواد الايض الثانوي، كما ان بعض جينات الايض الثانوي تقع على البلازميدات البعض منها بلازميدات مفتوحة توجد في خلايا الجنس، وهذا التوزيع يعطى لأفراد الجنس مرونة وسعة كبيرة في إنتاج مواد مختلفة التي يمتاز بها الجنس.

الهدف من دراسة الجينوم الشامل والمركزي

تكون الحاجة ماسة لتحديد تواليات السلالات المختلفة من النوع الواحد لفهم التنوع ضمن النوع الواحد وهذا لا يكون هدفاً نظرياً فقط ، فقد وجد ان اللقاحات الوقائية لبعض المكورات المسبحية يكون ممكناً في حالة التعرف على الجينوم غير الأساسي للسلالة. كما ان البحث في الجينوم غير الأساسي او غير الضروري يكشف عن وجود عوامل الضراوة مثل الشعيرات Pili وغيرها ولذا يتوقع ان تكشف الطرق الحديثة الكثير عن جينومات الميكروبات التي لا يعرف عنها الشيء الكثير في الوقت الحاضر.

وتلعب أنواع الجينومات دوراً مهماً في الانتخاب الايجابي في عمليات التطبع التي تحصل نتيجة لتغاير وتطور الجينومات وخاصة في الأحياء المرضية التي تظهر سعة كبيرة في التطبع للمضيف او المضاييف الجديدة التي تغزوها مثل التطبع تجاه الجهاز المناعي والتطبع للمضادات الحيوية. وهناك ثلاث آليات يمكن ان تؤدي الى التطبع منها:

- الانتخاب الايجابي حيث تميل الأحياء الى تثبيت الطفرات المفيدة لها
- تقبل المواد الوراثية التي تنقل اليها بالطريق الأفقي والسماح بتأسيبها Recombination مع جينوماتها.
- إظهار السعة في تنظيم الجينومات

والعاملان الأخيران تعد الأساس في تطبع الأحياء المرضية وان معدل التأشب - ربما- يكون أعلى من حدوث الطفرات. وقد أوجدت طرق إحصائية لحساب معدل الطفرات للاستبدال المترادف وغير المترادف. ويعد ارتفاع مستوى الاستبدال غير المترادف مقارنة بالاستبدال المترادف مؤشرا" للانتخاب الموجب (انتخاب دارون)، وهذا النوع من الاستبدال يقود الى تثبيت الطفرات المفيدة وهو وراء التغيرات التطوعية في الجينات والجينومات المؤدي الى نشوء الاختلافات بين الأنواع ، ومن خلال الطرق الإحصائية تم التوصل الى حقيقة ان الميل لاكتساب الجينات هي اكبر من فقد الجينات. وعليه فان زيادة أعداد الجينات يمكن ان تحدث من تضاعف الجينات نفسها ولكن الطريقة الأكثر شيوعا هي اكتساب الجينات الجديدة ومن ورائها الوظائف الجديدة نتيجة للانتقال الأفقي للجينات بين أحياء ربما لا تكون بينها علاقة، وبذا فان الانتقال الأفقي يمثل الطريق الأسرع لعمليات التطور ومساعدة الأحياء في التطبع للبيئات المتغيرة.

البلازميدات

جزيئات DNA مفصولة عن الكروموسوم الأصلي للخلايا، وتوجد في العادة في البكتريا والاركية وقليل منها يوجد في الأحياء المجهرية حقيقية النواة. تكون هذه القطع او الجزيئات بشكل أشربة مزدوجة البعض منها فائق الالتفاف وتتضاعف بشكل مستقل عن كروموسوم الخلية اي انها من المتضاعفات وتشكل جزءا مهما من مكنون العناصر المتحركة Mobilome.

وتكون الجزيئات بمثابة DNA عاري وتشفر في بعض الأحيان لما يساعدها في الانتقال الذي يرتبط بشكل وثيق بعملية الاقتران Conjugation اي ان انتقالها يحتاج الى تلامس فيزيائي بين الخلايا لذلك فهي تمثل جزءا مهما من عمليات انتقال الجينات الأفقي . والبلازميدات كما ذكر يكون اغلبها حلقي ولكن هناك العديد منها مفتوح فمثلا في البكتريا الحلزونية *Borrelia burgdorferi* الحاوية على 17 بلازميد، تسع منها على الأقل مفتوحة ، والبلازميدات المفتوحة تحمي أطرافها من التفكك بواسطة تراكيب تشبه أطراف الكروموسومات بتكوين عروات ماشية الشعر Hairpin loops او تكسد بعض البروتينات الخاصة عند الأطراف لمنع وصول إنزيمات القطع الخارجي اليها. ومن الجدير بالذكر ان اكبر بلازميد Megaplasmid تم تحديد تواليه في الجنس *Strep. clavuligerus* ذات القابليات المتعددة لإنتاج مواد الايض الثانوي.

أصناف البلازميدات

تقسم البلازميدات الى عدة أقسام اعتمادا على مؤشرات خاصة منها:

- طرق الانتقال: وفي هذه الحالة تقسم الى بلازميدات اقترانية وهي الحاوية على جينات *tra* التي تقوم بالعملية المعقدة للاقتران بمساعدة تراكيب أخرى. والقسم الثاني منها البلازميدات

غير الاقتترانية وهذه ليس لها القابلية على بدء الانتقال من مضيف الى آخر وإنما تنتقل بمساعدة البلازميدات الاقتترانية. وهناك مجموعة ثالثة من البلازميدات التي تقع بين هاتين المجموعتين وهي البلازميدات المحركة Mobilizable تمتلك جينات ولكن لا تستطيع الانتقال لذا تتطفل على البلازميدات الاقتترانية لنقل جيناتها.

● **التوافقية Compatibility** : وتعني إمكانية وجود أكثر من نوع من البلازميدات في الخلية نفسها وفي الوقت نفسه وهنا تقسم الى:

♣ **Incompatible** المتوافقة غير المتوافقة وهي الأنواع المتشابهة من البلازميدات ويمكن لنوع واحد منها البقاء في الخلية نظرا لتشابه مكائن تضاعفها واحتياجها للمستلزمات الخلوية نفسها للبقاء والتضاعف وفي حالة وجودها فان احدها يطرد بعد تضاعف الخلية ، والموضح بعضها في الجدول 4.3 .

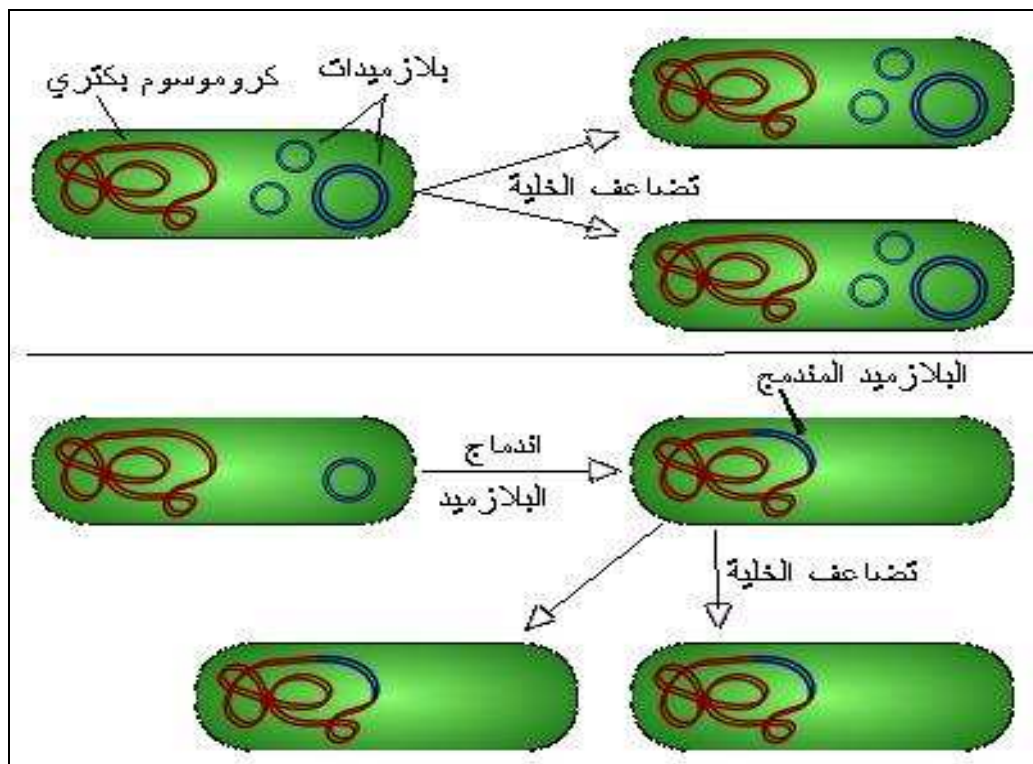
♣ **Compatible** المتوافقة عادة تكون البلازميدات غير متشابهة ويمكنها البقاء في الخلية الواحدة وفي الوقت نفسه

جدول 4.3 : مجاميع البلازميدات غير المتوافقة

البلازميدات	المجموعة غير المتوافقة
F, R386	FI
R1	FII
Col B-K99, Col B-K166	FIII
R124	FIV
R62, R64, R483	I
خمس مجاميع ثانوية على الأقل	J
R391	N
R46	●
R724	P
RP4, RK2	Q
RSF1010	T
R401	W
R388, S-a	

● **الوظيفة** : تمثل المؤشر الآخر لتصنيف البلازميدات ، والبلازميدات لها القابلية لإضفاء العديد من المزايا على الخلايا التي توجد فيها، وكان يعتقد سابقا انها تشفر لوظائف ثانوية ويمكن

- الاستغناء عنها وإنها لا تكون ضمن الجينوم المركزي للخلايا ولكن تبين فيما بعد ان هذا قد يكون صحيحا الى حد ما. ومن هذه البلازميدات:
- ◆ بلازميد الخصوبة او ما يسمى بـ F. factor يحوي البلازميد على جينات *tra* لذلك له القدرة على بدء عملية الاقتران.
 - ◆ بلازميد المقاومة ويعرف بـ R-factor من الناحية التاريخية ويساعد الخلايا في مقاومة المضادات الحيوية والسموم ، ويساعد الخلايا على تكوين الشعيرات الجنسية لبدء عملية الاقتران وقد تكون هذه المجموعة مفيدة جدا من وجهة نظر الخلية لجعلها تقاوم العديد من المضادات التي أنتجت ضدها ولكن من وجهة النظر المقابلة (الإنسان) كانت هذه البلازميدات كارثية التأثير.
 - ◆ Col plasmids بلازميدات تنتج البكتريوسينات او البروتينات المضادة لقتل البكتريا الأخرى . ومنها ما ينتج Colicins في *E. coli* . ويوجد العديد منها في البكتريات الأخرى مثل مجموعة بكتريا حامض اللاكتيك *Lactic acid bacteria* والتي استغلت في الحفظ الحيوي Biopreservation للأغذية ، ومن الجدير بالذكر ان مجموعة البلازميدات التي تشفر لسموم تقتل غيرها من الخلايا تشفر في الوقت ذاته لنظام المناعة من السموم لحماية نفسها وتكون عادة في اوبرون واحد .
 - ◆ بلازميدات المعايشة وتثبيت النتروجين وتوجد الأولى في بكتريا المثبتة للنتروجين بطريقة تعايشية وتكون مسئولة عن تكوين العقد الجذرية مع نباتات البقول، وجزء من الجينات المسئولة عن تثبيت النتروجين في هذه المجموعة يوجد على البلازميد. اما في البكتريا *Klebsiella pneumoniae* فان تثبيت النتروجين فيها موجود على بلازميد كبير.
 - ◆ بلازميدات حث الأورام في النباتات (Ti) Tumor induced plasmid موجود في البكتريا الزراعية *Agrobacterium* مثل *A. tumefaciens* وغيرها من الأنواع التي تصيب نباتات ذوات الفلقتين بعد حدوث جروح فيها وينقل البلازميد وخاصة قطعة منه T-DNA التي تحمل عددا" من الجينات المشفرة للهرمونات النباتية الى جينوم الخلية النباتية ويعبر عنها وتؤدي الى انقسام ونمو الخلايا بشكل خارج عن السيطرة مولدة الورم. وقد استغلت هذه البلازميدات بشكل كبير في هندسة النباتات وراثيا.
 - ◆ بلازميدات مفككة تساعد الخلايا في تفكيك المواد العvisية وغير الطبيعية مثل التولوين وغيرها عند شحة المواد الغذائية الطبيعية بصورة خاصة .
 - ◆ بلازميدات الضراوة تساعد الخلايا البكتيرية لتتحول الى أحياء ممرضة وذلك بزيادة التعبير عن جينات عوامل الضراوة فيها.
- المؤشر الآخر الذي تقسم على ضوءه البلازميدات هو طبيعة وجودها في الخلايا. فالبعض منها عند دخوله الى الخلايا يندمج مع كروموسومها ويسمى البلازميدات الاندماجية Integrative plasmid ويسمى ايضا الجسم الخارجي Episome ، والقسم الآخر يبقى حرا في الخلية ويتضاعف بشكل مستقل عن الكروموسوم وكما موضحة في الشكل 34.3:



شكل 34.3 : تضاعف البلازميدات الاندماجية وغير الاندماجية

وفضلا عما ذكر أعلاه فهناك بعض البلازميدات التي تم التلاعب بتركيبها لتواكب الاحتياجات المرغوبة، فمثلا تم بناء عدداً من البلازميدات من عدة قطع من بلازميدات مختلفة منها مجموعة pBR₃₂₂ ومجموعة pUC لتستعمل في الهندسة الوراثية. وكذلك تم بناء البلازميدات المكونة Shuttle plasmids التي تستطيع التضاعف في مضاييف متباعدة، إذ أضيف لها أكثر من أصل للتضاعف وتستعمل في الهندسة الوراثية، والأنواع الأخيرة وكذلك الأنواع المذكورة أعلاه متوفرة على النطاق التجاري.

مضاييف البلازميدات وعدد نسخها

بعض البلازميدات يمكنها التضاعف في عدد محدد من المضاييف (الأنواع البكتيرية) وتكون ذات مدى ضيق مثل Col E1 و pBR₃₂₂ و pUC18 المتخصصة ببكتريا *E. coli* والأنواع القريبة جدا منها. وبلازميدات تتضاعف في عدد كبير من الأنواع أي واسعة المضاييف وهذا يعني أن إشارات التواليات للتعبير الجيني والسيطرة عليه وكذلك إشارات التضاعف تكون مختلفة.

أما أعداد نسخ البلازميد في الخلية فتتناسب عكسياً مع حجم البلازميد الذي يتراوح بين 1-1000 كيلو قاعدة ، وصفة العدد تكون خاصة بالبلازميد وطريقة تنظيم تضاعفه التي تتم بعدة آليات وتكاد تكون متخصصة بكل بلازميد. فمثلا بلازميد الخصوبة يوجد 1 لكل خلية في حين pUC18 يوجد من 30-50 نسخة للخلية الواحدة، وبعض الأحيان وتحت ظروف خاصة يمكن

ان تزداد جرع البلازميدات الى الآلاف لكل خلية، ومن الجدير بالذكر ان البلازميدات تستعمل لتضخيم الجينات المرغوبة داخل الحي *In vivo*.

تضاعف البلازميدات

تتضاعف البلازميدات بشكل مستقل عن كروموسوم الخلية المضيفة نظرا لاحتوائها على نقطة أصل (Ori) خاصة بها. وطريقة تضاعف بلازميدات البكتريا السالبة لصبغة كرام تسمى تضاعف ثيتا θ replication اما في معظم بلازميدات البكتريا الموجبة لصبغة كرام فتكون بطريقة الدائرة المتدرجة Rolling circle replication. وأيا كانت طريقة التضاعف البلازميدي فهي تمثل الآلية التي يسيطر بها على عدد من صفات البلازميدات مثل مدى المضاييف التي توجد بها وعدد البلازميدات لكل كروموسوم Copy number ، والتوافقية المذكورة أعلاه وكذلك إمكانية تحركها وانتقالها من مضيف الى آخر.

فوائد البلازميدات

للـبلازميدات عدة فوائد قد تمثل جانبيين، الجانب الأول الخاص بالخلايا والجانب الثاني هو التعامل مع الخلايا ويعني العاملين في مجال الأحياء المجهرية وقد يتوافق الجانبان بعض الأحيان وتتضارب المصالح في أحيان أخرى وفيما يلي ذكر بسيط لبعض الجوانب.

• تساعد البلازميدات في تحسين الانتخاب للخلايا الملائمة تحت ظروف معينة مثل المقاومة للمضادات التي تنتجها الأحياء التي تشاركها البيئة نفسها والتي انسحبت لمقاومة المضادات العلاجية للأنواع الممرضة منها.

• تساعد في خلق العلاقات التعايشية وكذلك المساهمة في تثبيت النتروجين التي تعد العملية الأساس لدورة النتروجين على الأرض.

• تفكيك المواد العصبية والمشاركة في المعالجة الحيوية للبيئات الملوثة والتي استغلت في تصميم العديد من السلالات لهذا الغرض.

• استعمالها في الهندسة الوراثية اذ تكون أحد أحجار الأساس في هذا التوجه، اذ تستعمل في تكثير الجينات (أي زيادة جرع الجين) في مضاييف خاصة وهذه مستغلة على النطاق التجاري وفيها يتم بناء البلازميد المطلوب مع الجينات المرغوبة ويدخل الى الخلايا بطريقة التحول ، لتساعد في زيادة كميات البروتينات المطلوبة، وفي هذا المجال يحمل البلازميد كميات من DNA تتراوح بين 1-10 كيلو قاعدة واذا أريد تحميله كميات اكبر يتم اللجوء الى استعمال نواقل أخرى.

• تستعمل البلازميدات في العلاج الجيني وذلك بتحميل البلازميد بـ Zinc finger nuclease (ZFN) الذي يكسر شريطي DNA ويساعد في التآشب المتماثل Homologous recombination وعند تحميل مثل هذا البلازميد بالجين المعالج فانه يمكن ان يوصله الى المكان المطلوب من الكروموسوم وهذه الطريقة أفضل من استعمال عملية الدمج العشوائي وكذلك أفضل من الطرق المعتمدة على الفيروسات في إيصال الجينات الى المناطق التي تحتاجها.

بلازميدات الخمائر

تكون البلازميدات منتشرة في الخلايا بدائية النواة ولكن بعض الخمائر تحوي على البلازميدات ومنها:

- $2\mu\text{m}$ يوجد في نواة خميرة الخبز واستغل لأغراض كثيرة ضمن نطاق الهندسة الوراثية.
- **YIP** (Yeast integrative plasmids) مجموعة تتميز انها تعتمد على الاندماج مع كروموسومات المضيف لغرض البقاء والتضاعف ولها استعمالاتها الخاصة.
- **YRP** (Yeast replicative plasmids) وتقوم بنقل جزء من الكروموسوم الحاوي على نقطة أصل التضاعف وهي بلازميدات غير ثابتة وتفقد بسرعة خلال عملية تبرعم الخمائر.

طرق انتقال الجينات الأفقي

طرق الانتقال الأفقي تمثل الانتقال بين الخلايا غير المعتمد على تضاعف جينوماتها وتوريثها الى الأحياء الناتجة من الانقسام والتكاثر المعروف بالانتقال العمودي. وتتم عمليات النقل الأفقي Horizontal or lateral gene transfer في الأحياء بدائية النواة بطرق الاقتران والتحول Transformation، والتبنيغ Transduction الذي يتم بواسطة العاثيات . تساعد عمليات الانتقال في تبادل المواد الوراثية في بيئات مختلفة ومتباعدة. اما في حالة الاقتران والتحول فتحتاج الخلايا المستهدفة ان تكون بتماس مع الجزيئات المنقولة اليها اي انها يجب ان تكون في البيئة نفسها.

توزيع وضرورة الجينات في الجينومات البكتريا

يكون توزيع الجينات في الخلايا بدائية النواة ومنها البكتريا متعلقا بعدد من العمليات منها التضاعف والانحراف Skew (التي سيأتي ذكرها لاحقا)، ولكن الملاحظة العامة ان ضرورة الجينات وليس التعبير عنها هو الذي يملئ توزيع الجينات على الجينوم والأشرطة المتضاعفة وهذا ما لوحظ في كل من النماذج الدراسية *E. coli* (بكتريا سالبة لصبغة كرام) و *Bacillus subtilis* (بكتريا موجبة لصبغة كرام). وقد وجد ان الجينات الضرورية تميل ان تكون على الشريط القائد (أحد أشرطة DNA المفردة التي توجد في عملية التضاعف) وان كانت هذه الملاحظة ليست عامة اذ أثبتت دراسة التماثل الجزيئي Homology ان الجينات الضرورية تقع على الشريط المتخلف حتى في الجينات الضرورية التي لا يعبر عنها بشكل كبير ويمكن ان تشذ عن هذه القاعدة بعض البكتريا التي تتطفل داخل الخلايا.

وفكرة ضرورة الجينات هي فكرة عامة نظرا لان العديد من الجينات قد تعد غير ضرورية ولكن تحتاجها الخلايا تحت ظروف معينة، ولذلك تحدد الضرورة عند وضع الكائن في وسط غني جدا بكل المتطلبات الغذائية ومتطلبات الحياة الأخرى ، وتوفر تواليات الجينومات وتقنيات الطرق السريعة مكنت من التعرف على عدد من الجينات الضرورية في العديد من البكتريا كما في استعمال التطفير بالقافزات Saturation transposon mutagenesis واستعمال تقنيات الإسكات بـ Antisense RNA خاصة في البكتريات او الأحياء الأخرى التي يصعب

زراعتها على الأوساط الغذائية ولكن مع هذا فان هذه الطرق قد لا تعطي نتائج مضبوطة بدرجة عالية.

وتعد عملية تحديد ضرورة الجينات وتوزيعها مهمة لعدة أغراض منها البحث عن أهداف جديدة للمضادات الحيوية وكذلك تعد أساسية في دراسة عمليات التطور، كما ان نمط فقدان الجينات الضرورية تكون مهمة في فهم التطور المرافق الذي يحصل للبكتريا بعد استقرارها في مضاييف معينة.

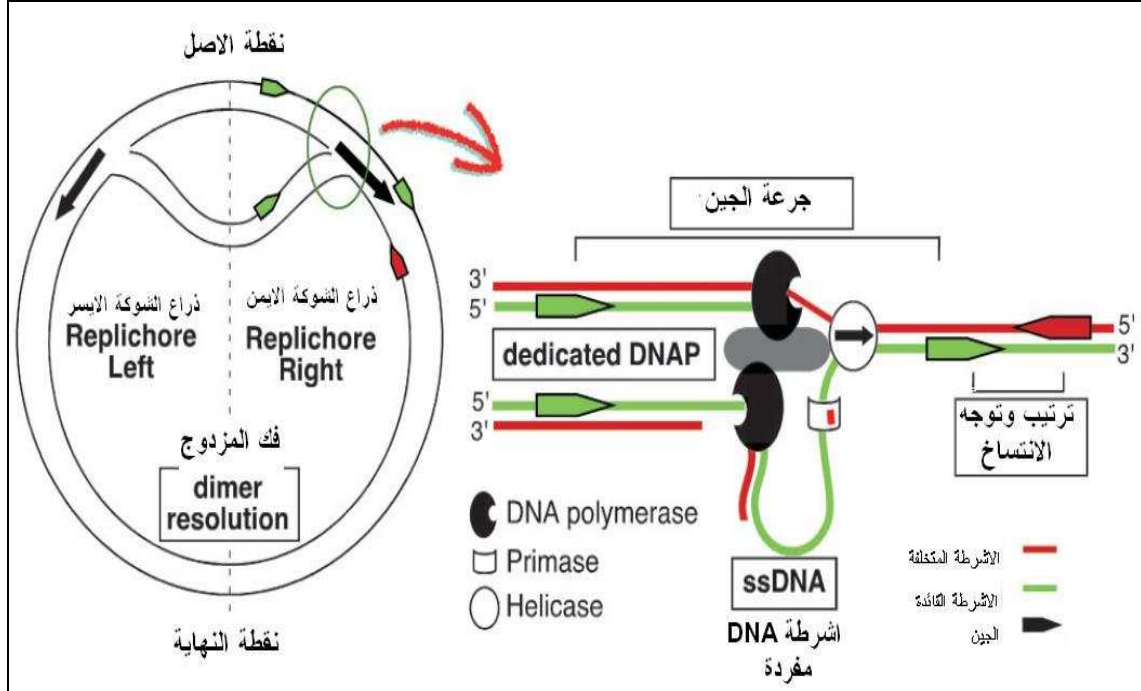
التواليات المتكررة في جينومات البكتريا

تحتوي بعض جينومات البكتريا على تواليات مكررة وان كانت قليلة مقارنة بجينومات الخلايا حقيقية النواة وتكون على شكل قطيقات توجد في مواقع متعددة تختلف في الحجم ونوعية القواعد النتروجينية والتعقيد ونمط التكرار ويطلق على البعض أحيانا Variable (VNTR) number tandem repeat (VNTR).

والمعقدة منها تسمى Multilocus variant repeats (MLVA) البعض منها يعمل بآليات خاصة للسيطرة على التعبير لبعض البروتينات الغشائية، كما ان البعض منها يعمل في مهمات الجينات الخاصة بالتشفير للمستضدات السطحية. ويستعمل بعضها في التشخيص الوبائي لعدد من البكتريا. وتصل نسبة المكررات في بعض البكتريا الى 25% من الجينوم وتعد مناطق غير مشفرة.

تضاعف جينومات بدائية النواة

درست في البكتريا كنموذج لبدائية النواة، ويبدأ عند نقطة الأصل Ori. وعملية التضاعف تنظم بدرجة عالية عند بدء العملية وترتبط بشدة بالكتلة الخلوية. وبعد عملية البدء يتم التضاعف بالاتجاهين الى حين الوصول الى نقطة النهاية Ter كما موضح في الشكل 35.3:



شكل 35.3: تضاعف جينوم البكتريا الحلقي

وتختلف سرعة التضاعف بين الأجناس والأنواع البكتيرية بشكل كبير جدا فمثلا في البكتريا *E. coli* يكون تقدم شوكة التضاعف بسرعة حوالي 1000 نيوكلويتيد \ ثانية في حين تكون في *Pyrococcus abyssi* بحدود 300 نيوكلويتيد \ ثانية وتنخفض الى 100 نيوكلويتيد \ ثانية في *Mycoplasma capricolum* ولذلك فان جينوم الأخيرة وان كان اصغر من جينوم *E. coli* بحوالي 7 مرات الا انه يأخذ وقت أطول لإكمال التضاعف. وارتباط بروتينات صغيرة في مواقع الانتهاء توقف حركة الشوكة ويتم فك الكروموسومات الجديدة الناتجة بمساعدة مجموعة اخرى من البروتينات.

شوكات التضاعف

ان عملية استطالة الأشرطة المخلفة حديثا يستوجب إزاحة شوكة التضاعف واذرعها Replichores على طول الكروموسوم ، وتكون بداية شوكة التضاعف بمثابة معقد من إنزيم التضاعف DNA polymerase (DNAP) بوحداته المختلفة الذي يقوم بعضها بعملية الكوثرية وأخرى تضمن الاهتمام بإضافة القواعد النيتروجينية الصحيحة وكذلك إنزيمات فتح الاشرطة . Helicases

ونظرا لكون كوثرية DNA تتم بالاتجاه من 5' الى 3' فان أحد الأشرطة وهو ما يسمى بالشريط القائد تتم مضاعفته بشكل مستمر اما الشريط الآخر فتتم مضاعفته بشكل متقطع وهو ما يسمى بالشريط المتخلف باستعمال شوكة التضاعف وإنتاج قطع او كازاكي. والأخيرة يتم تخليقها باستعمال بوادئ مكونة من 10-12 نيوكلويتيد من RNA يتم تخليقها باستعمال إنزيم Primase ، يليه ان الإنزيم المسئول عن كوثرية الأشرطة المتخلفة ينتقل من النهاية 3'-OH

من قطعة او كازاكي المكونة توا الى نهاية البادئ الجديد مؤدية الى تخليق قطع من DNA تتراوح من 1000-2000 نيوكلوئيد في *E. coli* وقد تكون اقل في أنواع أخرى وتنتهي الدورة بإزالة بوائى RNA ومليء الفجوات. وهذا يعني انه يوجد اثنين من DNAP كل واحد يضاعف أحد الأشرطة .

الانحراف في مكونات الأشرطة

مما ذكر أعلاه (وبشكل مبسط) يشير الى عدم تناظر عملية تضاعف الجينوم البكتري وهذا يؤدي الى الانحراف في مكونات الأشرطة المتضاعفة ولعل أهم مظاهرها هو انحراف GC skew الذي يمثل

$$\text{انحراف GC} = \frac{C - G}{C + G} \quad \text{GC skew}$$

وبالمقابل يمكن ان يحصل انحراف في القواعد الأخرى AT skew

$$\text{انحراف AT} = \frac{A - T}{A + T} \quad \text{AT skew}$$

ويحدث بكثرة في المناطق القريبة من بدء التضاعف ومناطق بدء الانتساخ. وتستعمل الظاهرة في التنبؤ بمناطق أصل تضاعف والانتهاؤ للجينومات وكذلك التنبؤ بمناطق بدء الانتساخ (TSS) Transcription start site. ويعزى هذا الانحراف الى بقاء الأشرطة لمدة من الزمن بشكل مفرد مما يعرضها لحصول الطفرات والتي تكون مختلفة المعدل بين الأشرطة القائدة والأشرطة المتخلفة ، كما ان التصادم بين DNAP والإنزيم الذي يقوم بعملية الانتساخ RNAP يشكل أحد الأسباب فضلا عن تعدد نقاط بدء تضاعف الكروموسومات. ويمكن ان تشارك ظاهرة تفضيل استعمال الشفرات الوراثية في هذه الانحرافات والذي يلاحظ من اختلاف الأحياء في مدى حصول الانحراف فيها. ومن الإمكانات الأخرى المشاركة هو الاختلاف في سعة التصحيح للإنزيمات القائمة بعملية التضاعف واتجاه عملية التضاعف، ومن المتوقع ان تكون هناك أسباب أخرى لا مجال لتفصيلها.

جينومات الخلايا حقيقية النواة

تمتاز جينومات الخلايا حقيقية النواة بكبرها فضلا عن تفرقها بين الجينوم النووي الذي يوجد في تركيب النواة المحاطة بغشاء يتواصل مع السايوبلازم لنقل المواد من وإلى النواة، وتحتوي أيضا على النوية وتحتوي على أرضية ليفية ومعدات من DNA والبروتينات. وجينوم العضيات التي فيها مثل المايكوكوندريا والبلاستيدات . وقد ذكر سابقا ان الجينات النووية لهذا النوع من الخلايا تكون موزائكية نظرا لوجود الانترونات والاكسونات. اما المايكوكوندريا فلها جينومها الخاص بها وتستخدم شفرات تختلف قليلا عن الشفرات النووية وتقوم بعملية الترجمة الخاصة بها بمساعدة بروتينات خاصة بها التي يشفر لها بواسطة DNA الموجود في النواة . كما ان جينومها لا يحوي على جينات خاصة بـ 5S rRNA ، ويتراوح عدد الجينات في

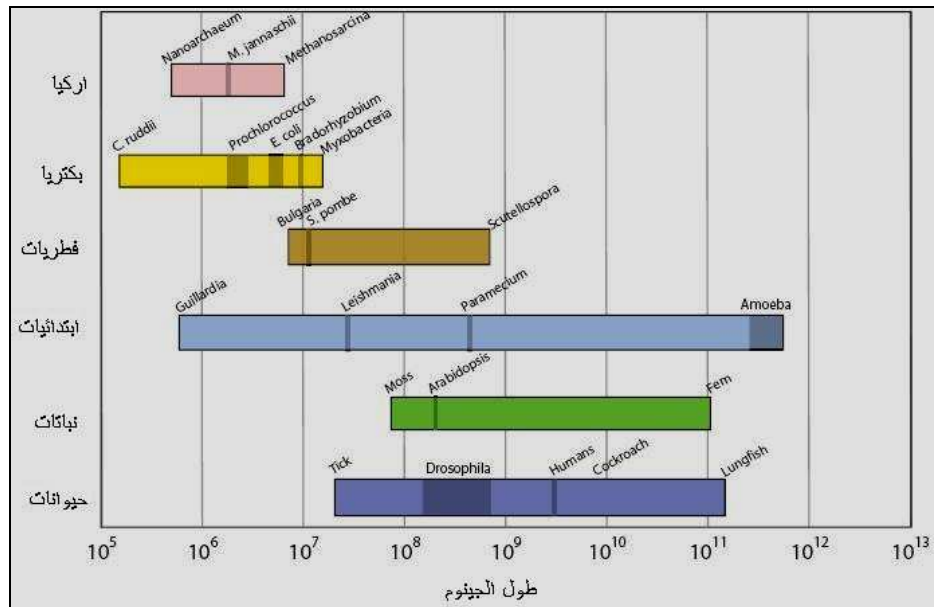
الميتوكوندريا الخاصة بالـ tRNA بين صفر - 27 وأي tRNA مفقود يتم توريده من الجينوم النووي.

وقد مر وصف وتشريح الجينات التي تكون عادة حاوية على الانترونات الأطول من الاكسونات ومر ذكر ان الطفرات سواء في الانترونات او الاكسونات او مناطق السيطرة الفاصلة بينهم المؤثرة في التعبير عن الجينات ، وتتم معالجة النسخ RNA بعدة طرق .

ومن الفروق بين جينومات بدائية النواة وحقيقية النواة ان 5'-cap الذي يضاف لجزيئات RNA المعالجة يقوم مقام rbs في بدائية النواة لتوجيه mRNA الى مواقع الارتباط على الرايبوزوم لبدء الترجمة من اقرب منطقة لشفرة البدء AUG وعليه فان التوالي في mRNA الذي يلي شفرة الميثايونين هو الذي يترجم.

وتختلف جينومات حقيقيات النواة في الحجم فتبدأ من حوالي ضعف معدل حجم جينوم البكتيريا الى مئات الأضعاف والشكل التالي (36.3) يعطي فكرة عن أحجام الجينومات.

كما ان الحجم يختلف بين أنواع الجنس الواحد، وكما موضح في الشكل فالمناطق المظلمة تشير الى الجينومات التي فيها اختلاف للجنس الواحد فمثلا جينومات ذبابة الفاكهة يختلف وفق المنطقة الجغرافية ، فالأنواع التي تعيش في مناطق ذات مستوى ارتفاع قريب من مستوى سطح البحر تحوي على جينومات اكبر وانوية اكبر مقارنة بتلك الأنواع التي تعيش بمستويات مرتفعة عن سطح البحر التي تكون جينوماتها صغيرة وكذلك انوية خلاياها صغيرة وهذا له علاقة بكمية الأوكسجين التي تنفذ الى الخلايا ويعتمد على ارتفاع المناطق .



شكل 36.3 : أحجام جينومات بعض الأحياء

المكررات في جينومات حقيقية النواة

تعد التواليات المكررة من ميزات جينومات الخلايا حقيقية النواة وهذه تكون غير مشفرة وتوجد بوفرة وتتعلق بحقيقة ان التواليات المشفرة تتناسب عكسيا مع تعقيد الكائن الحي . ومعظم التواليات غير المشفرة تشارك في عمليات تنظيم التعبير الجيني، فمثلا السمكة الكروية *Fugu rubripes* يتكون جينومها من 365 ميكاقاعدة يكون 10% من الجينوم مشفرا للبروتينات وهناك 22% انترونات والباقي في المناطق بين الجينات. ويلاحظ ان النسبة تتخفض في الإنسان الى حوالي 1.5% بالنسبة للتواليات المشفرة الذي يعد أكثر الأحياء تعقيدا. وتوجد بعض التواليات المتكررة في الأحياء بدائية النواة ايضا ، وفي احد التجارب على *E. coli* تم تجهيز هذه التواليات بما يناسبها للتعبير عن البروتينات وأدى ذلك الى موت الخلايا لذلك سميت بالمواد المعتمدة Dark matter. وأعطيت المكررات عدة تسميات تكون في بعض الأحيان متداخلة ومنها:

• المكررات المترادفة القصيرة (STRs) Short tandem repeats

وتتكون من تكرار 1 او 2 او 3 او 4 من القواعد المترادفة التي تكرر من 3 الى حوالي 10 مرات واحدة جنب الأخرى مثل: **ATTCATTCATTC** وتسمى ايضا التتابع الدقيقة ، وعندما يكون التكرار لاثنتين من النيوكليوتيدات تسمى المكررات ثنائية النيوكليوتيدات Dinucleotide repeats كما في جزر CpG ، وإذا كانت ثلاثة تسمى المكررات ثلاثية النيوكليوتيدات Trinucleotide repeats وهكذا.

• المكررات المتغيرة (VNTR) Variable number tandem repeats

مكررات يكون العدد المكرر منها غير معروف ومثل هذه المكررات تستعمل في الفحوص الوراثية لتحديد الأبوة والنسب وكذلك تستعمل للأغراض الجنائية. نظرا لان المكررات تكون حزم خاصة عند استعمال بصمة (DNA fingerprinting) DNA اعتمادا على الطول والمسافة التي تقطعها وبذلك تكون مختلفة بين الأشخاص، وبصمة DNA لا تعني التشابه بين الجينات .

ويمكن ان تقسم المكررات الى أقسام أخرى اعتمادا على انتشارها في الجينوم.

- المكررات المنتشرة Interspersed repetitive sequences وتنتشر في كل جينوم الخلايا حقيقية النواة وتساعد في نشوء الجينات الجديدة.
 - مكررات الكروموسوم Intrachromosomal repetitive sequence وهي الموزعة على الكروموسوم الواحد.
 - مكررات الكروموسومات Interchromosomal repetitive sequences وتكون موزعة على عوائل مختلفة.
- وهناك طريقة أخرى لتصنيف المكررات تعتمد على تكرارها في الجينوم ومنها:
- مكررات النسخة الواحدة Unique / single copy وتوجد مرة واحدة او مرات قليلة \ جينوم.
 - المكررات المعتدلة Moderately repetitive sequence ويوجد منها 10-10⁵ نسخة \ جينوم.

■ المكررات الكثيرة Highly repetitive sequence توجد أكثر من 10^5 نسخة \ جينوم. وأكثر المكررات دراسة هي المعتدلة. وهناك عدداً من الطرق للكشف عنها، والطرق الحديثة منها تعتمد على تضخيم القطع الحاوية عليها بتقنيات PCR المختلفة تم الترحيل الكهربائي على الهلام.

أطراف الكروموسومات Telomeres

مجموعة أخرى من التواليات المكررة توجد عند أطراف الكروموسومات المفتوحة. وفي الخلايا بدائية النواة التي تكون معظمها حاوية على كروموسومات حلقية لا تحتوي على هذه المكررات لذلك تكون دائمة التضاعف عند توفر الظروف الملائمة ، ولكن بعض الأجناس والأنواع مثل جنس *Streptomyces* والحلزونيات *Borrelia* تكون بعض عناصر جينوماتها مثل الكروموسومات والبلازميدات مفتوحة لذلك تمتلك أطراف كروموسومية ولكنها تختلف عن تلك الموجودة في الخلايا حقيقية النواة ، ويتم حماية الأطراف بواسطة معقدات بروتينية أو تكوين عروات الماشة من أشرطة مفردة عند نهاية الكروموسومات المفتوحة. ومكررات الأطراف مختلفة في الإنسان تكون الوحدة الأساسية مكونة من مكوثر سداسي Hexamer هو TTAGGG (3' → 5') الذي يتكرر بمعدل 2000 نسخة ليصل الى حوالي 15 كيلوقاعدة ولكن في عموم الأحياء يكون طولها بين 300-600 قاعدة والبعض منها في الأحياء الأخرى موضحة في الجدول 5.3 :

جدول 5.3 : مكررات أطراف الكروموسومات Telomeres في بعض الأحياء

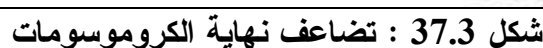
الكائن الحي	مكررات الأطراف 5' إلى 3'
Human, Mouse, Xenopus	TTAGGG
<i>Neurospora crassa</i>	TTAGGG
<i>Physarum, Didymium</i>	TTAGGG
<i>Dictyostelium</i>	AG (1-8)
<i>Trypanosoma, Crithidia</i>	TTAGGG
<i>Glaucoma Tetrahymena,</i>	TTGGGG
<i>Paramecium</i>	TTGGG (T/G)
<i>Oxytricha, Stylonychia, Euplotes</i>	TTTTGGGG
<i>Plasmodium</i>	TTAGGG (T/C)
<i>Arabidopsis thaliana</i>	TTTAGGG
<i>Chlamydomonas</i>	TTTTAGGG
<i>Bombyx mori</i>	TTAGG
<i>Ascaris lumbricoides</i>	TTAGGC
<i>Schizosaccharomyces pombe</i>	TTAC(A)(C)G(1-8)
<i>Saccharomyces cerevisiae</i>	TGTGGGTGTGGTG(from RNA template) or G ₍₂₋₃₎ (TG) ₍₁₋₆₎ T
<i>Saccharomyces castellii</i>	TCTGGGTG
<i>Candida glabrata</i>	GGGGTCTGGGTGCTG
<i>Candida albicans</i>	GGTGTACGGATGTCTAACTTCTT
<i>Candida albicans</i>	GGTGTA[C/A]GGATGTCACGATCATT
<i>Candida maltosa</i>	GGTGTACGGATGCAGACTCGCTT
<i>Candida guilliermondii</i>	GGTGTAC
<i>Candida pseudotropicalis</i>	GGTGTACGGATTTGATTAGTTATGT
<i>Kluyveromyces lactis</i>	GGTGTACGGATTTGATTAGGTATGT

وفضلاً عن مكررات القواعد المشار إليها هناك العديد من البروتينات المرافقة مثل TRF1 (Telomeric repeat binding factor 1) و TRF 2 و hRNP 1 وغيرها العديد التي تتناوب الوظائف والوجود في الخلية اعتماداً على المرحلة التي تمر بها من دورة الخلية وقد تكون مرتبطة بجزء الشريط المزدوج أو نهايته المفردة وبمجمليها يطلق عليها معقد الحماية Shelterin complex .

وتكون تواليات الأطراف بشكل أشرطة مزدوجة وينتهي أحد أطرافها بسلسلة مفردة 3' . والمكررات تكون إما سداسية أو ثمانية غنية بالكوانين والسلسلة المفردة المعلقة تكون مهمة في الحفاظ على طرف الكروموسوم، والتواليات تكون عروة تسمى T-loop ويساهم الجزء المفرد منها في تكوين عروة الإزاحة D-loop . وتكوين هذه العروات يحمي الأطراف من تأثير إنزيمات القطع الخارجية .

من أهم البروتينات ذات العلاقة بأطراف الكروموسومات والإنزيم يمثل معقد Ribonucleoprotein مكون من بروتينات وبادئ من RNA ، يكون الإنزيم معطلا في معظم الخلايا الجسمية ولكنه يكون نشطا في الخلايا الجنسية والخلايا الجذعية وبعض خلايا الدم وبصيلات الشعر، يعمل على إتمام مضاعفة الكروموسومات حتى نهاياتها ، وهو إنزيم نسخ عكسي.

تُعاني كروموسومات الخلايا حقيقية النواة من القصر بعد كل انقسام ويؤدي إلى تأكلها وهذا يؤثر في العشرات من وظائف الخلية. ويعتقد أن السبب في تآكل أطراف الكروموسومات يعود إلى طريقة تضاعف DNA وخاصة الشريط المتخلف وتكوين قطع أوكازاكي، حيث تملأ الفجوات الناتجة بعد إزالة بوائى RNA بالـ DNA ما عدا القطعة الأخيرة لذلك تفكك هذه المنطقة عند النهاية 5' من الشريط المتخلف بإنزيمات القطع الخارجي كما موضح في الشكل 37.3:



والفقد النظري يكون بحدود 20 قاعدة لكل دورة ولكن واقع الحال يشير الى فقدان يصل من 50-210 نيوكلوتيد في كل دورة انقسام وذلك لان الأطراف حساسة جدا للإجهاد التأكسدي الذي لا مفر منه في الأنظمة الحيوية وعندما يصل طول الطرف الى حد حرج يتوقف انقسام الخلايا وقد حدد Leonard Hayflick عدد مرات انقسام الخلايا وسميت "محدودية هايفليك" Hayflick limit. لذلك فان طول طرف الكروموسوم قد تكون له علاقة مع هرم وشيخوخة الخلايا وكذلك السرطانات (كما سيرد ذكره لاحقا). فمثلا بعض خلايا الدم تكون فيها الأطراف بحدود 8000 نيوكلوتيد عند الولادة ولكن بتقدم العمر تصل الى 3000 وبعض الأحيان يصل الى 1500 قاعدة ، وهذا يعني ان الخلايا يمكن ان تعاني 50-70 انقسام، وهذا يعتمد على الكائن، فمثلا محدودية هايفليك في خميرة الخبز تصل الى 25-30 انقسام.

وظائف أطراف الكروموسومات

يتضح مما سبق ان الأطراف تؤدي وظائف مهمة للخلايا منها:

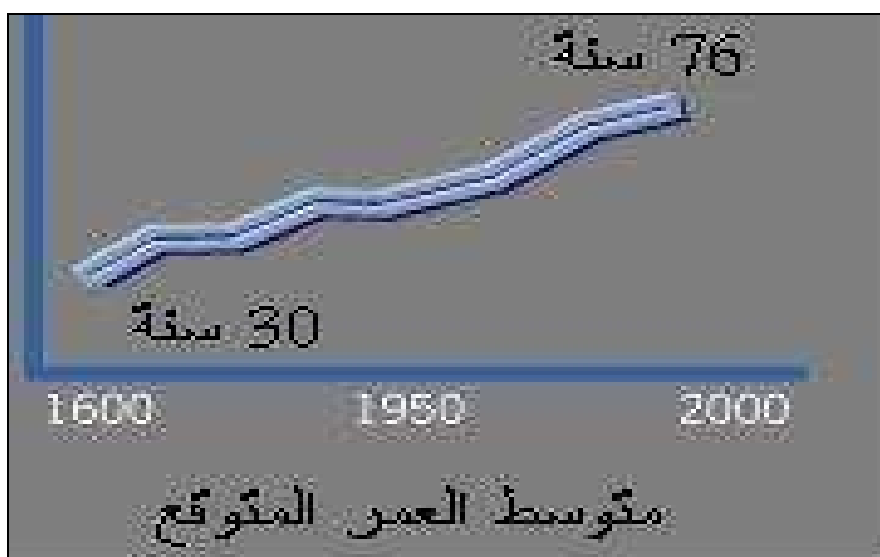
- حماية الجينات التي تقع على المناطق الطرفية من الكروموسومات لذلك فهي تكون بمثابة المضحي لعمليات التقصير التي هي ميزة أساسية لعمليات التضاعف وتحمي ما بعدها من الجينات اي انها تكون بمثابة تعويض والحفاظ على الجينات من الاضطراب بعد كل انقسام.
- تحمي أطراف الكروموسومات من التأشب المتماثل (HR) وربط النهايات غير المتماثلة Non-homologous end joining (NHEJ) وذلك لان الكروموسومات بدون تغطية هذه الأطراف ستعدها الخلايا بمثابة كسور في DNA وتقوم بإصلاحها ولصقها مع بعضها مما يؤدي الى دمج الكروموسومات وبالتالي يفقد الكائن البصمة Blueprint الخاصة به وخراب وظائف الخلايا وموتها ولذلك فهي تمثل دوايرء لنهايات الكروموسومات التي تستهلك أثناء الانقسامات لذلك يطلق عليها Disposable buffers.
- قصر الأطراف يكون مسئولا عن شيخوخة الخلايا على المستوى الخلوي اي تصبح غير قادرة على الانقسام ويمكن ان تدخل مسار الاستماتة Apoptosis او تحصل فيها تغيرات تحولها الى خلايا سرطانية. فضلا عن ان القصر يؤدي الى عدم ثبوت الجينوم.

المظاهر المتعلقة بأطراف الكروموسومات

مما ذكر أعلاه ان إنزيم Telomerase له الأثر الأكبر بطول أطراف الكروموسومات لذلك فان نشاطه يمنع تآكل الأطراف ولذا كان هدفا للتفكير في إطالة العمر ومعالجة السرطانات ببعض الأدوية التي تنشطه ، كما ان قلته او انخفاض نشاطه يؤدي الى فقدان الخلايا المناعية التائية T-cells لذا يمكن جعله هدفا في علاج بعض أمراض المناعة الذاتية. وقد وجد ان الإنزيم ينشط في حوالي 90% من الأورام ولكن الاستطالة قد تكون بعض الأحيان عن طريق آلية لا يشترك فيها الإنزيم . والإنزيم في حالة فعاليته يقوم بإضافة التواليات مما يساعد الخلايا في عبور محدودية هايفليك وعدم الدخول في مسار الهرم او الاستماتة لذا فان آليات التسرطن تكون قد غزت ودمرت هذه الوسيلة التدميرية المبرمجة (تقصير أطراف الكروموسوم) لتصبح الخلايا خالدة .

وقد وجد ان الأطراف المقصرة بشكل كبير يمكن ان تميز بالبروتين P₅₃ وهو أحد المكونات الفعالة في تنظيم دورة الخلية حقيقية النواة، ويتداخل البروتين مع العروة T لذلك فان فقدان او

عطب P_{53} يمكن ان يؤدي الى الأورام كما وجد في بعض الحالات السرطانية ان الخلايا تكون قصيرة الأطراف . ومن هنا يصبح من غير الممكن ان تعزى أسباب تقصير الأطراف الى هبوط فعالية إنزيم Telomerase وانما هناك أسباب أخرى منها الإجهاد التأكسدي (كما ورد ذكره آنفا) الذي يعد من أكثر أسباب الهرم، والجذور الحرة لا يمكن تجنبها وذلك انها تنتج بشكل طبيعي أثناء التنفس وكذلك في الالتهابات والإصابات بالأحياء الممرضة وتناول الكحول والتدخين، فضلا عن إضافة السكريات Glycation الآتية من الطعام ويقوم الكلوكوز وغيره من المواد بالارتباط ببعض البروتينات المرتبطة بالمواد الوراثية وإيقاف عملها، لذلك كانت قلة السرعات تطيل العمر. كل هذه الأسباب وغيرها يمكن ان تؤدي الى إطالة العمر كما يلاحظ في الشكل (38.3) ، اذ يلاحظ ان معدل العمر في القرن السابع عشر كان حوالي 35 سنة زاد في نهاية القرن العشرين الى 76 سنة والمتوقع ان يزيد الى 90 سنة وفي حالة القضاء على الإجهاد التأكسدي يتوقع ان يطول الى 1000 سنة ويكون الموت لأسباب أخرى ليست بايولوجية.



شكل 38.3 : متوسط عمر الإنسان المتوقع مع مرور الزمن

العوامل المؤثرة في أطراف الكروموسومات

يعد طول طرف الكروموسوم مؤشرا للعمر البايولوجي، فالملاحظ ان الأطراف القصيرة توجد في كبار السن ولذلك استعملت للأغراض الجنائية. وان الإجهاد التأكسدي يزيد من تقصير الأطراف لذلك كانت للمواد المضادة للأكسدة مثل فيتامين D أثر في الحفاظ على أطراف الكروموسومات فضلا عن ان تقليل الالتهابات والإصابات تؤدي الى النتيجة نفسها وتمنع تآكل الأطراف . كما ان طول الأطراف يتأثر بالجنس فهي في النساء أطول مما في الرجال ولذلك كانت النساء أطول عمرا" من الرجال في العموم.

أهتم البعض بإطالة العمر ومعالجة بعض السرطانات عن طريق إنزيم النسخ العكسي للأطراف وذلك عن طريق تعطيله مرة وتنشيطه مرة أخرى وفق الهدف المطلوب ولكن هذا له مخاطره

على الأقل في الوقت الحاضر الذي لا تزال العديد من الجوانب الحيوية غير واضحة، فإن غلق الإنزيم لمعالجة السرطان يمكن ان يؤدي الى العقم، ومنع النثم الجروح وعرقلة إنتاج بعض خلايا الدم والجهاز المناعي.

ومن جهة ثانية فان تنشيط الإنزيم يمكن ان يتسبب في تغيرات منها تثبيت التغيرات غير المرغوب فيها الحاصلة في الكروموسومات، وتشجيع نمو النسائل الخالدة ، إضفاء المقاومة للاستماتة عند الحاجة اليها كما انها يمكن ان تؤدي عند عدم التغطية ولو طرف على الأقل الى حصول تبادل كروماتيدي شقيقي او الاندماج الكروماتيدات الشقيقة Sister chromatid fusion.

العناصر القافزة Transposable elements

تحتوي جينومات الخلايا حقيقية النواة على العناصر القافزة كما هو الحال في ممالك الأحياء الأخرى مثل البكتيريا. وأول اكتشاف للقافزات كان في نبات الذرة، ثم تبين انها شائعة في الكائنات الباقية. وتعرف القافزات على انها قطع من DNA يمكنها الانتقال من مواقع محددة في الجينوم او بين مكونات الجينوم وبذا فهي تعد من المطفرات لان استقرارها في بعض الجينات يمكن ان يؤدي الى فقدان وظيفة الجين.

تصل نسب القافزات الى حوالي 50% من جينوم الإنسان في حين تصل هذه النسبة الى 70% في بعض أنواع الحشائش. وقد أدى اكتشافها الى تغيير بعض المفاهيم الوراثية لتبتعد بها عن المنحى التقليدي. وهناك بعض المفاهيم حول القافزات منها:

- تكون قادرة على التحرك داخل الكروموسوم او بين الكروموسومات ويمكن ان يكون استئصالها وإعادة إقامتها غير منضبط.
- تغير من التعبير الجيني الذي تستقر فيه.
- بعض العناصر تبقى في حالة سكون في الجينوم ويمكن إعادة تنشيطها ببعض الوسائل الحيوية او غير الحيوية وهذه الحالة يطلق عليها الإجهاد الجينومي Genomic stress .

أنواع القافزات

تصنف القافزات في الخلايا حقيقية النواة الى صنفين رئيسية اعتمادا على عمليات قفزها (التي سيأتي ذكرها لاحقا):

- الصنف الأول يكون قفزها بعد تكوين مركبات RNA وسيطة.
- الصنف الثاني الذي يقفز فيه DNA مباشرة اي ان القافز يتحرك بنفسه من موقع الى آخر.

وكلا الصنفان تحتوي على عناصر مستقلة وأخرى غير مستقلة Non autonomous ، فالأولى تمتلك أطر قراءة مفتوحة التي تشفر لمنتجات لازمة لعملية القفز.

* ويمثل الصنف الأول العناصر الارتدادية Retroelements التي تقسم بدورها الى مجموعتين اعتمادا على آلية قفزها وتركيبها ومنها LTR Retrotransposons ولها نهاية طويلة من المكررات (LTR) تصل الى حوالي 100 زوج من القواعد الى عدة كيلوات من القواعد . وتحتوي مجموعة الصنف الأول على جينات مثل gag و pol ، فالأخير يشفر لبروتينات مشابهة لقفصيات الفيروسات ، اما pol فيشفر لبروتين متعدد الوظائف. ان النسخ

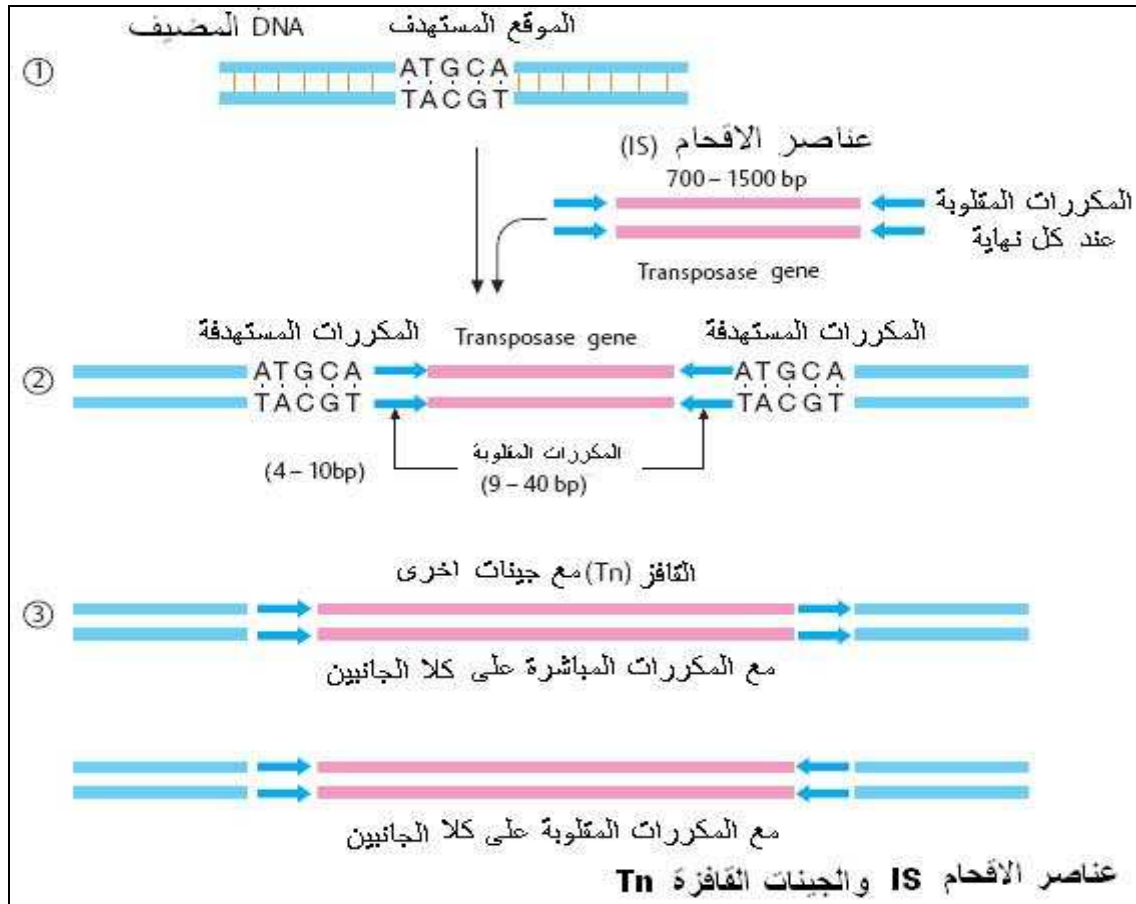
الناتجة من 5'LTR والمنتية بـ 3'LTR تنتقل الى السايكوبلازم وهناك تشغل دوراً مزدوجاً فهي اما تعمل كـ mRNA لتخليق البروتينات المذكورة تواءم لعمل كقالب لتخليق cDNA مزدوج الأشرطة ينقل الى النواة ويندمج مع الجينوم، وهذه المجموعة تؤلف الجزء الأكبر من جينومات النبات وهي المسؤولة عن الاختلافات الكبيرة في جينومات هذه المجموعة من الأحياء. اما المجموعة الثانية في الصنف الأول فهي Non-LTRs وهذه تقسم بدورها الى LINEs أنواع مستقلة يصل طولها الى 6500 قاعدة تنتشر في الجينوم ومنها L₁ في الجينوم البشري الغنية بـ AT، والأخرى غير مستقلة هي SINEs (Short interspersed nuclear elements) يصل طولها بين 90-300 زوج قاعدة التي يوجد منها 500,000 نسخة/الجينوم البشري.

والنوع الطويل يحوي على اثنين من أطر القراءة، الأول ORF₁ يعطي البروتين الرابط للـ RNA والآخر له فعاليات قطع الحوامض النووية الداخلية وإنزيم النسخ العكسي. وكل من LINEs و SINEs تنتهي بمكررات من توالي بسيط يكون في العادة من الأدينين (Poly (A).

• الصنف الثاني يكون بسيط التركيب DNA transposons وله تواليات نهائية مقلوبة قصيرة Terminal inverted repeat (TIR) يتراوح طولها بين 10-40 زوج قاعدة وقد تصل الى 200 قاعدة وتحوي على جين واحد يشفر لإنزيم القفز Transposase. ومن أبسطها تواليات الإقحام (IS) توالياتها النهائية المقلوبة تصل ما بين 9-13 زوج قواعد ويصل طولها الى 750-1500 قاعدة، وهناك أنواع معقدة تحمل بالإضافة الى الأساسيات تحمل بعض الصفات مثل المقاومة للمضادات الحيوية او مقاومة المعادن الثقيلة مثل Tn والذي توجد منه أنواع، وTns اما تكون مجنحة بمكررات مباشرة وبالاتجاه نفسه او مجنحة بمكررات مقلوبة اي بالاتجاه المعاكس كما موضح في الشكل (39.3).

ويكون الصنف الأول (Class I-non LTR) شائعاً في جينومات اللبائن وتشكل حوالي 5% من محتوى الجينومات من العناصر القافزة، اما في جينومات النباتات المزهرة ومنها ذوات الفلقة الواحدة مثل الحشائش والرز والذرة وبعض ذوات الفلقتين مثل رشاد الصخر والطماطة فتحتوي على تجمعات كبيرة من الصنفين وتشكل LTRs النسخ الأكبر في الأحياء المدروسة. وتختلف وظائف القافزات في الأحياء التي توجد فيها.

وتساهم القافزات في جينومات الأحياء المجهرية حقيقية النواة فخميرة الخبز يوجد فيها واحد من القافزات الارتدادية يسمى Ty فضلاً عن مكونات أخرى. والفطريات تم اكتشاف القافزات فيها حديثاً مقارنة بالأحياء الأخرى. فهذه المجموعة من الأحياء تنتشر فيها القافزات خاصة في المجتمعات الطبيعية لتساعدها في التطبع البيئي خاصة الأجناس التي ليس لها مراحل جنسية، والبعض منها نظراً لكثرة القافزات فيها طور أنظمة إسكات كأحد أنظمة الدفاعات ضد القافزات الغازية.



شكل 39.3: ترتيب وتركيب العناصر القافزة Tn

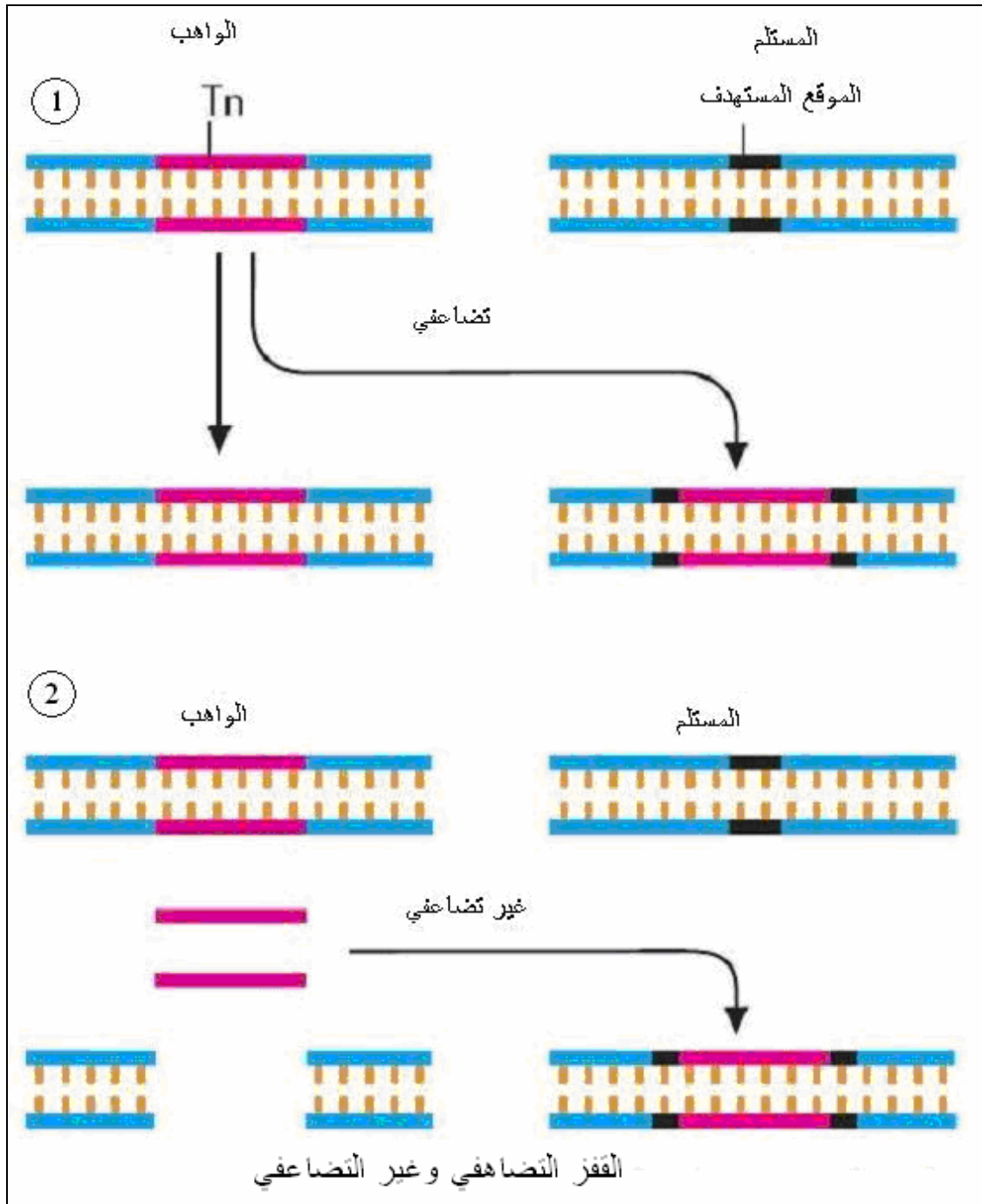
وأنظمة الإسكات وجدت في الفطريات التي لها ادوار جنسية مثل الفطريات الكيسية منها *Aspergillus nidulans* و *Neurospora crassa* التي تستعمل موديلات دراسية. ونظرا لأهمية الفطريات الصناعية والطبية زاد الاهتمام بدراسة هذه العناصر التي كشفت عن وجود عدداً من أنواع القافزات تغطي الأنواع المذكورة أعلاه وقد تصل الى 14% من جينومات بعض الأجناس واستغلّت هذه في تطوير السلالات، فضلاً عن ان تحديد توزيع القافزات في المجتمعات الطبيعية يعطي معلومات قيمة عن النواحي والاعتبارات البيئية والوبائية .

الفقر Transposition

يكون قفز العناصر بالتين هي القفز من مكان والإقحام في مكان آخر وهي آلية القطع واللصق Cut & paste والثانية يتم فيها تضاعف العنصر القافز ثم انتقال إحدى نسخه الى مكان آخر وتسمى آلية النسخ ثم اللصق Copy & paste.

الآلية الأولى تحصل في القافزات البسيطة التي لا تمر بمرحلة RNA وبعد تحديد الهدف الذي يقفز إليه العنصر يتم قطعه وانتقاله كوحدة فيزيائية الى DNA المستهدف دون ان يترك نسخة مكانه الذي اما ان يصلح في خلايا حقيقية النواة او يدمر كما في البكتيريا اذا كان هناك أكثر من نسخة من الكروموسوم. اما في الحالة الثانية فيتضاعف القافز ثم تدخل نسخة منه الى قطعة

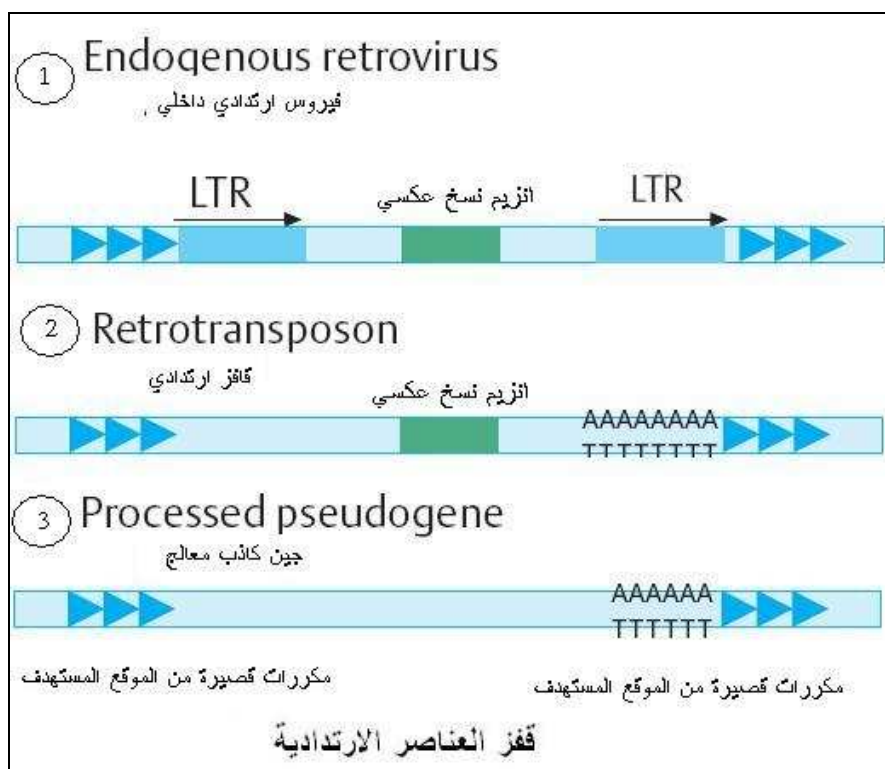
DNA المستلمة ويكون قد ترك نسخة منه في DNA الأصلي الذي جاء منه ، وهذا النوع من القفز يحتاج الى إنزيم القفز الذي يعمل على نهايتي الجين القافز الأصلي وإنزيم الحل Resolvase الذي يعمل على النسخ المتضاعفة، كما موضح في الشكل 40.3 :



شكل 40.3 : طرق قفز العناصر القافزة

قفز العناصر الارتدادية

هذه العناصر في الإنسان تشمل فيروس HIV وفيرسات الورم المكونة من RNA. وكخطوة أولى تحتاج هذه العناصر الى تخليق نسخ DNA من العنصر الارتدادي ثم إجراء عملية نسخ عكسي ثم إعادة إقحامه في DNA. وهناك ثلاثة أنواع منها هي الفيرسات الارتدادية الداخلية Endogenous retroviruses و القافزات الارتدادية ، والجينات الكاذبة المعالجة الموضحة تراكيبها في الشكل 41.3 :



شكل 41.3 : أنواع العناصر الارتدادية

فالمجموعة الأولى تكون بمثابة تواليات تشبه الفيرسات الارتدادية ويمكنها ان تصيب خلايا جديدة ولكنها محددة بجينوم واحد، اما المجموعة الثانية فتتقصرها LRTs والأجزاء الأخرى الخاصة بالفيرسات الارتدادية، والمجموعة الأولى والثانية تحوي على إنزيم النسخ العكسي لذلك تكون قادرة على القفز المستقل. اما المجموعة الثالثة وتسمى أيضا Retropseudogenes هذه لا تمتلك إنزيم النسخ العكسي لذلك لا تستطيع القفز بشكل مستقل ويمكن ان تستعمل إنزيمات الخلايا لهذا الغرض.

فعاليات وأهمية العناصر القافزة

وجود العشرات من العناصر القافزة وإدراك انها من القوى المهمة في تطوير الجينومات أدى الى إيجاد العديد من الطرق لاكتشاف الجديد منها فضلا عن تطوير طرق تحليلها ودراسة تداخلاتها مع المضيف.

ومن فعاليتها انها أدت الى تعقيد تركيب الجينومات وذلك بإعادة ترتيب الكروموسومات وإنتاج نسخ وتواليات جديدة بواسطة الإقحام في الجينات او الجينات المجاورة مؤدية الى التأثير في التعبير الجيني.

ويمكن ان يتطور الجينوم بإيجاد اكسونات جديدة التي تتم بعدة طرق:

- إدلاف الاكسونات او تضاعف الاكسون (وقد تكون هذه ليست لها علاقة بالقفزات ولكن تذكر هنا للضرورة)، وفيها يتم إقحام اكسون جديد في الجين الموجود أصلا بواسطة عملية التأشب، هذا فضلا عن تضاعف الجين المذكور توأً وعليه فان عمليات الخياطة البديلة Alternative splicing تعطي نسخا تحوي على الاكسون الجديد.

- تساعد في تفعيل الجينات الارتدادية، ويحدث في بعض الأحيان ان الجزيئات الناضجة من mRNA ان يعاد نسخها عكسيا بواسطة انزيم النسخ العكسي للعناصر القافزة ويعاد إقحامها في الجينوم الأصلي في المنطقة بين الجينات بواسطة عملية القفز الارتدادي منتجة ما يسمى بالنسخ الارتدادية Retrocopies وهي تكون خالية من الانترونات والممهدات والعناصر المنظمة التي تقع عادة الى يسار الجينات. وبعض النسخ الارتدادية يمكن ان تكتسب ممهدا "جديدا" وذلك بإقحام عنصر قافز الى يسارها ويمكن ان تنتسخ الى منتجات جين فعال، والجينات التي تنتج من النسخ الارتدادية تسمى الجينات الارتدادية كما موضح في الشكل (42.3).

- تساعد العناصر القافزة بتزويد الجينوم باكسونات جديدة للمضيف وبآليات مختلفة ومنها أكسنة Exonization العناصر القافزة. اذ يتم في البداية إقحام القافز في انترون الجين والبعض منها يحصل له طفرات تمكن ماكنة الخياطة للـ RNA من إقحام القطع الجديدة في جزيئات mRNA الناضجة. ومعظم عمليات الأكسنة تحدث في اكسونات الأولية (itexons) التي تخضع للخياطة البديلة والتي تحدث في خصى اللبائن بشكل رئيس. وعملية الأكسنة يمكن ان تحدث ليس فقط في الجينات المشفرة للبروتينات وانما في اي من الجينات التي تحصل فيها الخياطة البديلة.

- يمكن ان تقحم العناصر القافزة في المناطق غير المشفرة UTR مؤدية الى التأثير في تنظيم التعبير الجيني.

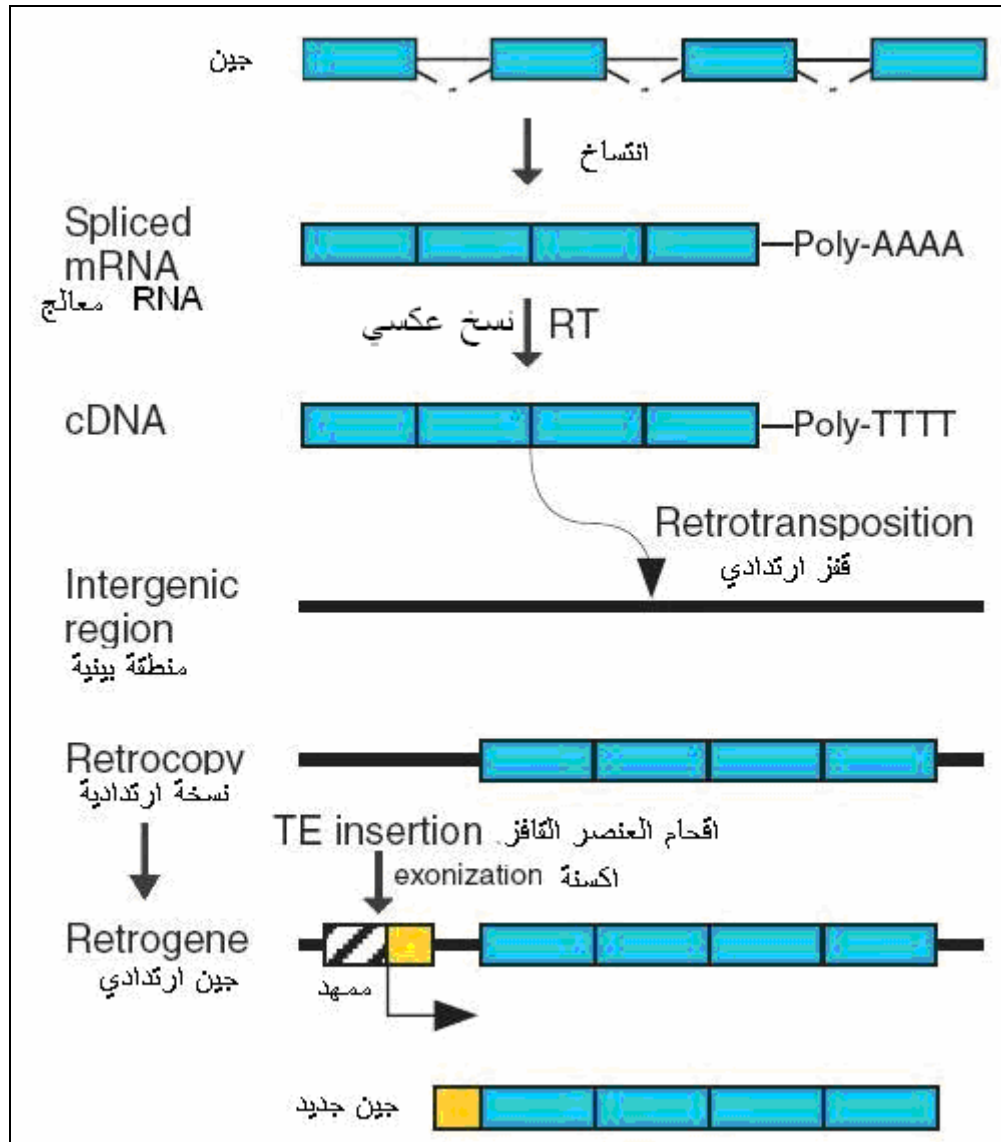
- كما ان تكاثر العناصر القافزة داخل الجينوم يؤدي الى زيادة التواليات المتكررة التي تساعد في:

0 تحفيز عمليات التأشب

0 تزيد من مواقع عمليات تنظيم الانتساخ

0 توفر مواقع لإضافة الأدينين

0 توفر إشارة خياطة Splicing



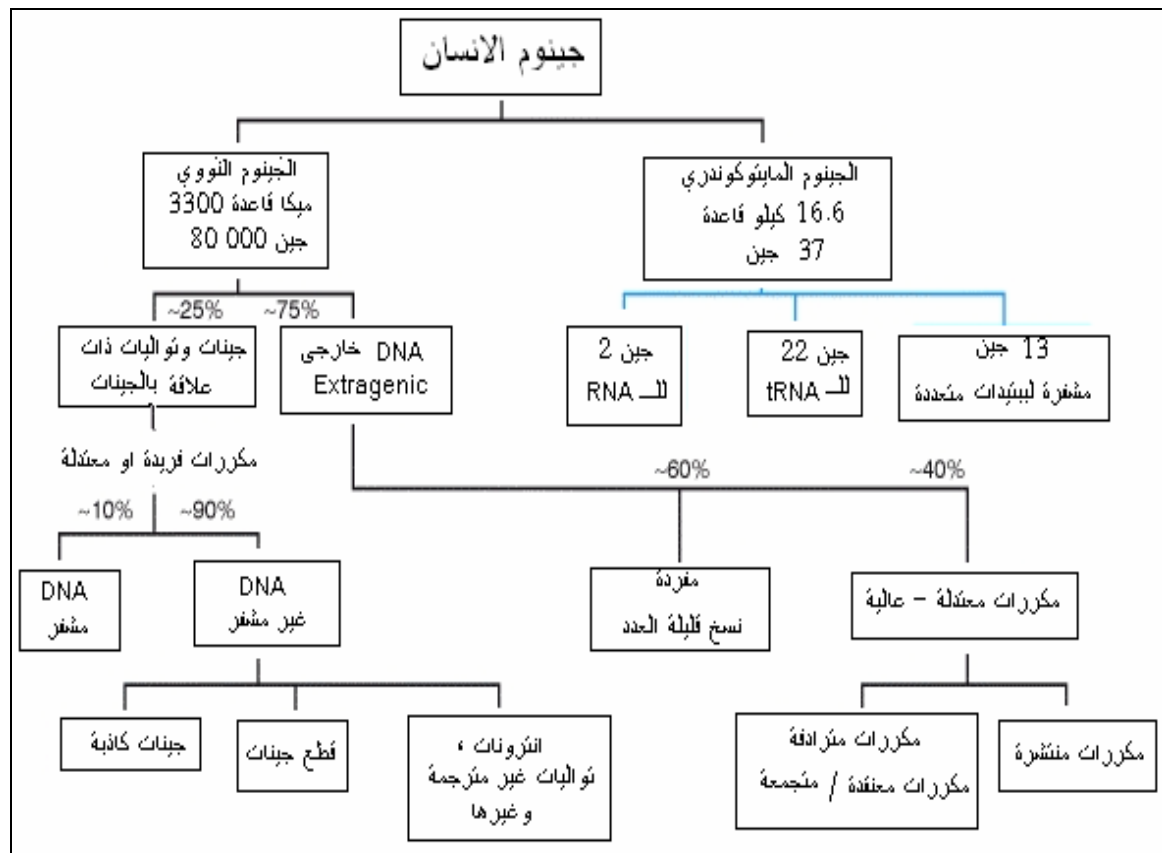
شكل 42.3 : تفعيل الجينات الارتدادية

جينوم الإنسان

عند الحديث عن جينومات الخلايا حقيقية النواة، لابد من التحدث عن الجينوم البشري الذي تم تحديد تواليه مطلع عام 2003 لما يمتاز به من التعقيد والفرق عن الكائنات الأخرى. جينوم الإنسان يحوي على 1.5-1.7 % من الجينات المشفرة للبروتينات و 98.5% هي تواليات غير مشفرة وخاصة من العناصر القافزة كما ورد في تقرير الجهات المختصة International Human Genome Sequencing Consortium عام 2001، وبهذا حل لغز C-value (C= Constant) وحل مصطلح C-value enigma محل C-value paradox، إذ إن الأخيرة تشير إلى كمية DNA في النواة الفردانية في حين الأول يوضح

نوع DNA غير المشفر ومن أين أتى والكيفية التي تنتشر فيها التواليات غير المشفرة والتعامل الزمني معها ووظائفها والتأثيرات التي تشارك فيها هذه القطع غير المشفرة سواء في الكروموسومات أو الانوية ثم في الخلايا ثم في الكائن.

والجينوم البشري يبلغ حوالي 3×10^9 قاعدة منها 17 كيلوقاعدة تخص المايكوكوندريا والباقي يمثل الجينوم النووي. والملاحظ أن الجينوم البشري أكبر من الجينوم البكتري آلاف المرات ولكنه أصغر مئات المرات من جينوم بعض أنواع الأميبا. ومن الإحصائيات الأخرى أن خلايا الإنسان تحوي عدد من الجينات يبلغ خمس أضعاف عدد الجينات البكتريا *Bacteroides fragilis* التي تقطن أمعاء الإنسان وإذا أخذ بنظر الاعتبار كل الأنواع البكتيرية الموجودة في أمعاء الإنسان التي تتراوح بين عدة مئات إلى أكثر من 40,000 نوع مكونة عدداً يصل إلى 100 تريليون خلية يلاحظ أن الجينات البشرية تشكل 1% والباقي تساهم به البكتريا بشكل رئيس فضلاً عن الأحياء الأخرى. في المناطق غير المشفرة المذكورة أعلاه هناك حوالي 62% منها مناطق بين الجينات تعج بتواليات غير مشفرة سواء طويلة أو قصيرة ، اما 36% فتشكل الانترونات و UTRs قبل وبعد الجين والجينات الكاذبة وقطع الجينات. ويمكن إجمال توزيع الجينوم البشري بالشكل 43.3 :



شكل 43.3: توزيع تواليات DNA في الجينوم البشري

وتمتاز جينات الإنسان بوجود الانترونات التي تكون أكبر من الاكسونات فضلا عن تكرار الأولى والمسجل ان أطول جين في الإنسان هو الذي يشفر للـ Dystrophin الذي يصل طوله الى 2.4 ميكاقاعدة والمشفرة منها 14 كيلوقاعدة.

وتنتشر أنواع المكررات والعناصر القافزة في جينوم الإنسان ومنها القافزات الارتدادية مثل LINEs المرتبطة بالجينات واطئة التعبير وطويلة الانترونات . و SINES مثل هذه التواليات قد توجد في الجينات ذات التعبير العالي قصيرة الانترونات والأخرى المشابهة للفيروسات الارتدادية Retrovirus-like elements . وتشير الدراسات الى ان عوائل المكررات تطورت في وقت يعد قصيرا (عدد قليل من ملايين السنين)، ولذا تقسم المكررات الى جديدة او قديمة وأخرى قديمة اعتمادا على دخولها للجينومات، وبعد اقتحامها للجينومات فان أغلب النسخ تجمعت فيها الطفرات جعلتها غير قادرة على التكاثر وهي تقحم في اغلب الأحيان في المناطق غير المشفرة وتميل المكررات الحديثة ان تقحم داخل المكررات القديمة مؤدية الى تكوين تراكيب عنقودية من المكررات .

Alu sequences من القافزات الارتدادية وهذه التواليات تعد مهمة في جينوم الإنسان والتي يعتقد انها نشأت من جزيئات RNA صغيرة (7SL RNA) سايتوبلازمية توجد في المناطق الغنية بـ GC، وهذه التواليات يصل طولها 300 قاعدة لذلك فهي ضمن مجموعة SINES وتحوي على موقع للانفلاق بأحد إنزيمات القطع الداخلي (Alu) Endonuclease .
'5'-AG/CT-3' المنتج من البكتريا *Arthrobacter luteus* . وهذه التواليات المتحركة هي الأكثر وفرة في الجينوم البشري وتصل بعض الأحيان الى مليون نسخة ١ جينوم بشري ونظرا لتكرار افراد عائلة Alu sequences لأكثر من مليون فهي تشكل حوالي 10.7% ويكون 0.5% متغايرا. مجموعة Alu تعود الى القافزات الارتدادية لا تشفر لمنتجات بروتينية وتعتمد على القافزات الارتدادية LINE في تضاعفها.

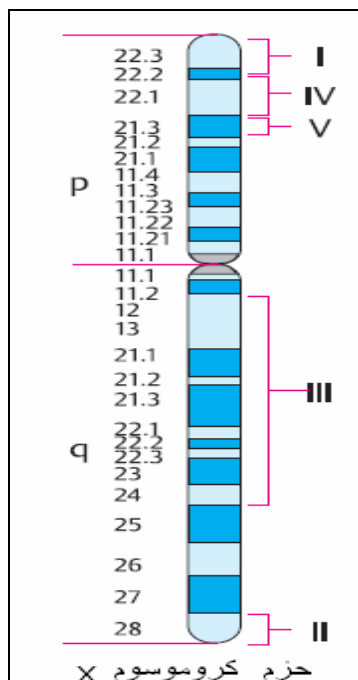
الاضطرابات التي تحدث في هذه التواليات تؤدي الى الأمراض المستورثة ومنها جعل الأشخاص مؤهلين للإصابة بالسرطان ، اذ ان إقحامها يؤدي الى الاضطراب ومنها احد أنواع سرطان الثدي Ewing sarcoma ونزف الدم الوراثي Haemophilia بنوعيه A و B ، داء السكري النوع الثاني ، ارتفاع كولسترول الدم العائلي، أمراض في الجهاز العصبي مثل Neurofibromatosis. اما التغاير الذي يحصل فيها فيمكن ان يكون مرتبطا ببعض الأمراض اذ تؤثر في مستويات الانتساخ ومنها مرض الزهايمر وسرطان الرئة وسرطان المعدة. ومن الجدير بالذكر ان هذه التواليات تكون مفيدة في تحديد العلاقة التطورية كما دراسة المتحجرات وذلك لان الأفراد التي تشترك في توالي Alu محدد هي التي تكون ذات أصل مشترك.

التميط النووي Karyotyping

يعني وصف للكروموسومات اي المجموعة الكاملة من كروموسومات النوع او الفرد الواحد مثل العدد الذي يحسب في الخلايا الجسمية والحجم والشكل وأي صفات أخرى مثل ظهور بعض التراكيب الداخلية مثل المراكز ومواقعها وهذه تسجل في الطور الاستوائي Metaphase

للانقسام الخيطي عند الفحص بالمجهر الضوئي ، تدرس هذه المواصفات ضمن Karyology وتنقسم كروموسومات الإنسان وفقا لموقع المركز Centromere الى عدة أقسام :

- وسطية المركز Metacentric
 - Submetacentric ويكون المركز منزاحا" عن وسط الكروموسوم
 - Telocentric ويكون المركز قرب نهاية الكروموسوم
 - Acrocentric ويكون المركز في نهاية الكروموسوم
- هذه المواقع تحدد أذرع الكروموسوم وهي ذراعان القصير p (من petit) وذراع طويل q (الحرف الذي يلي p في ترتيب الحروف).
- ويمكن ان تظهر حزم في الكروموسوم عند تصبغ الخلايا بصبغات خاصة بعد إيقاف عملية الانقسام باستعمال الكولجسين Colchicine مثل صبغة كزما Giemsa التي تتخصص بالتفاعل مع مجاميع الفوسفات في DNA واستعمال طرق تصبغ خاصة (طريقة التحزيم G-banding) اذ يستعمل فيها الترسين لهضم البروتينات قبل التصبغ والسماح للصبغة للارتباط بـ DNA. وهناك طرق أخرى لإظهار الحزم في DNA منها Q-banding باستعمال صبغة Quinacrine التي ترتبط بشدة الى المناطق الغنية بـ AT وفي هذه الحالة لا بد من استعمال مجاهر الفلورة الخاصة واستعمال الأشعة فوق البنفسجية وفيها تظهر الكروموسومات حزم مماثلة للطريقة أعلاه .
- وتكون الحزم الغامقة ممثلة للكروماتين المتباين الخامل من الناحية الوراثية في اغلب الأحيان وحزم فاتحة اللون تمثل الكروماتين الحقيقي الذي ينتسخ ويعبر عنه وتكون مناطق الكروماتين المتباين غنية بـ AT وتتم مضاعفته في أوقات متأخرة في مرحلة التخليق S phase من دورة الخلية حقيقية النواة، اما الكروماتين الحقيقي فهو غني بـ GC وتتم مضاعفته في أوقات مبكرة من مرحلة التخليق. والطريقة الأخرى لإظهار الحزم في الكروموسومات R-banding هي حالة العكس لعملية G-banding اذ تظهر المناطق الغامقة في حالة G-banding فاتحة اللون في هذه الحالة وكذا الحال مع المناطق الفاتحة.
- وهناك طرق تصبغ أخرى لدراسة الوراثة الخلوية Cytogenetics مثل T-banding الخاصة بإظهار نهايات الكروموسومات .
- اما التصبغ بأملح الفضة وهي طريقة خاصة جدا فتستعمل لصبغ النويات Nucleolar organization region والبروتينات المرتبطة بها حيث تصطبغ المنطقة بشدة مشيرة الى فعاليات الجينات الخاصة بـ rRNA.
- وهناك تقنيات متطورة تستعمل وسائل أخرى مثل الصبغات المختلفة ومجاهر تعمل بأسس مختلفة. وتظهر طرق التصبغ هذه الحزم الثانوية وكذلك الحزم الثانوية- الثانوية .
- ويبدأ ترقيم المناطق من المركز والمناطق تقسم الى حزم. والشكل التالي (44.3) يوضح طريقة الترقيم ومظهر الكروموسوم.



شكل 44.3 : طريقة ترقيم مناطق وحزم الكروموسومات (كروموسوم X)

وأفضل الخلايا المستعملة لتحديد النمط النووي لخلايا الإنسان هي لمفاويات الدم الدائر لسهولة تنميتها في المزارع الخلوية. ودراسة التتميط النووي تشير الى ان:

- حجوم او أطوال الكروموسومات غير متساوية
 - اختلاف في موقع المركز الذي يمكن ان يتغير مكانه بعملية الانتقال Translocation
 - الاختلاف في عدد ومواقع قطع التتابع Satellites
 - الاختلاف في توزيع الكروماتين المتباين والحقيقي.
 - الاختلافات في بنية الكروموسومات في النوع الواحد تعتمد على:
 - الجنس
 - الفروق بين الخلايا الجنسية والجسمية
 - اختلافات نتيجة التغيرات في الكروموسومات
 - الاختلافات الناتجة عن اختلاف المواقع الجغرافية
 - الاختلافات في العرق Race
- هذا فضلا عن الاختلاف في المجاميع الكروموسومية التي قد تكون حالات شاذة ومنها:
- Diploidy تمثل عدد مجموعة كاملة من الكروموسومات وهي الحالة الطبيعية .
 - Polyploidy حالة وجود أكثر من مجموعة من الكروموسومات وتكثر هذه في النباتات وفي الحالات المرضية للإنسان .

- Haplo-diploidy وفيها يكون أحد الجنسين من النوع الواحد (الذكر أو الأنثى) مزدوج المجموعة الكروموسومية والآخر يكون أحادي المجموعة كما يحصل في رتبة الحشرات Hymenoptera .
- Endopolyploidy تحصل في بعض الأحياء عندما تتخصص بعض أنسجتها وتتوقف عن الانقسام الخلوي ولكن دون توقف الانقسام النووي أي يحصل انقسام خيطي داخلي Endomitosis أو ما يسمى تكاثر داخلي للكروموسومات Endoreplication، وتفصل مجاميع الكروموسومات عن بعضها بغشاء نووي وتحصل في بعض الحشرات وبعض النباتات مثل الذرة، وتستغل هذه كإستراتيجية لزيادة الإنتاج حيث أن الأنسجة التي تحصل فيها زيادة في جرع الجينات تحصل فيها عمليات التخليق بمستويات عالية، والعملية تحصل في العديد من الكائنات وتساهم في التخصص والصفات المظهرية للكائنات.
- Aneuploidy وفيها تحصل إما زيادة أو نقصان لأحد الكروموسومات مؤدية إلى حالة غير طبيعية وتؤدي إلى حدوث الأمراض الوراثية مثل متلازمة داون Down's syndrome و Turner syndromes.

تغيرات النيوكلو تيد المفرد (SNPs) Single nucleotide polymorphisms

تغير يحصل في إحدى نيوكلو تيدات توالي معين إذ تحل قاعدة نيتروجينية محل أخرى كما في الآتي .



ويعد التغيير SNP فيما إذا حصل في 1% من المجتمع وتعد من أهم مصادر التغيرات بين أفراد النوع الواحد. وتوجد أنواع منها ما يقع في المناطق المشفرة من DNA وتسمى cSNP وقد تكون هذه غير مؤثرة لذلك تسمى بالتغيرات الصامتة sSNP نظرا لامتلاك بعض الحوامض الأمينية أكثر من شفرة وراثية (كما مر ذكره في موقع آخر). وبعض هذه التغيرات تحصل في المناطق المسؤولة عن التنظيم rSNP مثل الممهدات وأخرى تقع في مواقع خياطة الجينات مؤدية إلى التأثير في النسخ الناتجة . وتوجد أعداد كبيرة من المواقع في الجينوم البشري، وتحدث بمعدل واحد لكل 1000-2000 نيوكلو تيد في الجينوم البشري وهذا يعني أن هناك حوالي 1.42 مليون من المتغيرات المنتشرة في الجينوم، وهناك حوالي 60,000 تقع في الاكسونات.

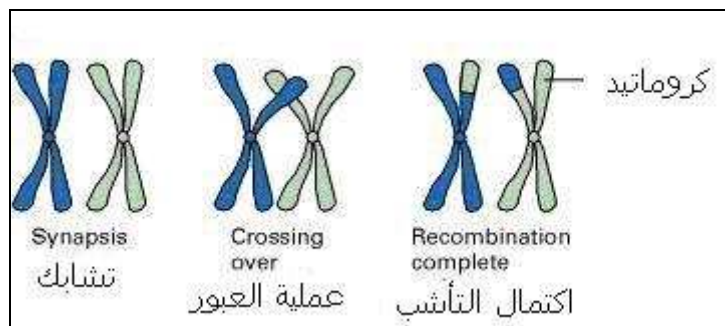
إن أغلب التغيرات تكون غير مؤثرة ومعتدلة ولكن هناك تغيرات يطلق عليها غير المترادفة (nsSNP) تؤثر في التعبير عن الجين أو التأثير في وظيفة البروتين الناتج عنه والذي قد يكون كارثيا بالنسبة للخلايا وتغير النمط المظهري ويوجد عددا كبيرا منها في الجينوم البشري وبمعدل 2400-60,000 للشخص الواحد أي حوالي 1-2 طفرة في البروتين، البعض منها يمكن أن يزيد من اللياقة الشخصية والمدمر منها يتم التخلص منه أثناء عمليات الانتخاب في حين تتم

المحافظة على الطفرات والتغيرات المفيدة وترحف الى جينومات المجتمعات لتصبح ثابتة وتساهم فيما بعد في تمييز النوع التي عندها يمكن تتبعه في الجينات وتوارث الصفات. وقد وجد ان التغيرات المدمرة nsSNPs تكثر في الحيوانات المدجنة أكثر من الحيوانات الطبيعية او البرية وهذا يعني ان حدوث هذه التغيرات يكون تحت تأثير الانتخاب المتراخي نظرا لتعرض الحيوانات المدجنة الى ظروف اقل قسوة من تلك الموجودة في الحياة البرية والتي سمحت لهذه التغيرات المدمرة بالثبوت فيها.

ونظرا لأهمية هذه التغيرات وتسببها في إحلال حامض أميني مكان آخر وجدت عدة برامج وطرق للتعرف عليها (كما سيأتي ذكره في الفصول القادمة)، وإيجاد قواعد بيانات خاصة بها. ومما سهل دراسة التغيرات انها لا تحتاج الى عمليات ترحيل كهربائي او خطوات معقدة للكشف عنها ويمكن ان يكشف عنها باستعمال طريقة سوزرن Southern blot او RFLP (Restriction fragment length polymorphism)، اذ في الطريقة الأخيرة يمكن ان يؤدي وجود التغيرات الى اختلاف مواقع التميز بالإنزيمات القاطعة بين النسخ او الاليلات المختلفة، كما يمكن استعمال طرق التهجين للكشف عنها باستعمال تواليات قصيرة وعند الاختلاف في نيوكلوديد واحد يؤدي الى فشل عملية التهجين.

التأشب Recombination

عملية طبيعية تكسر فيها أشرطة DNA ويعاد لحمها لإنتاج تشكيلة جديدة من الجينات، حيث تكون تشكيلة الجينات في الآباء لا تشبه الأبناء الناتجة إي أنها تؤدي الى التغير الوراثي وأفضل الأمثلة عليها عمليات العبور الموضحة في الشكل 45.3 :



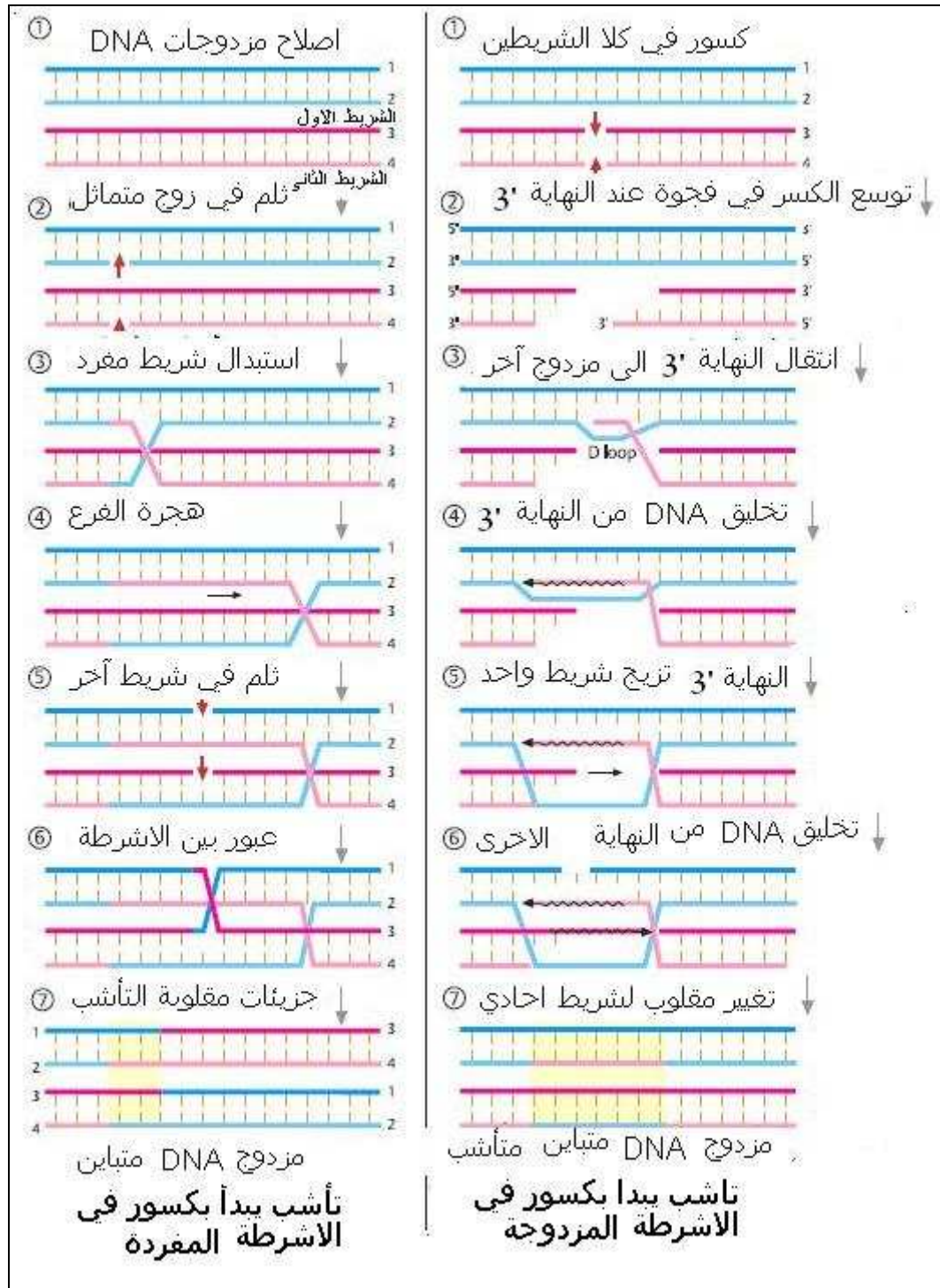
شكل 45.3 : عمليات العبور والتأشب في الكروموسومات

والإنزيمات العاملة فيها هي إنزيمات إعادة الالتحام او التأشب Recombinases وعند انتقال الأشرطة الى المواضع الجديدة يؤدي الى استبدال نسخ الجينات (أليلات) بأخرى . ويوفر التأشب مرونة للجينوم وبدونه فان الجينات تبقى ثابتة على كروموسوم واحد والتغير الوحيد الذي سوف تعانيه هو الطفرات التي تكون خطرة في معظم الأحيان ، ولذلك فان التأشب يوفر وسائل لإعادة التركيب والتخلص من الطفرات غير الملائمة والمحافظة ونشر الطفرات الملائمة وبذا يمكن للفرد الحصول على مجموعة من المعلومات الوراثية. ويزيد التأشب من إمكانية تطور الجينات . ويحدث التأشب في العادة بين تواليات متماثلة لضمان عدم إضافة او حذف للقواعد النروجينية. ويوجد أنواع من التأشب:

- العام أو المتجانس والذي يحدث في الأحياء حقيقية النواة أثناء عمليات الانقسام الاختزالي
 - التأشب المتخصص الموقع
 - نوع آخر يحدث أثناء عمليات قفز العناصر القافزة أو المتحركة لإقحام تواليات DNA في جزيئات أخرى دون مراعاة التشابه بين التواليات.
- وفي العموم يمكن ان يحصل التأشب نتيجة لكسر أحد الأشرطة أو كسر الشريطين. ففي الحالة الأولى تحدث كسور في المواقع المتقابلة من أشرطة متماثلة أي تكون التواليات من أصول مختلفة بواسطة أحد إنزيمات القطع الداخلية، وهذا يسمح لأحد النهايات الحرة للشريط المكسور في أحد الجزيئات بالارتباط مع النهاية الحرة للشريط المكسور في الجزيئة الثانية. اما في حالة كسر الشريطين فان الفجوة بين الكسر ستتضخم بواسطة إنزيمات القطع الخارجي Exonucleases، وبعملات متوالية وعمليات إصلاح وإعادة ترتيب تؤدي الى إنتاج مزدوج حلزوني متباين في كل المنطقة التي عانت من التأشب والتي تظهر نتائجها بعد إعادة تضاعف الأشرطة . والعمليات موضحة في الشكل (46.3) .

تضاعف المواد الوراثية

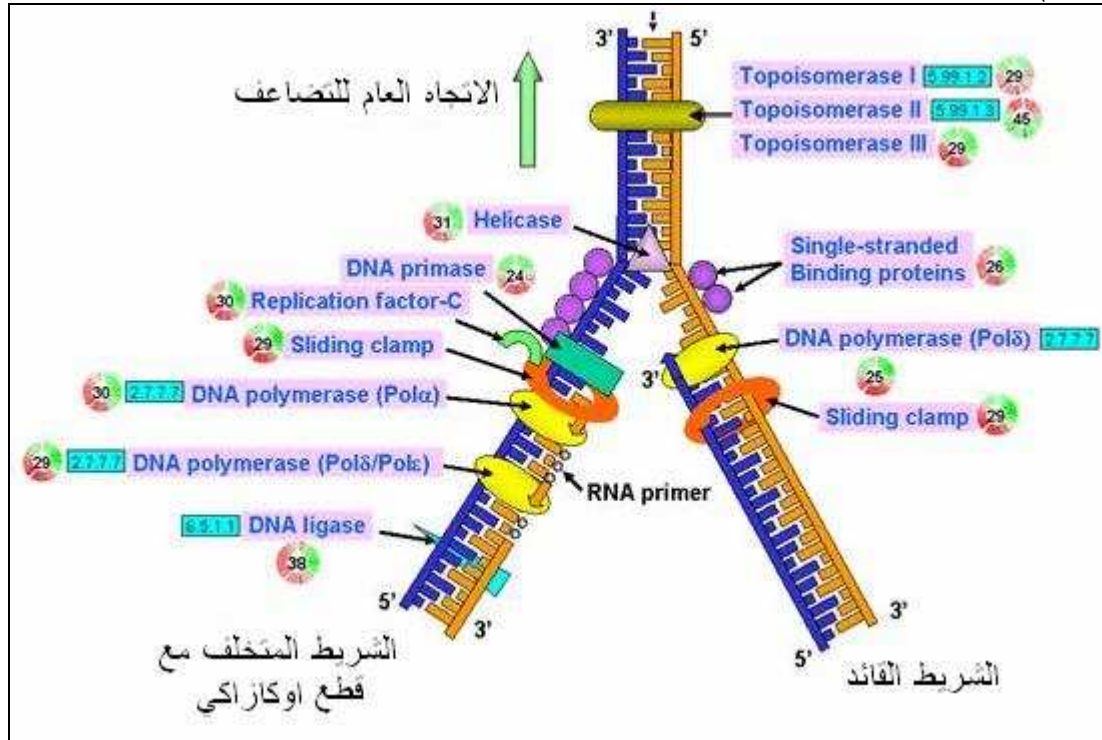
تتضاعف المواد الوراثية أو الكروموسومات وغيرها من العناصر الوراثية قبل انقسام الخلايا وهذا يكون ضمن الأنظمة الحية ، ولكن الطرق الحديثة مكنت من تصنيع بعض الجينات الصغيرة مثل تلك المسؤولة عن تخليق الببتيدات العصبية والببتيدات المضادة للميكروبات والبروتينات المكيفة Adaptors وبعض البروتينات المنظمة. ومن العودة للأنظمة الحية ففي الخلايا حقيقية النواة يحصل هذا في طور التخليق S phase. وتبدأ عملية التضاعف بفك ارتباطات Unzipping للأشرطة الأبوية وكسر الأواصر الهيدروجينية بينها وتعمل قواعد كل شريط كقوالب لإقحام القواعد في الأشرطة الجديدة تحت التخليق، نظرا لكون السلاسل أو الأشرطة مكملة إحداها للآخرى لذا يكون تضاعف DNA شبه محافظ أي تتكون سلسلة جديدة مقابل سلسلة شريط قديم . وعند ارتباط النيوكليوتيدات ثلاثية الفوسفات في أثناء تخليق الشريط الجديد الى ذرة الكربون 3' يفقد النيوكليوتيد مجموعتين من الفوسفات اللاعضوية .



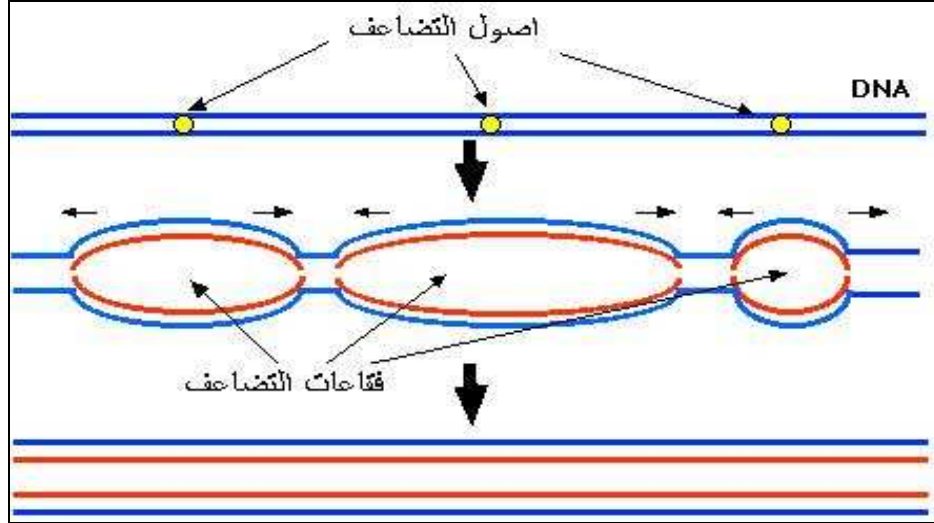
شكل 46.3: أنواع التآشب Recbinations

وتشمل عمليات التضاعف عدداً من الإنزيمات والبروتينات ومن الإنزيمات العاملة هي إنزيمات فتح الأشرطة التي تقوم بفك الأشرطة المزدوجة لتكون شوكة التضاعف وترتبط هذه الإنزيمات مع غيرها من الإنزيمات والبروتينات وأهمها إنزيم الكوثر DNAP ضمن تركيب معقد يدعى جسيم التضاعف Replisome وترتبط جزيئة من إنزيم كوثر DNAP إلى أحد الأشرطة وتبدأ بالتحرك من الاتجاه 3' إلى 5' (بالنسبة للشريط القالب) لتصنع الشريط الجديد بالاتجاه من 5' إلى 3' ، ونظراً لاختلاف قطبية الشريطين يكون أحد الأشرطة شريطاً قائداً ويضاعف بدون عرقلة في الخلايا حقيقية النواة بالإنزيم دلتا (DNAP δ) أما الشريط الآخر الذي يكون معاكساً لاتجاه فعالية DNAP وهو الشريط المتخلف فتتم مضاعفته في خلايا حقيقية النواة بالإنزيم إبسلون (DNAP ϵ) ويكون تخليقه بشكل قطع مفصولة هي قطع أوكازاكي ثم تزال من القطع بوادئ RNA المستعملة وتملاً الفجوات الداخلية وتلحم بإنزيمات مثل DNA ligase I ماعداً مكان البادئ الأخير الذي يكون في نهاية الكروموسوم الذي يقصر بعد كل عملية انقسام (كما مر ذكره في موقع آخر) كما موضح في الشكل 47.3 .

والتضاعف في البكتريا يبدأ من نقطة واحدة هي نقطة الأصل ويسير بالاتجاهين بمعدل 1000 نيوكليوتيد \ ثانية ويستغرق في *E. coli* حوالي 40 دقيقة ويكون مرافقاً لعمليات التصحيح. أما في كروموسومات الخلايا حقيقية النواة الكبيرة جداً فهذه تستغرق وقت طويلاً جداً إذا بدأت من منطقة واحدة لذلك توجد العديد من نقاط بدء التضاعف التي تسير بالاتجاهين مكونة فقاعات التضاعف تلنقي في النهاية وتحدث في طور التخليق من دورة الخلية كما موضح في الشكل (48.3) .



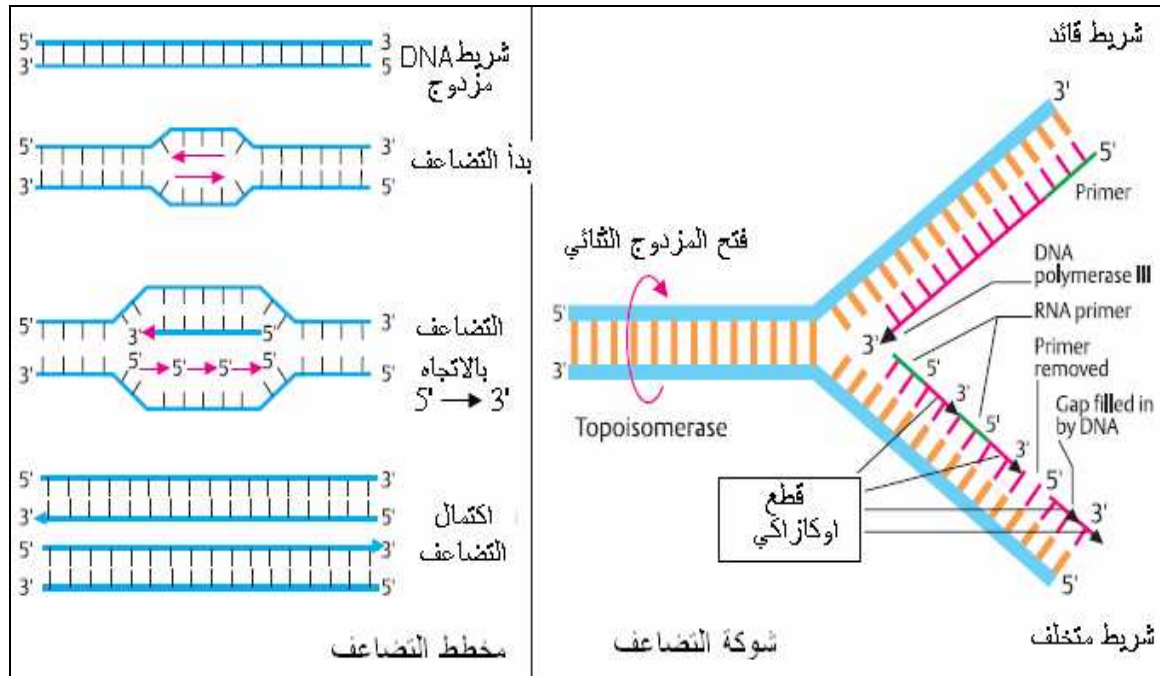
شكل 47.3: تضاعف أشرطة DNA



شكل 48.3 : تضاعف الكروموسومات في حقيقية النواة

مخطط عملية التضاعف

ان أشرطة DNA الجديدة تبدأ بالتخليق بالاتجاه 5' الى 3' وعند نقطة التضاعف يعمل ذراعي شوكة التضاعف كقوالب لتخليق شريط جديد وتكوين شوكة التضاعف يحتاج الى إنزيمات من المجموعة التي تغير طوبوغرافية الأشرطة وهي Topoisomerases وبما ان الأشرطة الأصلية متعاكسة القطبية، فهناك شريط قائد يبنى بالاتجاه 5' الى 3' والشريط المتخلف الذي اتجاؤه من 3' الى 5' لذلك فان شريط DNA الجديد الذي يخلق على قالب الشريط المتخلف يخلق بشكل قطع صغيرة تصل من 1000-2000 نيوكلوثيريد وهي Okazaki fragments تبنى على بوادي من RNA لبدء التضاعف وتنتج من قبل أحد إنزيمات كثررة RNA المسمى Primease (كما مر ذكره) وبعد ذلك تزال قطع البوادي ويتم إقحام قطع من DNA مكانها بواسطة DNAPI ثم يتم لحمها مع باقي الشريط بإنزيم اللحم Ligase. والإنزيم المسئول عن كثررة DNA هو DNAPIII معقد التركيب يتكون من عدة وحدات ثانوية، كما انه في الخلايا حقيقية النواة يوجد تخصص لهذا الإنزيم منه ما يعمل للشريط القائد والآخر يعمل للشريط المتخلف كما موضحة في الشكل 49.3 :



شكل 49.3 : مخطط لعملية تضاعف DNA

وأثناء عملية التضاعف تحصل بعض الأخطاء يمكن التخلص منها بواسطة الفعالية المعقدة المعروفة بآلية تصحيح القراءة Proofreading mechanism التي تزيل أي قاعدة أدمجت بالخطأ وتشد عن حقيقة تكامل القواعد وتصنع مكانها القاعدة الصحيحة.

السيطرة على عمليات التضاعف في حقيقيات النواة

الأطر العامة للسيطرة تكون إما سيطرة موجبة أو سالبة :

1- الموجبة وهي ما يطلق عليها بالسيطرة الموجبة المرخصة Licensing positive control وتتم بمساعدة معقدات بروتينية منها:

- معقدات تميز الأصل Origin recognition complex proteins (ORC) وهذه تبقى مرتبطة بالـ DNA طول عملية التضاعف.

- بروتينات مساعدة وتسمى عوامل الرخصة Licensing factors تتجمع في مرحلة G_1 من مراحل دورة الخلية وأغلبها من المدورات Cyclins.

2- الطريقة الثانية للسيطرة هي السالبة Gemini negative control وتتم بمساعدة بروتينات Gemini التي تتجمع في الانوية أثناء المرحلة G_2 من مراحل دورة الخلية وتقوم بمنع الكروموسومات المتضاعفة حديثاً من التضاعف مرة ثانية لحين اكتمال الانقسام الخلوي ثم تتحلل هذه البروتينات لتسمح لعوامل الرخصة بالعمل وبدء عملية التضاعف في طور التخليق من الدورة القادمة للخلية. وبعض الخلايا تسمح لطور التخليق وتضاعف الكروموسومات بالبدء دون إكمال عملية انقسام الخلايا وهو ما

يسمى بالتضاعف الداخلي اي يلغى في هذه الحالة دور السيطرة السالبة او دور بروتينات او عوامل Gemini .

التعبير الجيني

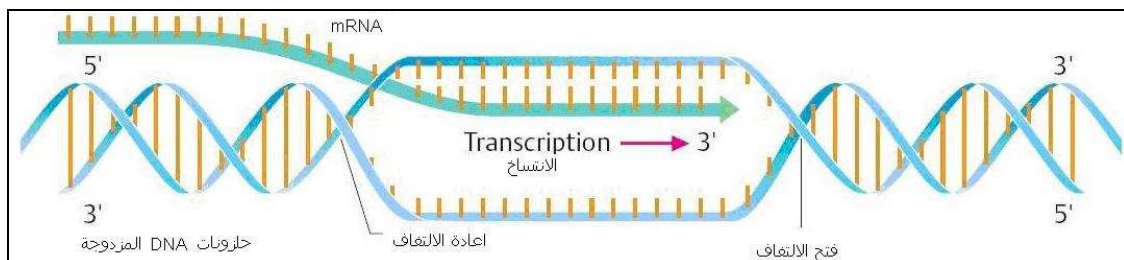
المعلومات الوراثية الموجودة في مكنون جينومات الأحياء يجب ان تحول الى شكل يستفاد منه اي يجب ان تسري بشكل يستفاد منه في النواحي الحيوية ويتم ذلك بتخليق البروتينات التي تمثل الطريق المباشر للاستفادة من سريان المعلومات الوراثية، وقد تكون الاستفادة من سريان المعلومات بشكل غير مباشر كما في عمليات تنظيم الجينات (كما سيأتي ذكره في موقع آخر). والواضح ان ليس كل المعلومات الوراثية الموجودة في الجينوم يعبر عنها في الوقت نفسه. ان سريان المعلومات يكون باتجاه واحد من DNA الى RNA ثم الى البروتينات. بواسطة عمليات الانتساخ Transcription التي يتم فيها تخليق جزيئات من RNA باعتماد أشرطة DNA كقوالب للبناء وفق حاجة الخلايا ثم بعد ذلك ترجمة المعلومات Translation التي حملت بجزيئات RNA الى بروتينات بمساعدة الرايبوزومات الواقعة في الساييتوبلازم ويتناول هذا الجزء بعض جوانب عمليات الانتساخ والترجمة.

الانتساخ يعني تخليق جزيئات RNA حاملة للمعلومات من مناطق محددة من جينوم الخلية بعد انصهار الأواصر الهيدروجينية وفتح الأشرطة ليبدأ إنزيم كثررة RNA (RNA polymerase) بإضافة القواعد النتروجينية المكملة للشريط الذي ينسخه وهذا يعني ان الأشرطة الناتجة تكون نسخة من شريط DNA ما عدا استعمال اليوراسيل (U) مكان الثايمين (T) وبذلك تكون الأشرطة المزدوجة DNA/RNA duplex اقل استقرارا من أشرطة DNA المزدوجة (dsDNA) وسرعان ما تنفك منها بعد أداء مهمتها لتعود أشرطة DNA للازدواج ثانية، وبما ان أشرطة RNA هي تقريبا صورة للـ DNA الذي نسخته لذلك سميت العملية انتساخ او استنساخ وتسمى هذه المرحلة بالتعبير الجيني . في حين أطلق على العملية التي تليها وهي الترجمة بالتعبير البروتيني Protein expression لان هوية المواد الناتجة لا تشبه هوية المواد الكيماوية في الجزيئات المراسلة. وعلى العموم فالنواتج النهائية اي البروتينات او الببتيدات متعددة تحدد وظائفها ومكان وجودها وأي صفات أخرى بتواليات قطع DNA التي شفرت لها.

لذلك كان هناك فرقا بين التعبير الجيني والتعبير البروتيني والتي يمكن الكشف عنها وتمثيلها بأطالس التعبير Expression atlas ، فلإنتاج بروتين ما فلا بد من إنتاج mRNA ولكن ليس كل mRNA سوف ينتج بروتين فضلا عن ان العمر النصفى للـ mRNA قصير مقارنة بالبروتينات التي تعطيها وعليه فانه يمكن الكشف عن البروتينات ولكن جزيئات mRNA غير موجودة، كما ان كميات جزيئات البروتينات و mRNA هي كميات غير متكافئة.

عملية الانتساخ

يتم فتح الأشرطة للمنطقة المراد انتساخها او الجين لإنتاج او بناء جزيئات RNA الذي يتحول بعد عدة عمليات في الخلايا حقيقية النواة الى mRNA ويستعمل كقالب لصف الحوامض الامينية كما موضح في الشكل 50.3 :



شكل 50.3 : ملخص لعمليات الانتساخ والترجمة

وعملية الانتساخ هي التي تكون مكنون النسخ Transcriptome بأجمعه ويكون جزءا منها جزيئات RNA القابلة للترجمة (mRNA) اما الباقي فتكون جزيئات غير قابلة للترجمة مثل rRNA و tRNA وجزيئات tmRNA والأخيرة تساعد مع جزيئات RNA صغيرة (sRNA) أخرى في عملية تشجيع الترجمة المتعثرة لسبب ما. ومن هنا يلاحظ عدم صلاحية فرضية القضية المحورية Central dogma في علم الحياة الجزيئي والتي تشير الى

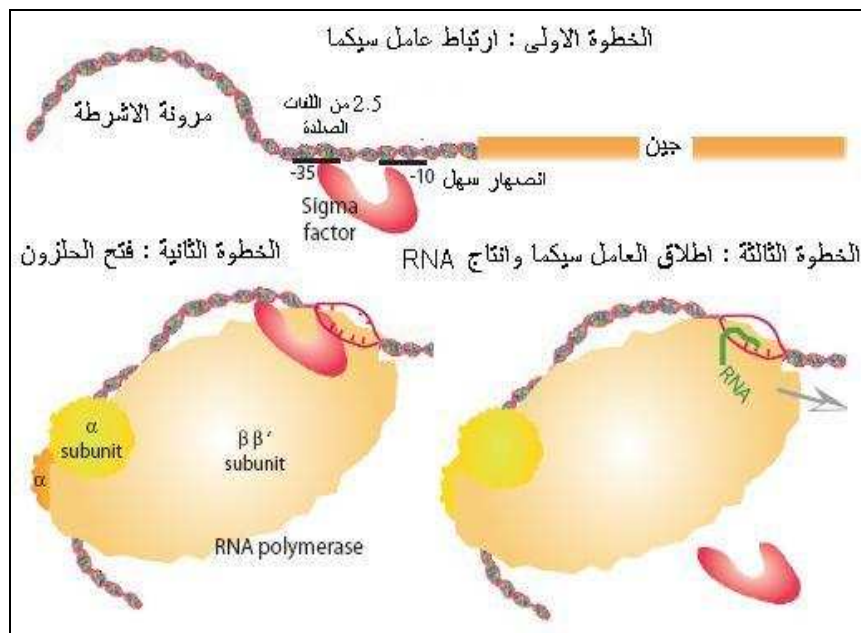
DNA ← RNA ← بروتين

وهذه قد تكون صالحة في معظم الجينومات البكتيرية حيث ان أغلبية DNA المنتسخ يترجم الى بروتينات.

ان جزيئات mRNA وحتى الناضجة منها تكون أطول قليلا من الجينات التي يعبر عنها الى حوامض أمينية وذلك لانها تحوي على جهة اليسار Upstream على قطع بدء الانتساخ التي تحدد بالممهدات Promoters، كما تحوي على مناطق في جهة اليمين Downstream بعد شفرة الوقف. اما في حالة الخلايا حقيقية النواة تكون نتيجة الانتساخ hnRNA (heterogeneous nuclear RNA) الحاوي على الانترونات التي يجب ان تزال والتي لا يمكن لمكائن الترجمة قراءتها وتحدث عملية الإزالة في النواة ثم تصدر mRNA الناضجة الى السايوبلازم (كما سيأتي ذكره في موقع آخر).

الممهدات Promoters

مناطق من تواليات تقع أمام التوالي المراد انتساخه وهذه المناطق لا تنتسخ وإنما يتم التعرف عليها. في الخلايا بدائية النواة تتم عملية التعرف باستعمال العامل سيكما (σ) الذي يرتبط الى إنزيم النسخ RNAP المكون من 4 وحدات ثانوية ويوجهه الى المكان المراد انتساخه الذي في العادة يشمل عدة جينات (أوبرون) الموضحة في الشكل 51.3 :



شكل 51.3 : بدء عملية الانتساخ في البكتريا

ويبدأ إنزيم كثررة RNA (RNAP) بمساعدة العديد من البروتينات بنسخ الشريط الحساس أو المطلوب Sense strand أو يسمى الشريط المشفر ويبدأ بالكثرة من 3' بالاتجاه 5' أي ان RNA يخلق بالاتجاه 5' الى 3' ، اما الشريط المقابل فهو الشريط النقيض أو غير الحساس . وتعد تفاصيل تركيب الممهدات (منها المذكورة في الشكل أعلاه) من الأساسيات في وضع برامج الحاسوب للتعرف عليها. وفي العموم تمتاز الممهدات القوية بوجود قطعة من توالي DNA منحنية يتبعها وجود منطقة صلبة هي الواقعة بين (-35) و (-10) ، ثم منطقة سهلة الانصهار وهي (-10) ومن وجود هذه الصفات يمكن تحديد قوة الممهدات من دراسة تواليها.

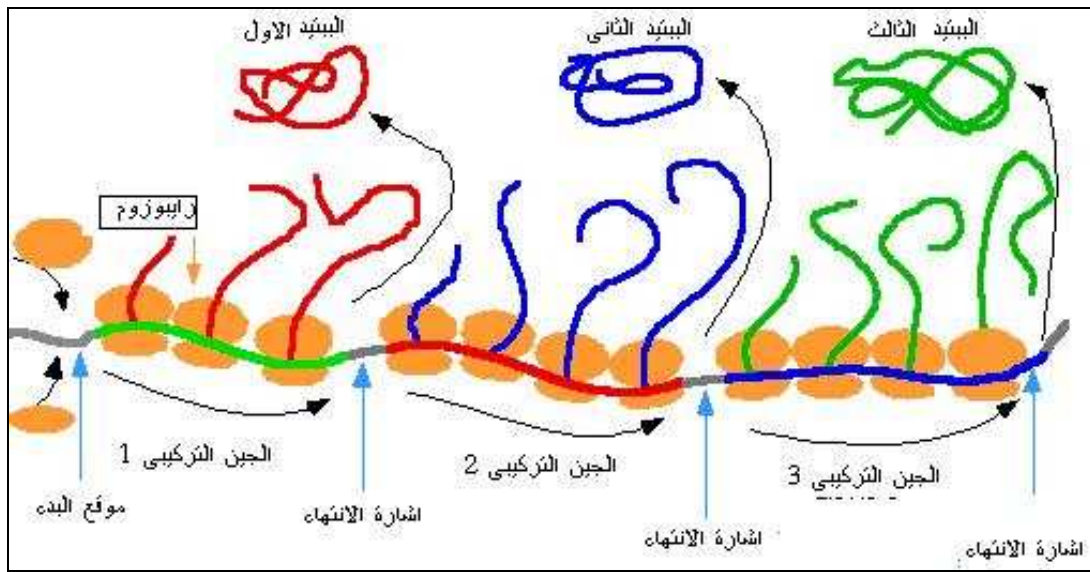
تحويلات نسخ RNA

تجري على جزيئات RNA الناتجة من نسخ الجينات المراد ترجمتها العديد من التحويلات ومنها ما هو خاص بالخلايا بدائية النواة وهذه تكون بسيطة ويمكن للنسخ ان تذهب لعملية الترجمة مباشرة ، ولكن معالجة الجزيئات في الخلايا حقيقية النواة تكون معقدة ومنها بخطوط عريضة :

- إضافة غطاء Capping الى النهاية 5' والتي تقوم مقام موقع الارتباط rbs في نسخ بدائية النواة.
- إضافة ذيل الأدنين الى النهاية 3' Poly(A)tail
- إزالة الانترونات وما يحصل بها من تفاضل (موضحة في عملية الخياطة البديلة) ومعالجة جزيئات RNA .
- إجراء التصحيحات RNA editing .

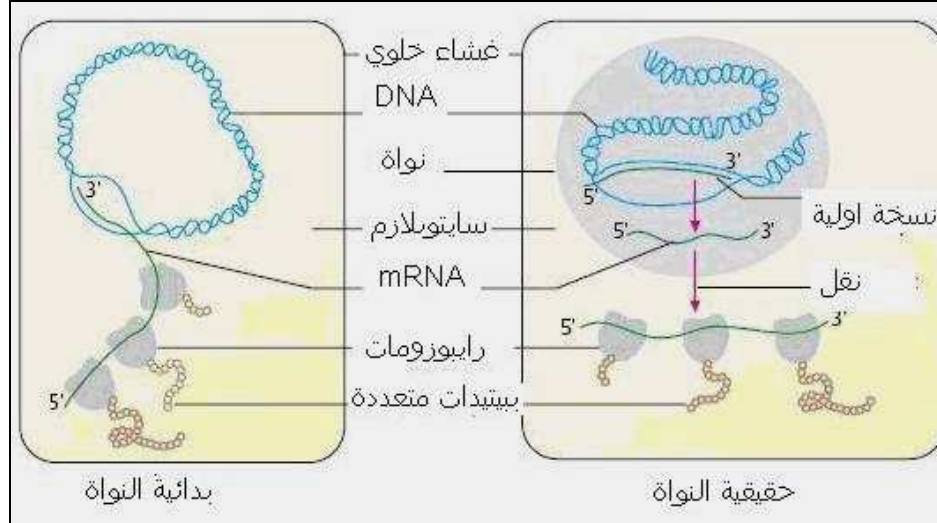
الترجمة Translation

بالمنطق البسيط هو وضع الحوامض الامينية في الببتيد الناشئ بمساعدة الرايبوزوم مقابل كل شفرة وراثية أي ان العلاقة تكاد تكون 1:1 أي ان كل من DNA و RNA تكون متوافقة Collinear لذلك يمكن ان تزوج عمليات الترجمة مع عمليات الانتساخ في الخلايا بدائية النواة وتفصل الرسائل المتعددة في mRNA في الخلايا بدائية النواة عن بعضها بشفرات الوقف كما موضحة في الشكل 52.3 :



شكل 52.3 : تزامن الانتساخ مع الترجمة في البكتريا

وفي الخلايا حقيقية النواة تكون الرسائل في العادة أحادية وكل رسالة تعطي بروتين واحد ولا يكون هناك ارتباط بين عملية الانتساخ والترجمة لان كل منهم يتم في جزء منفصل عن الآخر، ويمكن إجمال العديد من الفروق في الشكل 53.3 :

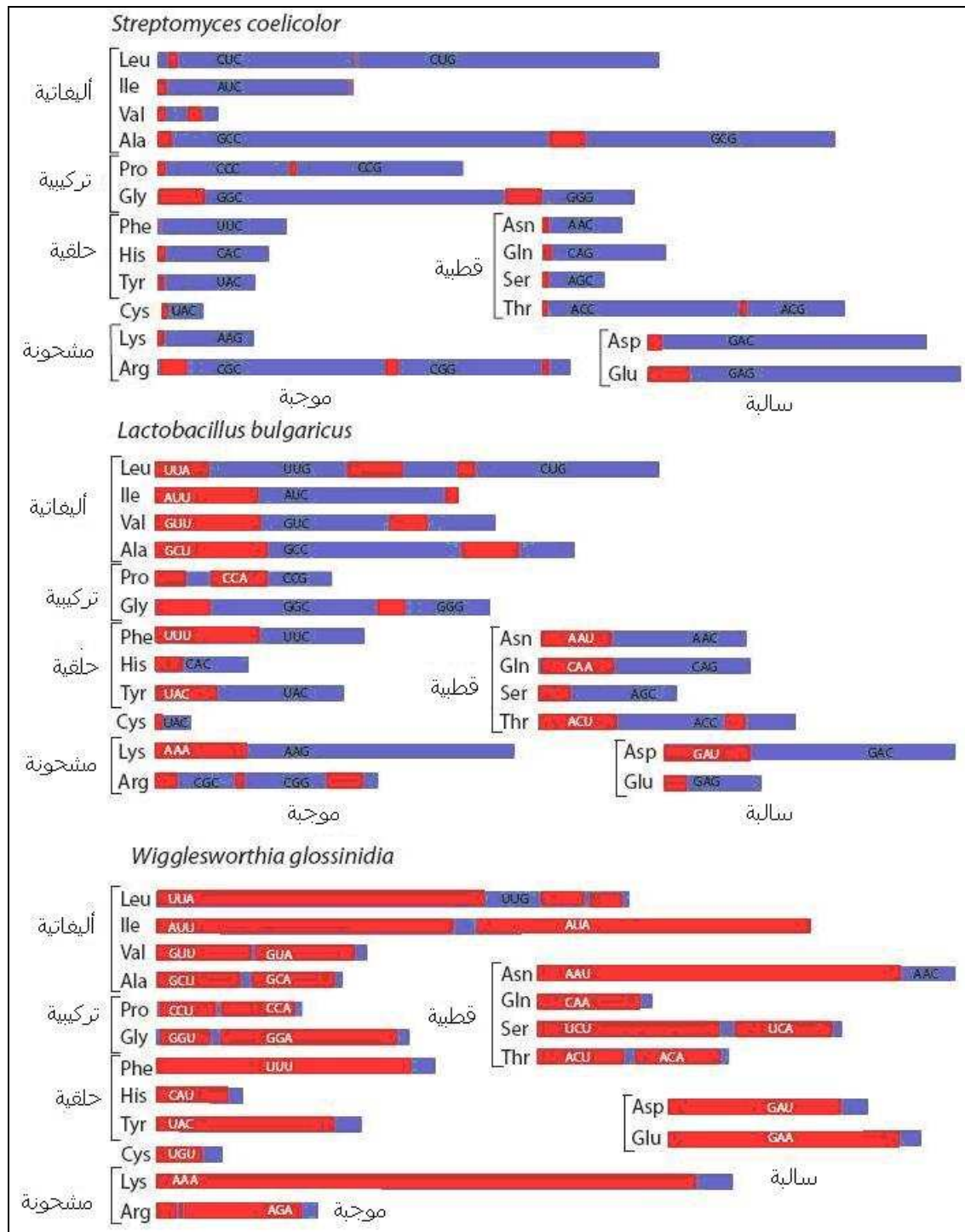


شكل 53.3 : الفروق بين الانتساخ والترجمة بين الخلايا حقيقية النواة وبدائية النواة

وبما ان mRNA للجين لا يكون مشفرا بشكل تام (كما ذكر آنفا) نتيجة لوجود بعض التواليات غير المشفرة للبروتينات التي تتركز عند النهايات 5' و 3' وكذلك المناطق بين الجينات في حالة الرسائل المتعددة ، وتكون هذه المناطق غنية بـ AT لذلك يقوم الرايبوزوم بعملية مسح باحثا عن rbs (في الخلايا بدائية النواة) التي تقع أمام شفرة البدء وهي شفرة الميثايونين في اغلب الأحيان . وفي حالة تعثر عملية الترجمة نتيجة لحصول الطفرات او اي سبب آخر تقوم جزئيات tmRNA بالمساعدة في الترجمة (كما ذكر في موقع آخر).

ان إضافة الحوامض الامينية مقابل كل شفرة خاصة بها هي ليست عملية مباشرة واضحة نظرا لوجود بعض الحوامض الامينية التي يشفر لها بعدة شفرات ، فهناك بعض الأحياء تميل الى استعمال شفرات محددة للحوامض الامينية متعددة الشفرات اي يكون هناك انحياز في استعمال الشفرات المتعددة والتي تختلف في الغالب في القاعدة النروجينية الثالثة وهذا الانحياز يكون له علاقة بمحتوى الجينوم من AT. والشكل التالي (54.3) يوضح تفضيل استعمال الشفرات في بعض أجناس بدائية النواة.

والبكتريا الظاهرة في الشكل *Streptomyces coelicolor* المنتجة لعدد من المضادات الحيوية يكون محتوى جينومها من AT بحدود 27.9% ، اما بكتريا اللبن *Lactobacillus delbrueckii* subsp *bulgaricus* فمحتواها من AT يصل الى 50.3% وفي البكتريا *Wigglesworthia glossinidia* من *Proteobacteria* وتعيش متطفلة في حشرات مرض النوم فمحتواها عاليا من AT ويصل الى 77.5% من الجينوم.



شكل 54.3: تفضيل استعمال الشفرات في بعض البكتريا

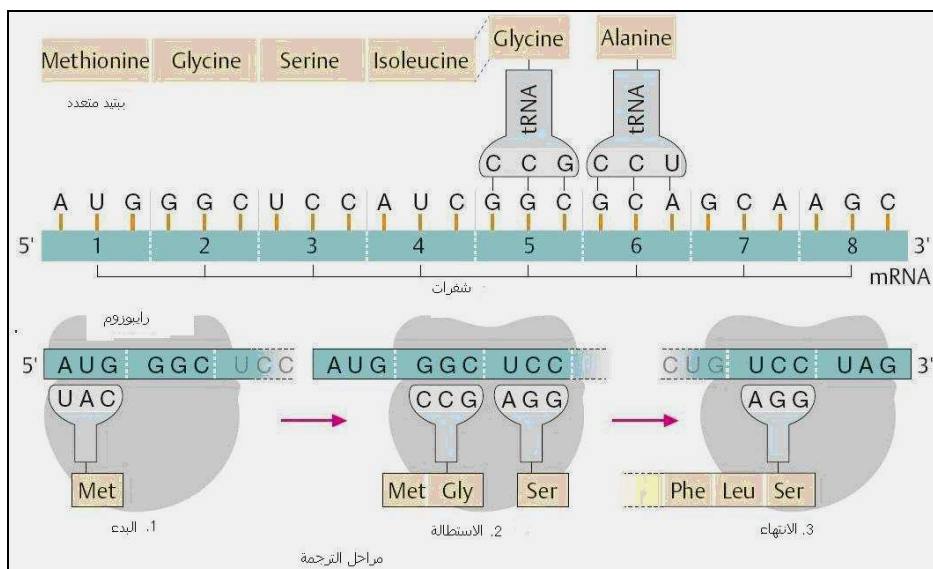
ومن الملاحظ من الشكل أعلاه ان أكثر الحوامض الامينية استعمالا هي ليست نفسها في الأحياء ، فمثلا *Strep. coelicolor* يكثر فيها الالنين ، والليوسين في *Lb. bulgaricus* والايزوليوسين في *W. glossinidia* ، وهذا ليس مستغرب لأنه لابد ان يكون هناك توازن بين الحوامض الامينية وبالتالي شفراتها مع محتوى الجينوم من AT ، فمثلا البكتريا الأخيرة لو استعملت الالنين بكثرة لابد ان يكون محتواها من AT اقل (اذ ان شفرات الالنين الأربعة يكثر فيها G و C). وكذلك الحال مع *Strep. coelicolor* التي يكثر الالنين ويكون متوافقا مع محتوى الخلايا مع القواعد النتروجينية ولا تلجأ مثلا الى استعمال الايزوليوسين الذي يحتاج الى نسب عالية من A حيث تكون شفراته غنية بهذه القاعدة النتروجينية، والحوامض الامينية المذكورة في المثال هي من مجموعة الحوامض الامينية الاليفاتية وبذلك كانت الأحياء تفضل بعضها على بعض ولكن ضمن المجموعة، ومثل هذا التفاضل يسري بالنسبة للحوامض الامينية المشحونة كما يلاحظ في الشكل أعلاه ، وكذلك الحال مع الحوامض الامينية سالبة الشحنة.

ان انحياز استعمال الشفرات يعني الانحياز في تكرار ووجود الشفرات المترادفة في DNA المشفرة وتفضيل الشفرات يعكس التوازن بين انحياز الشفرات والانتخاب الطبيعي للوصول بعملية الترجمة الى الحالة المثلى . فالشفرات المثالية في الأحياء المجهرية السريعة النمو مثل *E. coli* او خميرة الخبز تعكس وجود مجموعة من tRNAs الذي يساعد على زيادة سرعة معدلات الترجمة وبدقة عالية. اما في الأحياء بطيئة النمو مثل الإنسان و *H. pylori* و *C. elegans* والدروسوفلا ورشاد الصخر فتظهر الأحياء ميل قليل من تغاير الشفرات ولا يعتقد انه يعتمد على وفرة tRNA. وعلى العموم فهناك العديد من العوامل التي تؤثر في انحياز استخدام الشفرات منها ما هو متعلق بالكائن الحي اي عوامل داخلية ومنها ما هو متعلق بالظروف الخارجية وتطرفها مثل الحرارة والملوحة وغيرها.

ولكن التكرار الطبيعي للشفرات (الحوامض الامينية) الموجودة في mRNA يمكن ان يحدد بنوع جزيئات tRNA اللازمة لعمليات الترجمة ووفرتها تحدد معدل عملية الترجمة . وتحصل عملية الترجمة للإطار القابل للقراءة Reading frame من قبل ماكنة الترجمة والتي تبدأ بشفرة البدء وبمساعدة جزيئات tRNA الحاملة للشفرات المقابلة للحامض الاميني المشفر له بـ mRNA.

والترجمة في الخلايا بدائية النواة تحصل في الساييتوبلازم مباشرة لعدم وجود حدود بين المواد الوراثية وما يحيطها. اما في حقيقيات النواة فتحصل في الساييتوبلازم بعد تصدير الجزيئات المشفرة من النواة. وتقوم الرايبوزومات بهذه المهمة وهي مكونة من عدد من البروتينات و rRNA . والمراحل الأساسية لعملية الترجمة هي:

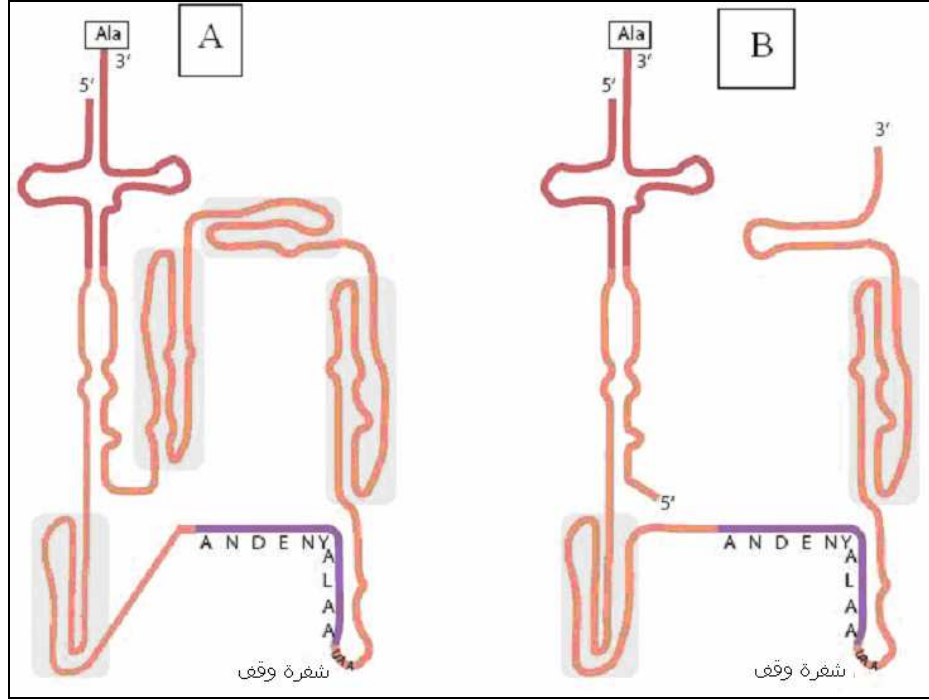
- **البدء** وفيها يتكون معقد البدء من mRNA والرايبوزوم و tRNA ويتم تجميع المعقد بمساعدة عدد من عوامل البدء Initiation factors مثل IF₁ ، IF₂ ، IF₃ وغيرها.
- **مرحلة الاستطالة** وفيها يضاف الحامض الاميني المحدد بإملاء من الشفرات الموجودة في mRNA ويتم انتقال الرايبوزوم على mRNA بالاتجاه 3' للـ mRNA وتستمر العملية الى حين الوصول الى شفرة الوقف وهي UAA ، UGA ، UAG ويترك اليبنتيد الرايبوزوم كما موضح بشكل مبسط في الشكل 55.3 .



شكل 55.3: انتقال الرايبوزوم الى حين الوصول الى شفرات الانتهاء

ومن هذا يتضح ان mRNA يعد حلقة الوصل بين النيوكليوتيدات والحوامض الامينية وان tRNA هي جزيئات فك الشفرات، وتنتهي العملية كما أشير أعلاه عند الوصول الى شفرات وقف وعملية إطلاق الببتيدات من الرايبوزومات تتم بمساعدة عدد من عوامل الإطلاق التي تميز شفرة الوقف على mRNA وترتبط بها وتجبر الرايبوزوم للابتعاد عن mRNA . وفي حالة حدوث خطأ من mRNA يؤدي الى عدم قابلية عوامل الإطلاق الارتباط مثل حصول كسر في الجزيئة او خطأ في عملية الانتساخ يؤدي الى إنهاء خاطئ للترجمة فان الرايبوزوم المتعثر Stalled ribosome لا يمكنه إعادة الدورة لإجراء عملية ترجمة جديدة . ومثل هذا الوضع تستطيع الخلايا تجنبه في البكتريا مثلا هناك جزيئات خاصة هي tmRNA تقوم بهذه المهمة.

وجزيئات tmRNA جزيئات غير مشفرة بطول حوالي 200 قاعدة لها ألفة عالية للرايبوزومات المتعثرة والمتوقفة لتقوم بتصحيح الخطأ بعملية تسمى Trans-translation والجين المسئول عنها يسمى SSrA يكون ثابتا في البكتريا وكان مثل هذا الجين يهمل في حالة تهيمش الجينومات في الماضي ولكن الدراسات الدقيقة ووضع البرامج الملائمة أدت الى توضيحه وتحديد تواليه، وأشارت الدراسات الى ان الجين المسئول عن tmRNA يوجد في البكتريا الحقيقية وبعض العائيات والمائتوكوندريا والبلاستيدات الخضراء ولكنه لم يسجل لحد الآن في جينومات الاركيا. ويوضح الشكل التالي (56.3) تركيب جزيئة tmRNA الثنائي.



شكل 56.3 : جزيئات tmRNA

ويلاحظ من الشكل ان التركيب يحوي ما يشبه mRNA و tRNA ومنها أُشتق اسم الجزيئات وتكون لها صفات mRNA و tRNA وبعض الأحيان تمتلك بعض البكتيريا جزيئات شبه منفصلة كما يلاحظ من الشكل أعلاه (B) وهو ما موجود في بكتيريا مثل *Prochlorococcus marinus* و *Caulobacter crescentus*.

وجزيئات tmRNA مثل tRNA يتم تحميلها بالحامض الاميني مثل الالانين (الموضح في الشكل) ولكن الجزيئة لا تحوي على شفرة مقابلة وبدلاً من ذلك تحوي على إطار قراءة قصير ينتهي بشفرة وقف في الجزء المماثل للـ mRNA وهذا يعمل بمثابة شرك لإيقاع عوامل الإطلاق لفتح الرايبوزوم المتوقف. وترتبط جزيئة tmRNA المحملة بالحامض الاميني المطلوب الى الرايبوزوم وكأنه جزيئة tRNA وبذا تزيل جزيئات tmRNA تأثير جزيئات mRNA المعطوبة.

RNA

يمثل RNA أحد الجزيئات العملاقة المهمة في الأنظمة الحيوية وجزيئاته سواء القابلة للترجمة أو غير القابلة للترجمة التي تشترك بشكل فاعل في عمليات التنظيم تقع ضمن مكنون النسخ للخلايا الذي يمثل مرآة للفعالية الوراثية للخلية. ويختلف عن DNA :

- احتوائه على سكر الرايبوز العادي الذي يحوي على مجموعة OH- على ذرة الكربون رقم 2، في حين يحوي DNA على H- في هذا الموقع وهذا ما رشح ان تكون جزيئة الأخير مستودعا" للمعلومات الوراثية لأنها أكثر ثباتا.
- والاختلاف الآخر هو احتواء RNA على القاعدة النتروجينية اليوراسيل الذي لا يحوي على مجموعة مثيل CH_3 - على الذرة الخامسة ويكون بديلا للثايمين.
- يكون RNA في الأغلب بشكل أشرطة مفردة.

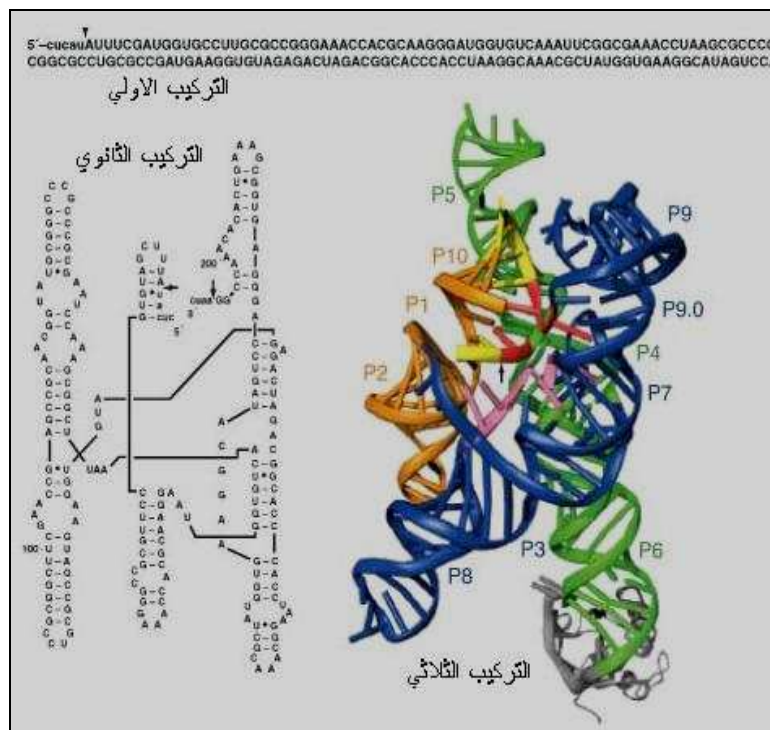
ان فهم تركيب جزيئات RNA يعد مهما في فهم الوظائف التي تقوم بها ومنها التعبير الجيني وتنظيمه. يحدد تركيب الجزيئة باستعمال الأشعة السينية X-ray crystallography والرنين النووي المغناطيسي Nuclear magnetic resonance (NMR) ، وهي طرق مكلفة وتحتاج الى وقت طويل.

توجد ثلاث أنواع رئيسية من جزيئات RNA وهي mRNA الذي يقوم بتوجيه عمليات تخليق البروتينات، و rRNA الذي يقوم بوظائف تركيبية وهي بناء الرايبوزومات وكذلك المشاركة في عمليات تخليق البروتينات. اما الجزيئة الناقلة tRNA فهي متخصصة بكل حامض أميني وتقوم بنقله اثناء عمليات تخليق البروتينات . وهناك عدد لا يحصى من جزيئات RNA الصغيرة وغير المشفرة للبروتينات التي تساهم في عمليات التنظيم كما سيأتي ذكره مستقبلا. وبصورة عامة فان الصفة المميزة لجزيئات RNA انها أشرطة مفردة ولكنها تعمل هجائن ذاتية في بعض المناطق جاعلة منها أشرطة مزدوجة، لذلك كان فهم تركيبها الأولي والثانوي مفتاحا لمعرفة وظائفها، فضلا عن ان تحليل تركيب جزيئات مثل rRNA يكون أساسا في إيجاد العلاقات التطورية بين الأحياء .

مستويات تركيب جزيئة RNA

تركيب هذه الجزيئات يمكن ان يوصف على ثلاثة مستويات :

- **التركيب الأولي** وهو توالي مستقيم يحوي على أربعة من القواعد النتروجينية U,G,C,A.
- **التركيب الثانوي** ويشير الى المستوي الذي يحوي على بعض المناطق المزدوجة ضمن شريط مكون من سلسلة مفردة من تواليات القواعد النتروجينية. في المناطق المزدوجة تشبه مزدوجات Watson & Crick التقليدية الخاصة بأشرطة DNA وفيها تزدوج U-A و G-C ، ولكن قد تكون هناك ازدواجات غير تقليدية اذ يمكن ان تزدوج U-G ، والازدواج الأخير يكون غير ثابت عادة في الأشرطة المزدوجة.
- **التركيب الثلاثي**: ويمثل تركيب مجسم ثلاثي الأبعاد لسلسلة RNA والمستويات الثلاثة من التراكيب موضحة في الشكل 57.3:



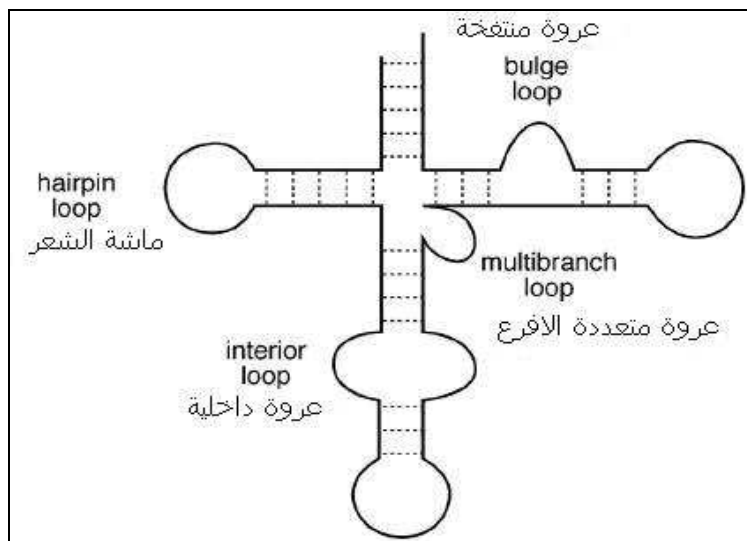
شكل 57.3 : مستويات تراكيب RNA

واعتمادا على ازدواج الأشرطة والحلزونيات في التركيب الثانوي يلاحظ وجود أنواع من التراكيب الثانوية وهي:

- تركيب دبوس او ماشة الشعر
- تركيب بوقي او بشكل فقاعة Bulge loop
- العروات الداخلية Interior loops
- عروات متعددة الأذرع

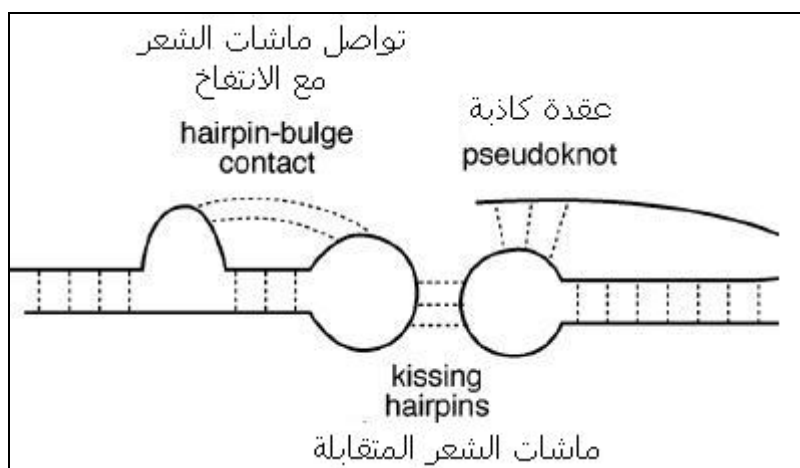
وهذه التراكيب موضحة في الشكل 58.3 .

ففي تركيب الماشة ترتبط بعض مناطقه مكونة ما يسمى الساق كما موضح في الشكل أعلاه. اما تركيب الفقاعة او البوق فان ساقها يحوي على منطقة مفردة الأشرطة تنشئ بين منطقتين من الأشرطة المزدوجة وبذلك تكون بشكل فقاعة على جانب واحد من الجزء المزدوج من الأشرطة . اما العروات الداخلية فهي مناطق مفردة الأشرطة ترتبط الى مناطق مزدوجة الأشرطة وتكون فقاعة على الجانبين.



شكل 58.3: التراكيب الثانوية في بعض جزيئات RNA

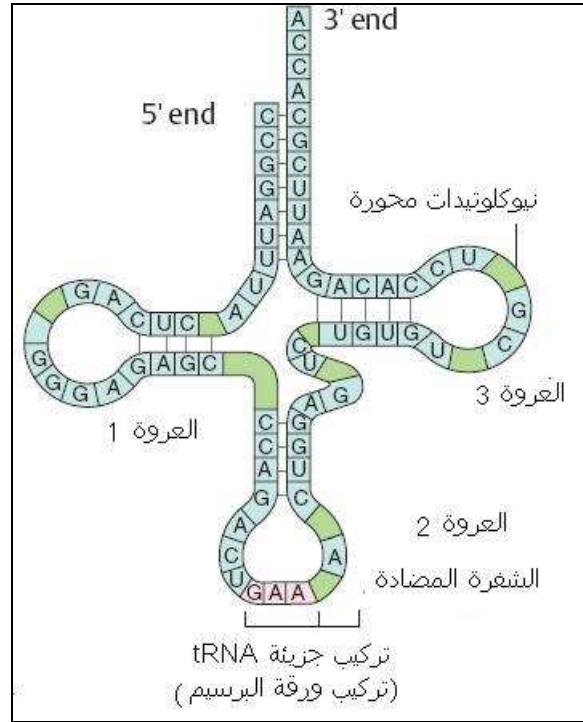
وهناك عروات التفرع وتسمى أيضا مناطق الوصل للحلزونات Helical junctions وهي العروات التي تربط بين ثلاث قطع من تراكيب مزدوجة أو أكثر مكونة تراكيب متعددة الأفرع. وفضلا عن التراكيب الثانوية المذكورة أعلاه فإن العروات يمكن أن تزود قواعد مكونة تراكيب ذات مستويات تعقيد أعلى مثل العقد الكاذبة Pseudoknots وماشيات الشعر المتقابلة Kissing hairpins وتراكيب مشتركة من الفقاعات وماشيات الشعر كما موضح في الشكل : 59.3



شكل 59.3 : تراكيب RNA المعقدة الفائقة Super secondary structures

اذ يلاحظ أن العقد الكاذبة تتكون من ازدواج القواعد من ثملات عروة ما مع ثملات من خارج تركيب الماشية، في حين تتكون ماشيات الشعر المتقابلة نتيجة تكوين أوامر هيدروجينية بين

ثمالات العروات لاثنتين منها. في حين تتكون تراكيب مشتركة من الفقاعات والماشات من تداخل الثمالات لكلا التركيبين. وهذه التراكيب تمثل تراكيب ثانوية فائقة Supersecondary structure وهي نادرة في العادة لذلك تهمل في خوارزميات او برامج التنبؤ التقليدية. واغلب التراكيب المذكورة يمكن ان تظهر في جزيئات tRNA الذي يكون تركيبه بشكل ورقة ثلاثية الفصوص مكونة من حلقة مفردة الأشرطة مع ساقها، والتركيب معقد وفيه مناطق مميزة لأداء الوظائف مثل العروة التي تحمل موقع التعرف على شفرات mRNA المراد ترجمتها والتي يطلق عليها عروة مضادة الشفرة ، والشكل التالي (60.3) يوضح تركيب tRNA الخاص بنقل الحامض الاميني الفيل-النين في الخميرة.

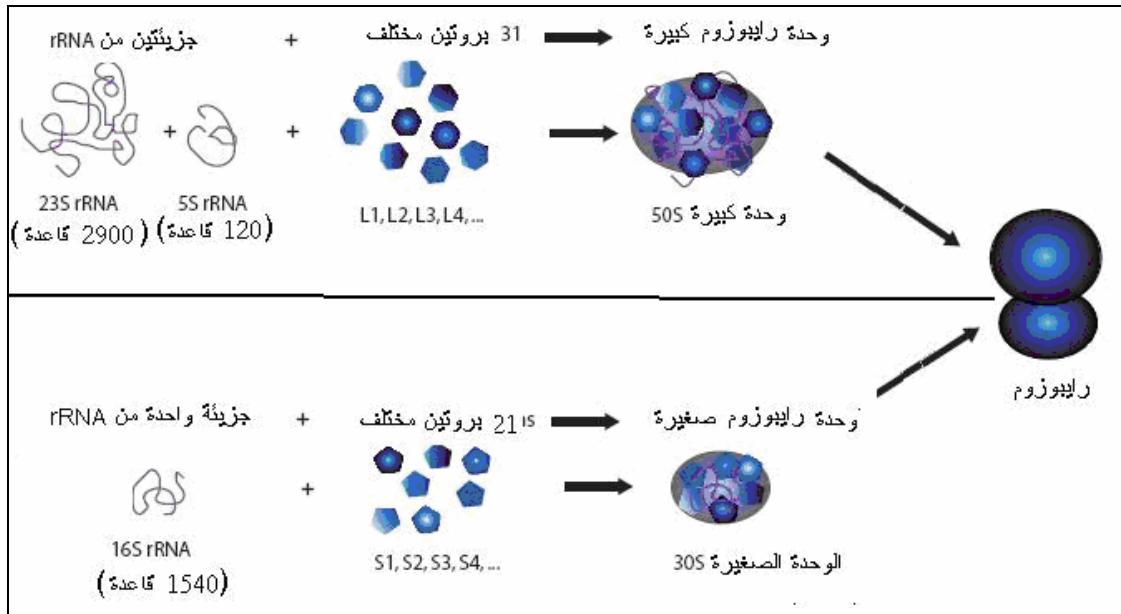


شكل 60.3: تركيب tRNA المشابه لورقة البرسيم

أهم جزيئات RNA العملاقة الثابتة

من المعروف ان جزيئات RNA الناقلة للرسائل مثل mRNA و tmRNA تكون ذات وجود عابر، ولكن هناك جزيئات قد تستمر مع طول حياة الخلية ومنها:

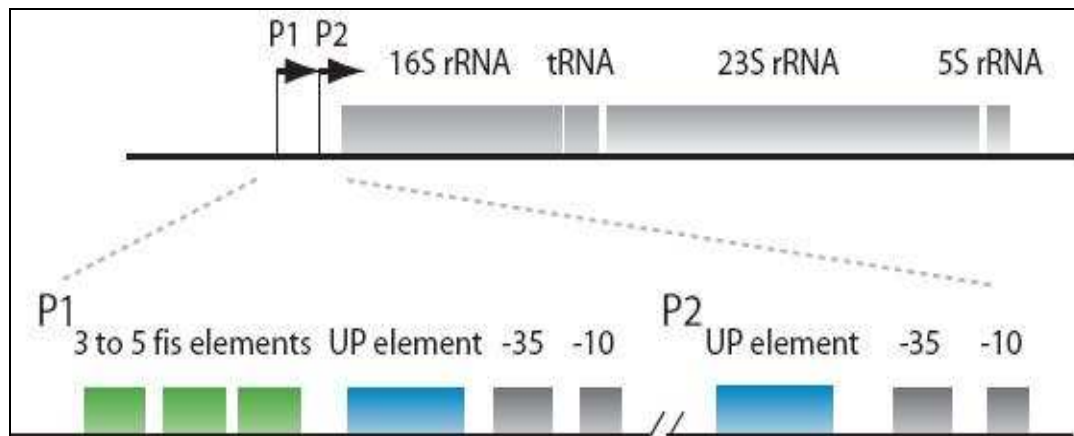
جزيئات RNA الرايبوزومي (rRNA): تشارك في بناء الرايبوزومات مع العديد من البروتينات التي تصل حوالي 50 فضلا عن ثلاث من جزيئات rRNA لتشكل الوحدات الصغيرة او الكبيرة من وحدات الرايبوزوم كما موضح في الشكل (61.3) في الخلايا البكتيرية .



شكل 61.3 : تراكيب وحدات الرايبوزوم في الخلايا البكتيرية

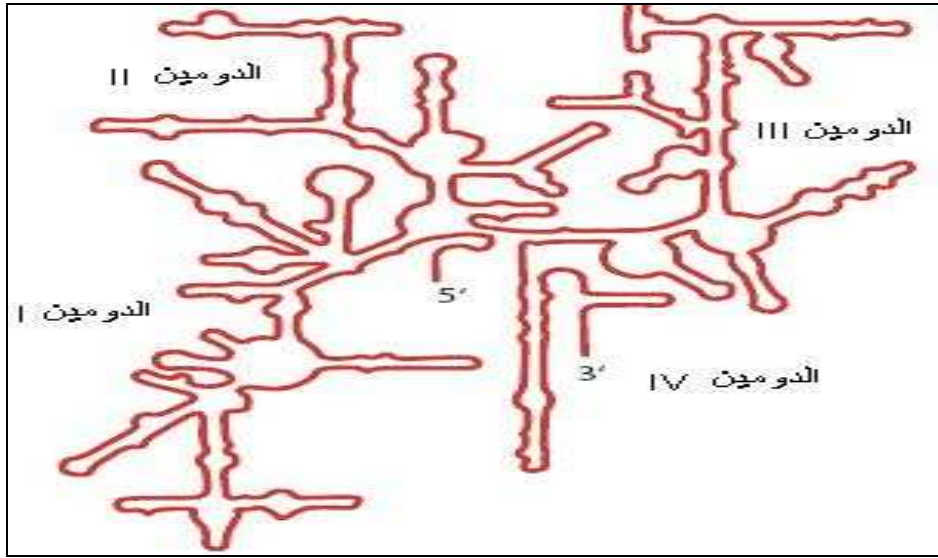
S) Svedberg Unit هي وحدة معامل الترسيب للطرد المركزي الفائق التي تزداد بشكل لا خطي مع الوزن الجزيئي).

في البكتيريا يعد وجود اوبرون واحد للـ rRNA هو الشائع ولكن بعض البكتيريا تمتلك 10 اوبرونات اعتمادا على معدل النمو ويصل العدد ذروته (ربما) في البكتيريا *Photobacterium profundum* التي تحتوي على 15 اوبرون وهي بكتيريا تعطي وميض حيوي ومتعايشة مع بعض القشريات البحرية. كما يمكن لـ aوبرونات rRNA ان تحوي على أكثر من مهده كما في *E. coli* وبعض البكتيريا القريبة منها كما موضح في الشكل 62.3 :



شكل 62.3 : اوبرون rRNA الحاوي على أكثر من مهده في *E. coli*

وهي p1 و p2 وعمل اي منهما يتحدد بطور النمو وظروف التتمية، ففي *E. coli* يلاحظ ان البكتريا تستعمل الممهد p1 عندما تكون في حالة نشاط ونمو، في حين تستعمل p2 في طور الركود Stationary phase وفي البكتريا عموما فان الرايبوزومات تحوي الوحدات الكبيرة (50S) LSU ، والوحدة الصغيرة (SSU) Small subunit . الكبيرة تحوي على 23S rRNA ، 5S rRNA . اما الوحدة الصغيرة فتحتوي على 16S rRNA المكون من 1540 نيوكلوثيريد ويكون تراكيب من عروات وماشات متعددة قد تصل الى أربعة دوميئات كما موضح في الشكل 63.3 :



شكل 63.3 : تركيب 16S rRNA

وهذه التراكيب ثلاثية الأبعاد للـ 16S rRNA تكون مهمة لوظيفة الجزيئة ولذلك كان هناك ثبوت او صيانة لتواليها في البكتريا وربما كان ذلك عائدا الى ان هذه الجزيئة ترتبط بعدد من البروتينات في الوقت نفسه اللازمة لعملية الترجمة، وثبوت الجزيئات رشحا في كثير من الأحيان الى استخدامها في عمليات التصنيف وإيجاد العلاقة التطورية. ولكن مع هذا تحدث بعض التغيرات مثل الطفرات ولكنها تكون مقبولة وذلك لان حدوث الطفرات في موقع ما فإنها تستوعب في الجزء المكمل من المناطق المزدوجة (Stem) للجزيئة وبذا تحصل محافظة على التركيب العام للجزيئة. فضلا عن ان مناطق العروات تكون حرية حدوث الطفرات فيها محدودة نظرا لتداخلها مع البروتينات الرايبوزومية وبذا يكون هناك انتخاب طبيعي تطوري نحو الثبات وصيانة هذه الجزيئات التي تشارك فيها البروتينات بطرق مختلفة.

ومن جهة ثانية فان جينات rRNA في بعض الكائنات غير البكتريا تحوي على الانترونات التي تزال بعمليات الخياطة لتبقى مناطق الاكسونات لتدمج في الرايبوزوم . وعند انتساخ جينات rRNA الثلاث في البكتريا تبدأ من أمام 16S rRNA وتستمر لتعطى جزيئة متعددة الرسالة الثلاث التي تنتهي بعد 5S rRNA ، وتتفلق الجزيئة الناتجة بإنزيمات

خاصة هي Endo-RNase ثم تشذب من ناحية 3' بإنزيمات أخرى . ان موقع اوبرون rRNA وغيرها من الاوبرونات يمكن ان يختلف ولكنها دائما تكون على الكروموسوم الأصلي، ولكن في بعض الأحيان يمكن ان توجد على البلازميدات، فضلا عن ان بعض الجينومات يمكن ان تحوي نسخ إضافية للـ 16S rRNA و 5S rRNA اما للـ 23S rRNA فتكون الحالة نادرة. والملاحظ ان جينات 16S rRNA و 23S rRNA غالبا ما تكون مفصولة بواحد او أكثر من جينات tRNA ، وغالبا ما تكون جينات rRNA واقعة على الشريط القائد من شريطي الكروموسوم (كما سيأتي ذكره في موقع آخر) خاصة في الأحياء التي تحوي على نسخ متعددة فمثلا البكتريا *Photobacterium profundum* وهي بكتريا بحرية تحوي على 15 اوبرون للـ rRNA (المذكورة سابقا) وهذه تكون بمثابة إشارة لمعدلات نموها العالي تحت الظروف المثالية . و *Bacillus cereus* وهي بكتريا تربة مغتتمة الفرص تسبب التسمم الغذائي تحوي على 13 جين من جينات rRNA لذا كانت لها القابلية على التضاعف في دقائق قليلة وتكون الجينات في هذه الحالة بشكل مترادف Tandem وبعض الأحيان تكون مفردة ولكن في العموم تكون قريبة من منطقة بدء تضاعف الكروموسوم.

مقارنة تواليات rRNA في البكتريا

في هذا الموضع لابد من ذكر بعض الحقائق عن جينات rRNA في الأنواع المختلفة، ومنها ان هذه الجينات تكون ثابتة نوعا ما في النوع والتي على أساسها يتم تقسيم أفراد الجنس الى أنواع Speciation. وفي حالة كون تواليات 16S rRNA مفصولة عن 23S rRNA ببعض التواليات التي يطلق عليها فواصل النسخ الداخلية (ITS Internal transcribed spacer) ، فان مثل هذه المناطق تظهر بعض التباين في بعض المجاميع البكتيرية مثل α -Proteobacteria وتستعمل في عملية التمييز الجيني Genotyping . ولوحظ ان حدوث طفرات في هذه الفواصل الداخلية يضفي صفة المقاومة لبعض المضادات الحيوية مثل مجموعة Macrolides و Amino glycosides.

وكما ذكر أعلاها فان المنطقة الخاصة بـ 16S rRNA استعملت في التصنيف وكذلك إيجاد المسافات في العلاقات التطورية للأحياء Long-distance evolutionary relationships ، وتبين من هذه الدراسات ان الاختلاف الرئيس بين البكتريا الحقيقية والاركية يعود الى الاختلاف في rRNA (فضلا عن الاختلافات في تفاصيل عمليات ومكان التضاعف والانتساخ)، ووفقا للقاعدة التي تشير الى ان الاختلاف بنسبة 5% تكون كافية لاعتبار الكائنات أنواعا مختلفة ولكن مع الأخذ بنظر الاعتبار ان جزيئات 16S rRNA هي ليست دائما ثابتة تماما ضمن الأنواع خاصة وان انتقال هذه الجينات يمكن ان يتم أفقيا والذي يؤدي الى إضعاف قوة 16S rRNA كمؤشر تصنيفي وعليه فان وضع تواليات 16S rRNA كحجر أساس لعمليات التصنيف لا تزال محض اختلاف ومع هذا فعند استخدام الجزيئات الأخيرة لإغراض مختلفة مثل التصنيف يفرض الى إعطاء معلومات ذات معنى حول الكائن الحي بذاته، في حين ان استعمالها لإيجاد العلاقات التطورية تهدف الى وصف العلاقات التطورية بين الجينات او الكائنات يوضح ان التصنيف والعلاقات التطورية يخدمان هدفين مختلفين الأغراض وان أساساتهما ليست دائما متداخلة.

علاقة جينات rRNA ومحتواها من AT

بما ان جينات rRNA تعد من المناطق الثابتة الى حد ما من بين باقي جينات الجينوم وعليه فان محتواها من AT يكون قليل التباين مقارنة بباقي جينوم البكتيريا فضلا عن ان نسبتها قليلة مقارنة بالجينوم الكلي او جينات tRNA ، ويكون التباين اقل في الاركيا مقارنة بتلك الخاصة بالبكتيريا، فضلا عن ان 5S rRNA تحوي النسبة الأقل من AT ولكن الأكثر تباينا. ان محتوى جينات RNA من القواعد النيتروجينية يعكس وظيفتها وان الاختلافات في محتوى الجينات من AT يعكس القوى المؤدية الى التطور والتي لا تكون هي نفسها لكل جين في الجينوم.

جزيئات RNA الناقل (tRNA): الجزيئات التي تقوم بنقل الحوامض الامينية لذلك تحتاج الخلايا في الأقل الى 20 جين للـ tRNA (جزيئة لكل حامض أميني) وبما انه هناك وفرة في الشفرات الوراثية للحوامض الامينية (61 شفرة) لذلك وجد مبدأ المراوغة للقاعدة النيتروجينية الثالثة في الشفرة الوراثية التي تسمح لجزيئة tRNA واحدة خدمة عدد من الشفرات من خلال عدم التلاؤم المنضبط. وفي العموم فان البكتيريا الحاوية على عدة اوبرونات للـ rRNA تحوي على عدد من جينات tRNA وبذا يكون تضاعفها عاليا تحت الظروف المثلى (كما ذكر أعلاه)، وبما ان وقت تضاعف البكتيريا غير محدد لكل الأنواع ولكن الحقيقة المذكورة توضح حول علاقة عدد اوبرونات rRNA و tRNA تتضح من ان البكتريات التي تمتلك اوبرون واحد من rRNA وعدد قليل من جينات tRNA تأخذ أياما لتضاعف كما في البكتيريا *Mycobacterium tuberculosis* . وعليه فمن حساب عدد جينات tRNA و rRNA يمكن معرفة نمط حياة الكائن على الأقل من ناحية معدل النمو وان كان هذا التوجه يصيبه بعض الوهن لان هناك عوامل أخرى تحدد عدد النسخ.

وجينات tRNA توجد بأعداد مختلفة ويحصل في بعض البكتريات تضاعف لها ليصل الى 150 جين، ووجود هذه الوفرة من جينات tRNA لحامض معين له علاقة بمدى استعمال الحامض الأميني في بروتينات الكائن. كما ان وجودها قرب مناطق التضاعف في البكتيريا يعني انها تحت تعبير عالي. وفي البكتيريا تكون جينات tRNA مكانا لرسو عدد من العناصر المتحركة مثل بعض القافزات Transposons التي تندمج في هذه الجينات. ومن الملاحظ في الأغلب ان جينات tRNA توجد بشكل مترادف ولو ان بعضها يكون منشرا ومشتتا في الكروموسوم ، وفي *E. coli* فان النوع المترادف يكون له مهاد واحد وينتج رسائل متعددة تفلق فيما بعد بإنزيمات Endo-RNases لغرض إطلاق رسائل او جزيئات منفصلة وتشذب الجزيئات الناتجة بواسطة إنزيمات القطع الداخلي لإعطاء جزيئات tRNA ناضجة وتكون عملية النضج بإزالة التواليات من جهة 5' ابسط من عمليات النضج عند النهاية 3'، وعلى العموم فان عمليات الإزالة والتشذيب تحتاج الى وجود تواليات غنية بالقواعد النيتروجينية اليوراسيل والادينين (AU elements) التي تكون ثابتة تقريبا في كل جينومات الأحياء .

جزيئات tmRNA

جزيئات RNA صغيرة ومزدوجة مشابهة لكل من tRNA و mRNA ولذلك سميت بهذا الاسم لأنها تمتلك صفات كل من tRNA و mRNA ويطلق عليها تسميات أخرى مثل

RNA SsrA أو 10Sa RNA نظرا لوجود توالي يعد طمغة من 10 حوامض أمينية وتشكل جزءا من حقل دراسات RNA (RNAomics). وهي تساهم بالدرجة الرئيسية في عمليات الترجمة غير الطبيعية المسماة Translation-trans-translation فضلا عن مشاركتها في تنظيم فعاليات الجين ودورة الخلية Cell cycle. وفي عملية الترجمة المخالفة Trans-translation التي تكون عمليات الترجمة فيها مضطربة تدخل هذه الجزيئات بشكل مشابه للـ tRNA الى الرايبوزومات ثم تقحم توالياتها المشابهة للـ mRNA في التوالي المسئول (Ribosome decoding sequence) ويقوم الرايبوزوم بترجمة الحوامض الامينية المشفر لها بـ tmRNA على النهاية الكربوكسيلية للبيتيدات الناتئة (قبل توقف الترجمة). وفي أحيان أخرى فانها — بصورة عامة — تقوم بإزالة معقدات الترجمة المتعثرة في الخلايا البكتيرية مثلا (والتي مر ذكر بعض جوانبها في موقع سابق).

وجزاء mRNA الذي يشارك في العملية أعلاه (وهي عملية ترجمة شاذة) ينطلق من الرايبوزوم ويتفكك بسرعة، فضلا عن وجود بروتينات عدة تحلل البروتين الناتج من منطقة التوقف وبذا يتم التخلص من نواتج عملية الترجمة الشاذة، لتستمر بعدها الحالة الطبيعية. وجزيئات tmRNA وملحقاتها من العوامل المشاركة في الترجمة المخالفة موجودة في بعض الأنواع البكتيرية وتكون مهمة في أمراضية الأنواع الحاوية عليها وكذلك تشارك في عمليات التمايز والاستجابة للاجهادات وتنظيم بعض المسارات الايضية. وفي *E. coli* تكون الجزيئات اكبر من جزيئات tRNA العادية وأكثر وفرة ولوحظ ان هناك عملية ترجمة مخالفة واحدة تحصل لكل 250 عملية ترجمة عادية عند نمو *E. coli* في المزارع الطبيعية. وهذا يعني ان كل رايبوزوم يقوم بعملية ترجمة مخالفة او شاذة لكل دورة خلية.

جزيئات RNA المنظمة

هذه الجزيئات في معظمها جزيئات صغيرة ، وإنتاجها او مراحل عملها تخالف القضية المحورية في علم الحياة البايولوجي والتي تشير الى "ان المعلومات الوراثية تنساب من DNA (الجينات) الى الجزيئات الوسطية RNA ثم تترجم الى بروتينات " (المارة الذكر) . ولكن هذه الجزيئات الصغيرة التي يطلق عليها في العموم جزيئات غير مشفرة ncRNA يمكن ان تنتسخ من أجزاء متعددة من الجينوم وليس جينات محددة وتقوم بعمليات التنظيم ضمن عملية شاملة يطلق على بعضها تداخل RNA interference (RNAi) . وبطبيعة نشوءها فهي تشابه كل من rRNA و tRNA التي تتوقف رحلتها بعد الخطوة الأولى من الانتساخ. وقد اكتشفت أعداد وأنواع كثيرة منها التي تساهم في تنظيم التعبير الجيني والتي يمكن ان ترتبط بالعديد من البروتينات مثل عائلة Ago proteins وسيأتي ذكرها في المواقع الخاصة بها.

وفي العموم فان كل ما يتعلق بجزيئات RNA هذه من حيث التركيب والوظيفة واي مواصفات أخرى تدرج تحت دراسة مكنون النسخ Transcriptomics الذي يتعامل بشكل رئيس مع ما يقابله من مكنون النسخ .

هناك العديد من الجزيئات غير المشفرة الأخرى التي تعمل في مجال التنظيم او غيرها من الوظائف التي لم تعرف لحد الآن وتطلق عليها تسميات مختلفة وهي تعد من الجزيئات الفعالة وتطلق عليها تسميات مختلفة مثل npcRNA (non-protein coding RNA) او

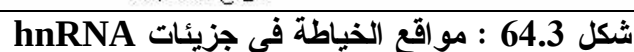
nm RNA (non-messenger RNA) أو (Small non-coding messenger RNA) sRNA snmRNA (Small RNA) وهذه تستعمل عادة مع الجزيئات غير المشفرة الموجودة في البكتيريا والجدول (6.3) يوضح بعض هذه الجزيئات ومجالات عملها:

جدول 6:3 بعض أنواع جزيئات RNA غير المشفرة

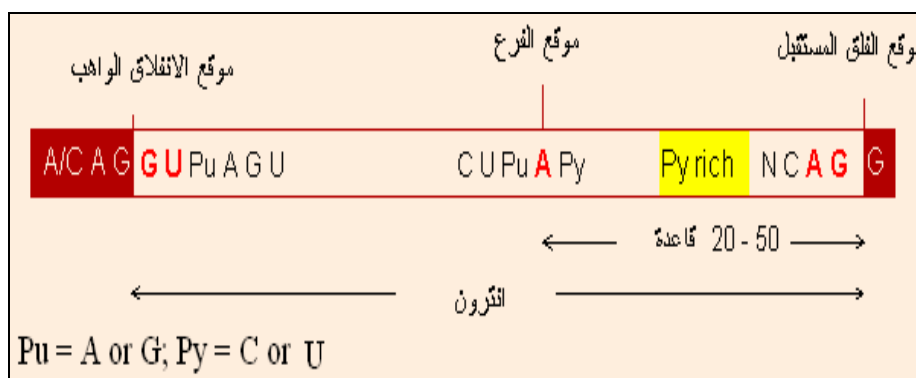
نوع جزيئة RNA	المختصر	الوظيفة	التوزيع
جزيئات تميز الإشارات Signal recognition particle RNA	7SL RNA or SRP RNA	الاستقرار في الأغشية	كل الأحياء
الجزيئات المشاركة في التحويرات بعد الترجمة أو تضاعف DNA			
الجزيئات الصغيرة النووية	snRNA	عمليات الخياطة ووظائف أخرى	الاراكيا وخلايا حقيقية النواة
جزيئات النويات الصغيرة	snoRNA	تحويل نيوكليوتيدات RNA	الاراكيا وخلايا حقيقية النواة
Smy RNA	Smy	الخياطة للـ mRNA trans-splicing	الديدان
Guide RNA الدليل	gRNA	تحويل جزيئات mRNA	الميتوكوندريا و kinetoplastids
Ribonuclease P	RNase P	نضوج جزيئات tRNA	كل الأحياء
Ribonuclease MPR	RNase MPR	نضوج جزيئات rRNA ، تضاعف DNA	الخلايا حقيقية النواة
Y RNA	Y RNA		الحيوانات
RNA نهايات الكروموسومات		تخليق نهايات الكروموسومات Telomeres	معظم الخلايا حقيقية النواة
جزيئات RNA المنظمة			
Antisense RNA النقيض	aRNA	توهين الانتساخ، تفكيك mRNA، تثبيت mRNA في بعض عمليات الترجمة	كل الأحياء
Cis-natural antisense transcripts		تنظيم الجينات	
Clustered regulatory interspaced short palindromic repeats	CRISPR or crRNA	مقاومة المتطفلات باستهداف جزيئات DNA العائدة لها	البكتيريا والاراكيا
جزيئات RNA طويلة غير مشفرة	Long ncRNA	وظائف متعددة	خلايا حقيقية النواة
micro RNA	miRNA	تنظيم الجينات في الخلايا حقيقية النواة	معظم الأحياء

الخياطة Splicing

لغرض إنتاج جزيئات mRNA جاهزة للترجمة تصدر من النواة الى السايטوبلازم لابد من إجراء عملية الخياطة . والخياطة بخطوطها العريضة تتمثل بإزالة الانترونات ولحم الاكسونات المتبقية. وتتم عمليات فلق الانترونات لأزالتها بناءا على وجود إشارات الفلق، فمعظم الانترونات تبدأ بالتوالي GT (GU في RNA) ويسمى معطي الانفلاق وتنتهي بالتوالي AG ويسمى مستلم الانفلاق باتجاه من 5' الى 3' كما موضح في الشكل التالي (64.3) وبعض هذه الإشارات قد يكون مغطى ببروتينات منظمة تساعد في الخياطة البديلة كما موضح في الشكل



وعلى العموم فإن وجود هذه التواليات في قطعة ما لا يعني وجود الانترونات ولكن هناك توالي آخر يسمى موقع الفرع Branch site هو الذي يحدد موقع الفلق الذي يقع على بعد 20-50 قاعدة الى يسار مستلم الانفلاق وتوالي الموقع في RNA هو $CU(R/C) A (C/U)$ حيث تكون A ثابتة في كل الجينات كما موضح في الشكل 65.3 :



شكل 65.3 : توضيح موقع الفرع Branch site في مواقع الخياطة

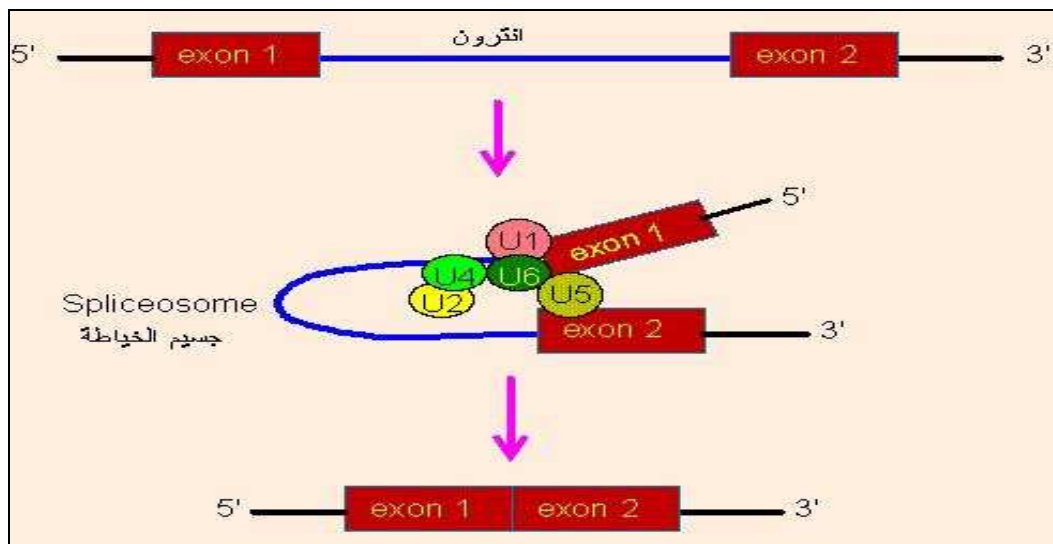
(R يعني البيورينات)

وفي 60% من الجينات القابلة للخياطة تكون (A/C) AG من الجهة اليسرى ومن الجهة اليمنى هو G كما موضح في الشكل أعلاه.

آلية الخياطة

عملية الفلق والخياطة عملية معقدة تشمل اشتراك في الأقل 5 من جزيئات RNA الصغيرة النووية snRNA هي U_1, U_2, U_4, U_5, U_6 و U_3 لا يشترك في عملية الخياطة وإنما يعمل في معالجة pre-rRNA ، وفضلا عن هذه الجزيئات يشارك عدد من البروتينات الرايبوزية مكونة جسيم الخياطة Spliceosome الذي له معدل ترسيب 60S، وتحدث عدداً من التفاعلات الإنزيمية يكون للقاعدة A المذكورة أعلاه دوراً رئيساً فيها ويتم إزالة الانترونات ثم ربط الاكسونات المتجاورة مع بعضها كما موضح في الشكل (66.3) .

وفي حالة حصول طفرات في إشارات الفلق تعيق المسار التقليدي للخياطة يتم اللجوء الى آلية أخرى هي تجاوز الاكسون Exon skipping (وهي إزالة الاكسون والانترون في عملية خياطة واحدة) .



شكل 66.3 : آلية خياطة جزيئات RNA

الخياطة البديلة Alternative splicing

تمثل أحد آليات التنظيم وهذه الآلية تعد الأساس في الاقتصاد الجيني Gene economy. فبهذه الآلية يمكن ان تنتج أنواعاً مختلفة من mRNA التي تعطي بروتينات مختلفة الوظائف وتعتمد على وجود إشارات الفلق وكذلك مواقع إضافة ذيل الأدينين.

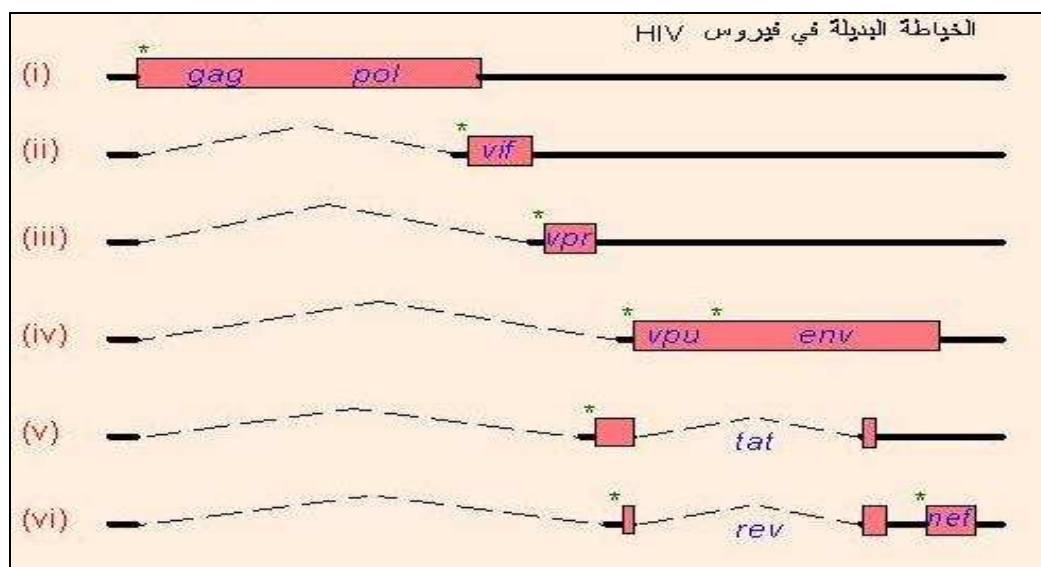
الاقتصاد الجيني

تمثل أحد أوجه التعقيد الوراثي في الأحياء الراقية ، ففي الإنسان هناك أكثر من 100,000 بروتين معبر عنه اعتماداً على تحليل طمغات التوالي المعبر عنها EST analysis (Expressed sequence tag) وهذا العدد من البروتينات ينتج من حوالي 30,000 جين ، ولو كان كل جين يعبر عن بروتين إذن لتطلب الأمر وجود عدد مقابل لعدد البروتينات، ولكن ظاهرة اقتصاد الجين هي الإستراتيجية التي تمثل القوة التفسيرية الإضافية التي تستعملها الخلايا حقيقية النواة للوصول الى العشرات من الأنماط المظهرية المتعددة والمختلفة ، وتعتمد في ذلك على عدة آليات أهمها الخياطة البديلة، وتعني ربط اكسونات مختلفة من جين واحد لإعطاء نسخ مختلفة وبالتالي بروتينات مختلفة وهي قريبة من الآلية المعروفة بإدلاف الاكسونات التي تربط اكسونات مختلفة من جينات مختلفة لإنتاج نسخ الجينات الشائعة في الخلايا حقيقية النواة. وفي الإنسان هناك أكثر من ثلثي الجينات تعاني من الخياطة البديلة تؤدي الى إنتاج حوالي 90% من البروتينات الكلية ، وفي ذبابة الفاكهة مثلاً الجين DSCAM يحوي على 115 اكسون وتظهر فيه الخياطة البديلة وتؤدي الى إنتاج 38,000 بروتين مختلف. ولذلك فإن القابلية الاستثنائية لهذا التنوع البروتيني الكبير وبوظائفها المختلفة توضح مدى تعقيد الجينومات وتوضح أهمية الانترونات من النواحي التطورية والتي تعمل بمثابة فواصل بين القطع المشفرة (الاكسونات).

ومن المظاهر الأخرى الخاصة باقتصاد الجين هو وجود الخياطة المتباعدة Trans-splicing ، وتحصل هذه عندما يكون هناك انتساخ لأكسونات جين واحد من شريطي DNA المتقابلة (أي الحساسة والنقيضة) وبتشكيلات مختلفة وترتبط لتكون جزيئات mRNA لإعطاء بروتينات مختلفة.

وبالعودة الى الخياطة البديلة فهناك حوالي 80% من جينات الإنسان تعاني من الخياطة البديلة، ومن الدراسات الجينومية المقارنة في اللبائن والحشرات أمكن تحديد مواقع الخياطة البديلة بطريقة مباشرة من تحديد تواليات الجينومات، ومن دراسة ثبوت التواليات وأطوال الأكسونات تبين الترابط بين الخياطة البديلة ومختلف العمليات التطورية مثل أكسنة تواليات Alu (Alu exonization) التي تؤدي الى إيجاد أكسونات جديدة أو فقدها، كما أشارت هذه الدراسات ضمن مجال المعلوماتية الحيوية ان الجينات التي لا تحصل فيها خياطة بديلة وتشكل 20% من الجينات في الإنسان تحصل فيها حالة تجاوز الأكسونات.

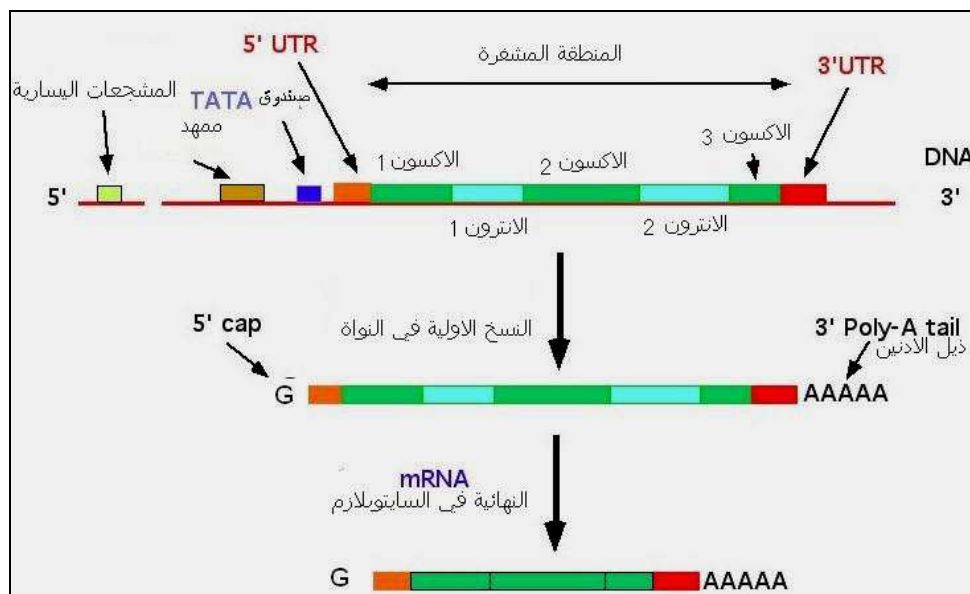
وفي الفيروسات مثل فيروس HIV فان جزيئات pre-mRNA يمكن ان تحوي عدة إشارات للخياطة البديلة مؤدية الى إنتاج عدد من جزيئات mRNA الناضجة، فالرسالة الأولية تكون حاوية على الجينات التسعة للفيروس وهي tat ، rev ، nef ، env ، vpr ، vpu ، vif ، pol ، gag وبالتالي يمكن الحصول على البروتينات من هذه الجينات بالخياطة البديلة، او بالمسح غير الدقيق لشفرة البدء او بواسطة انحراف عملية ارتباط الرايبوزوم كما موضح في الشكل 67.3 :



شكل 67.3 : الخياطة البديلة لجينات الفيروس HIV

وهذا يشير الى إمكانية إنتاج بروتينات مختلفة بطريق غير طريق الخياطة البديلة عند الحاجة. وبعد كل هذه المعالجة الأولية لإنتاج mRNA يتم تحويل الأخيرة وذلك بإضافة مثبت التركيب Cap عند النهاية 5' بواسطة معقد إضافة الغطاء المرتبط بـ RNAP وهو إذن يميز جزيئات mRNA ويساعد في تصديرها من النواة وتميزها من قبل الرايبوزوم ومنع RNases من

تفكيكه. وعلى الطرف يتم إضافة ذيل الأدينين عند النهاية 3' بواسطة الإنزيم Poly(A)polymerase وتضاف حوالي 200 قاعدة وتكون إضافتها ضرورية لإنهاء الانتساخ والتصدير من النواة والترجمة الى البروتين والحماية من التفكيك بإنزيمات القطع الخارجي. ومجمل العمليات أعلاه موضحة في الشكل 68.3 :



شكل 68.3 : خطوات معالجة جزيئات RNA

تأثيرات الخياطة البديلة الوظيفية

يمكن معرفة التغيرات الحاصلة في تواليات النسخ الناتجة من الخياطة البديلة باستعمال وسائل المعلوماتية الحيوية ولكن لمعرفة التغيرات الوظيفية لابد من إجراء العمل المختبري لمعرفة الكيفية التي تغير إليها البروتين الناتج من الترجمة وتداخلاته مع الجزيئات الأخرى. فالخياطة البديلة يمكن أن تؤثر في دومينات البروتينات وأخطرها إزالة دومين معين من البروتين إزالة كاملة وخاصة في الدومينات التي تشارك في تداخل البروتينات مع بعضها وبالتالي تلغي تداخل البروتينات الأصلي Default في الشبكات الخلوية مما يؤدي إلى تغير المسار الأيضي الذي تشترك فيه. ومن التأثيرات الأخرى إزالة الدومين الذي يثبت البروتينات الغشائية في الغشاء مما يحول البروتين إلى بروتين مفروز أو ذائب. ومن التأثيرات الأخرى أنها قد تؤثر في التفاصيل الدقيقة والصغيرة مثل إضافة قطعة صغيرة من تواليات حوامض أمينية وحتى إضافة حامض أميني واحد وهذه وجدت بنسبة 5% في جينات الإنسان. ومثل هذه التغيرات يمكن أن تؤثر في الهيئة التركيبية للبروتين ومن ثم وظيفته وبذلك تؤثر الخياطة البديلة في مكون البروتين للكائن.

ومن جهة أخرى يلاحظ أن الخياطة البديلة تساهم في عمليات التطور كما هو الحال في أكسنة تواليات Alu والتي تكون لها تأثيرات إيجابية ومن جهة ثانية قد تكون لها تأثيرات سلبية مثل حدوث الأمراض. وعند دراسة تأثير الضغط الانتخابي الذي يتعرض له الكائن في الخياطة

البديلة، ووجد ان بعض الجينات التي تتعرض للخياطة البديلة تتفاوت ثبوتيتها بين الكائنات، فالأحداث التي تؤدي الى إنتاج بروتينات غير ثابتة تحدث بمعدل أوطأ مقارنة بالتغيرات الأخرى وبذا تلعب الخياطة البديلة دوراً مهماً في التطور من خلال الأكسنة أو إيجاد اكسونات جديدة وهذه الفعاليات تتأثر بالضغط الانتخابي وكذلك بطول الاكسون الناتج من اي آلية الذي يجب ان يكون طوله مضاعفات لثلاث نيوكلوديدات (يعني شفرة وراثية) بحيث لا تؤثر في أطر القراءة عند الترجمة.

البروتينات

تمثل النواتج النهائية لعمليات التعبير الجيني (الانتساخ) والتعبير البروتيني (الترجمة) ، وفي هذا الجزء سيتم مناقشة بعض الجوانب الخاصة بالبروتينات .

البروتينات مكوثرات حيوية هي الأكثر تنوعا ، وحدات بناءها الحوامض الامينية، وتعد من الجزيئات الحيوية العملاقة ولكنها تبقى ضمن مدى الجزيئات النانوية Nanoparticles (1-100 نانومتر) اعتمادا على حجمها الفيزيائي، وتسمى حصان الشغل لأنها مسئولة عن وظائف متعددة مثل الوظائف التركيبية والنقل والتنظيم وتضخيم الإشارات وغيرها.

تركيبها الأساسي (الأولي) يتكون من سلسلة من الحوامض الامينية التي ترتبط بأصرة ببتيديّة التي تعرف كل منها بالشمالة Residue بعد ارتباطها مع بعضها، والجزيئات المكونة من 40 شمالة يطلق عليها الببتيدات . وتحتوي البروتينات العديد من المجاميع الفعالة مثل الكحولات والثايولات Thiols واسترات الثايول ومجاميع الكربوكسيل ومجاميع Carboxamide فضلا عن وجود المجاميع القاعدية، وهذه المجاميع يمكن ان تحدد الأطر العامة لفعالية البروتينات.

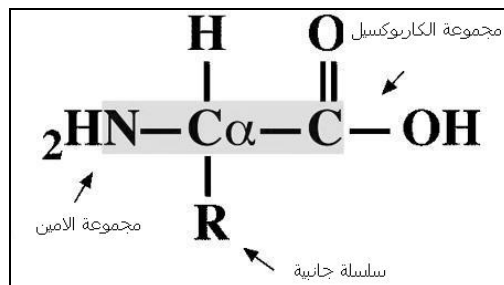
والمجال الذي ضمننت فيه في مجال المعلوماتية الحيوية هو المكنون البروتيني ويعني كل مجاميع البروتينات المعبر عنها في الخلية اي انها المنتجات المترجمة للجينوم ، و Proteomics يعني دراسة المكنون البروتيني وبشكل أخص هو كل التحليلات المترامنة للبروتينات المترجمة في الخلية، ويشمل تحديد كميات البروتينات ومواقعها في الخلية والتحويلات التي تجري عليها بعد الترجمة والتداخلات فيما بينها ، فضلا عن تحديد وظائفها، ومثل هذه الدراسات تؤدي الى معرفة وفهم الوظائف الخلوية للبروتينات وبالتالي إدراك وظائف الجينات، وذلك لان مكنون النسخ لوحده لا يعطي أجوبة واضحة عن الوظائف في الخلية لانه ليست هناك علاقة واحد بواحد بين mRNA والبروتينات في الخلية، كما ان الجينات في جينوم الخلايا حقيقية النواة تنتج نواتج مختلفة لوجود عملية الخياطة البديلة وكذلك عمليات التصحيح RNA editing وغيرها من العمليات . وهذا يعني ان هناك عدة بروتينات متميزة يمكن ان تنتج من الجين الواحد . فضلا عن ان تعقيد البروتينات يمكن ان يزداد بعمليات التحويل بعد الترجمة وتوجيه البروتينات وتداخلها مع بعضها. وعليه فان عدم ارتباط mRNA مع البروتينات يعني ان التعبير البروتيني هو الذي يعطي الصورة الأوضح لدراسة وظائف الجينات اذ تمثل البروتينات المعالجة النهائية سريان المعلومات من DNA الى نهايتها وهو تخليق البروتينات المتكاملة.

تقوم البروتينات بكل الفعاليات الحيوية (تقريبا) التي تجري في الأنظمة الحيوية وبصفة رئيسة كإنزيمات ، ولكن هذه الفعاليات تتم بعد ان تجري العديد من التحويلات على الجزيئات الأولية التي تعاني من الطوي Folding التي تتم بعدد من التداخلات مثل الأواصر الهيدروجينية والتداخلات الكارهة للماء وغيرها من التداخلات (كما سيأتي ذكره لاحقا). ومن مجمل الفعاليات التي تقوم بها المشاركة الفاعلة في عمليات انقسام ونمو الخلايا والأنسجة وتمايزها، والسيطرة على نقل الإشارات من خارج الأنظمة الحيوية الى داخلها، وكذلك السيطرة على التوازن الأيوني للجسم بالسيطرة على نقل المواد من وإلى الخلايا، وتساهم ايضا في الوظائف الميكانيكية كما في الجلد والأوعية الدموية وغيرها من المناطق فضلا عن مساهمتها في بناء المنصات والهياكل الخلوية وكذلك في الأنسجة الرابطة والأنسجة الواقية.

وتكمن أهمية دراسة البروتينات في كونها تمثل الناتج المباشر للتعبير عن المعلومات الوراثية. فضلا عن انها عصب الحياة في التفاعلات التي تجري وان الجزيئات الحيوية الأخرى تكاد تكون نواتج لفعاليتها.

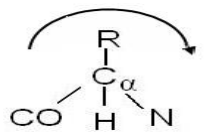
الحوامض الامينية

جزيئات حيوية تركيبها الأساسي ذرة كربون يطلق عليها $C\alpha$ ترتبط بأربعة مجاميع، أبسطها الهيدروجين، وترتبط الى مجموعة الأمين ($-NH_2$) وكذلك مجموعة الكربوكسيل ($-COOH$) وهذه المجاميع عادة تكون ثابتة، اما الموقع الرابع فيرتبط الى السلسلة الجانبية R والتي تكون مختلفة وهي التي تضيف الصفات على الحامض الاميني، والتركيب الأساسي موضح في الشكل 69.3 :



شكل 69.3 : تركيب الحامض الاميني

والتركيب الأساسي للحوامض الامينية متشابه. والتركيب اليساري L-form (الشائع في بروتينات الأنظمة الحيوية) يمكن ان يقرأ وفق قاعدة CORN ($CO-R-N$) كما في المخطط الآتي:



وتقرأ باتجاه عقرب الساعة والملاحظ ان ذرة $C\alpha$ هي الذرة الرئيسة التي ترتبط اليها المجاميع الأخرى وهي التي بها يحدد رقم موقع الحامض الاميني في التواليات البروتينية والذرة $C\beta$ هي الذرة الثانية من السلسلة الجانبية وليس ذرة الكربوكسيل وتختلف كثافتها الالكترونية وتمثل إحدى الصفات المستعملة كأساس لبعض برامج التنبؤ بتركيب البروتينات (كما سيأتي ذكره لاحقا). وتسري تسمية الذرات الأخرى على الذرات المرتبطة فالهيدروجين المرتبط بذرة $C\alpha$ يسمى α -hydrogen والمرتبط بـ $C\beta$ يسمى β -hydrogen وهكذا. وتفضل الجهات المعنية مثل International union of pure and applied chemistry (IUPAC) استعمال الأرقام لترقيم ذرات الكربون بدلا عن الحروف الإغريقية. ويرمز للحوامض الامينية اما بشفرة ثلاثية مكونة من أول ثلاث أحرف لاسم الحامض مثل Valine تكون Val او باستعمال الشفرة الأحادية V كما موضحة في الجدول 7.3:

جدول 7.3 : الحوامض الامينية وشفرات قراءتها وكتابتها

الشفرة الثلاثية Three letter code	الشفرة الأحادية One letter code	الحامض الاميني	
ala	A	Alanine	ألانين
arg	R	Arginine	أرجنين
asn	N	Asparagine	اسبارجين
asp	D	Aspartic Acid	حامض الاسبارتيك
cys	C	Cysteine	السيستئين
glu	E	Glutamic Acid	حامض الكلوماتيك
gln	Q	Glutamine	كلوتامين
gly	G	Glycine	كلايسين
his	H	Histidine	هستيدين
ile	I	Isoleucine	ايزوليوسين
leu	L	Leucine	ليوسين
lys	K	Lysine	لابسين
met	M	Methionine	ميثايونين
phe	F	Phenylalanine	فنيل النين
pro	P	Proline	برولين
ser	S	Serine	سيرين
thr	T	Threonine	ثريونين
try	W	Tryptophan	تربتوفان
tyr	Y	Tyrosine	تايروسين
val	V	Valine	فالين

ويتراوح معدل الوزن الجزيئي للحوامض الامينية بحدود 110 دالتون ، ولذلك تكون الأوزان الجزيئية للبروتينات العامة التي تشارك فيها والتي تتراوح أطوالها بين 50-200 ثمانية بين 5500-200,000 دالتون. والتركيب الموصوف أعلاه يكون عاما للحوامض الامينية ما عدا البرولين حيث ان ذرة الهيدروجين استبدلت بأصرة مربوطة الى السلسلة الجانبية. وتكون ذرة الكربون ألفا هي مركز التركيب الجسم لكل حامض أميني. واعتمادا على السلاسل الجانبية وطبيعتها وإمكانية تداخلها مع البيئة المحيطة تحدد صفات الحوامض . ومن هذه الصفات الحجم، محبة الماء، الشحنة، المجاميع الفعالة، وهذه بدورها يمكن ان تقسم الى أقسام ثانوية. والسلاسل الجانبية تكون مهمة جدا في إضفاء الصفات على البروتينات التي تدخل في تركيبها فمثلا السلاسل الكارهة للماء تميل ان تكون مدفونة الى الداخل (في اغلب الأحيان) في حين تكون السلاسل المحبة للماء الى الخارج عند وجود البروتينات في البيئات المائية التي تمثل البيئات العامة للأنظمة الحيوية .

الحوامض الامينية الكارهة للماء Hydrophobic

تشمل الليوسين والايزوليوسين والفنيل-النين والفالين، وبدرجة اقل التايروسين ، الالنين والتربتوفان والسستئين . وهذه تقسم بدورها الى حوامض أمينية كارهة للماء اليفاتية Aliphatic مكونة من سلسلة كربونية مستقيمة مثل الالنين والفالين والليوسين والايزوليوسين والميثايونين، في حين تكون المجموعة الثانية هي الحلقية والتي تحوي على حلقات مختلفة مثل الفنيل-النين والتربتوفان ويلحق بها البرولين.

الحوامض الامينية المشحونة

تقسم الى حوامض مشحونة بشحنة سالبة كما في حامض الكلوتاميك وحامض الاسبارتيك لانها تحوي على Amide group وعند الرقم الهيدروجيني الفسلجي تكون مشحونة بشحنة سالبة. اما المشحونة بشحنة موجبة فهي الارجنين والهستيدين واللايسين وهي قاعدية التفاعل والحوامض الامينية المشحونة تكون في الغالب محبة جدا للماء ولذلك توجد في المناطق الخارجية من البروتين.

الحوامض الامينية المحبة للماء Hydrophilic

تضم السيرين والثريونين التايروسين والاسبارجين والكلوتامين. وقد تكون هذه المجموعة قطبية او غير قطبية.

الحوامض الامينية ذات الصفات الخاصة

الملاحظ ان هناك تداخل في تصنيف الحوامض الامينية، فقد يكون أحد الحوامض حاملا لصفيتين فهو يمكن ان يدخل في مجموعتين. كما ان هناك بعض الحوامض الامينية التي لها خصوصية مثل الميثايونين والسستئين الحاوي على الكبريت ويكون الأخير أواصر ثنائية الكبريتيد مع ثمانية مماثلة له في موقع آخر من البروتين ، كما ان الكلايسين يمثل ابسط الحوامض الامينية كون سلسلته الجانبية مكونة من ذرة هيدروجين فقط، ولذلك يكون أكثر الحوامض الامينية مرونة ويؤدي الى تكوين تشكيلات تركيبية لا تظهرها الحوامض الأخرى. وكذا الحال مع البرولين الذي له تركيب خاص فسلسلته الجانبية تكون أصرة مع مجموعة الأمين الخاصة به مكونا تركيبا حلقيا غير مرن ولذلك لا يمكن ان يوجد في بعض مفاصل البروتين.

ومن الحوامض الامينية ذات الصفات الخاصة هو الهستيدين الحاوي على حلقة Imidazole ويمكن ان يكون متعادلا (بدون شحنة) او مشحونا بشحنة موجبة اعتمادا على البيئة المحيطة، وهو من الحوامض التي تضيف صفاتها والدفع بها ان تكون في المراكز الفعالة من البروتينات حيث يقوم بدور تبديل الأواصر وغيرها من الفعاليات.

ويمكن إجمال تراكيب وبعض صفات الحوامض الامينية بالجدول 8.3 :

جدول 8.3 : مجمل تركيب وصفات الحوامض الامينية

$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{H} \end{array}$ كلايسين Gly (G)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array}$ الالانين Ala (A)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array}$ فالين Val (V)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array}$ ليوسين Leu (L)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}-\text{CH}_2 \\ \quad \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$ ايزوليوسين Ile (I)	
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_2\text{N}^+-\text{C} \\ \\ \text{Cyclopentane ring} \end{array}$ برولين Pro (P)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$ فيلالانين Phe (F)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{CH} \\ \quad \\ \text{C}_6\text{H}_4 \quad \text{NH} \end{array}$ ترينوتروفان Trp (W)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$ سيسثيون Cys (C)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$ ميثايونين Met (M)	
Neutral amino acids, nonpolar side chains					حوامض امينية متعادلة ، سلاسل جانبية لا قطبية
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2\text{OH} \end{array}$ سيرين Serine Ser (S)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CHOH} \\ \\ \text{CH}_3 \end{array}$ ثريونين Threonine Thr (T)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_4 \\ \\ \text{OH} \end{array}$ تايروسين Tyr (Y)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{H}_2\text{N} \end{array}$ اسباراجين Asparagine Asn (N)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{H}_2\text{N} \end{array}$ كلوتمين Glutamine Gln (Q)	
Neutral amino acids, polar side chains					حوامض امينية متعادلة ، سلاسل جانبية قطبية
1. Basic (positively charged)					2. Acid (negatively charged)
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH} \\ \\ \text{C}=\text{NH}_2^+ \\ \\ \text{H}_2\text{N} \end{array}$ ارجينين Arg (R)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2=\text{NH}_3^+ \end{array}$ لايسين Lys (K)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_5\text{H}_4\text{N}^+ \end{array}$ هستيدين Histidine His (H)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$ حامض الاسباريك Asp (D)	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$ حامض الكلوتميك Glu (E)	
Charged amino acids					حوامض امينية مشحونة

أطراف جزيئة الحامض الأميني

أتضح مما سبق ان جزيئة الحامض الأميني غير متناظرة، والمهم في هذه الفقرة هي الأطراف مجموعة الأمين ومجموعة الكربوكسيل.

الطرف الأميني: وهو الطرف الذي ترتبط به مجموعة الأمين NH_2 - ويطلق عليه Amino terminus أو NH_2 -terminus أو N-terminal end أو Amine-terminus كل هذه المسميات توضح هذا الموقع والذي تبدأ منه كتابة تواليات الحوامض الامينية في البروتين، اي تشير الى بداية البروتين او الببتيد الذي تكون مجموعة الأمين فيه حرة، وعند تخليق البروتين وترجمة RNA الخاص به فهي تبدأ من النهاية الامينية لتنتهي عند الطرف الثاني وهو الطرف الكربوكسيلي.

يحتوي الطرف الأميني من البروتينات على إشارات الاستهداف Targeting signal وعادة يحتوي على الجزء من البروتين الذي يسمح له بالتوجه الى المكان المطلوب من الخلية، وإشارة الهدف هذه تفلق من البروتين بعد ان يصل الى مكانه المطلوب بإنزيمات حالة للببتيدات Peptidases.

وهناك بعض التحويلات التي تجري على الطرف الأميني بعد الترجمة منها إضافة جزيئة Myristoyl anchor ضمن عملية N-myristoylation والبروتينات التي تحور بهذه الطريقة تحوي على قطيقات متوافقة Consensus motifs عند الطرف الأميني كإشارة لإحداث التحوير، وكذلك إضافة مشتقات Palmitoyl group كمرسى. وهذه عادة تحدث للبروتينات التي وجهتها الأغشية الخلوية بدون احتوائها على دومين يستعرض الأغشية.

كما ان بعض البروتينات لا يمكن ان تطوى الا بوجود ببتيدي أولي يقع في الطرف الأميني من توالي البروتين. وفي اغلب الأحيان تبدأ عمليات طوي البروتين من النهاية الامينية، لذلك كانت هذه النهايات هدفا ملائما لوضع برامج التنبؤ بتراكيب البروتينات الثانوية .

الطرف الكربوكسيلي: وتطلق عليه تسميات أخرى - كما مر مع الطرف الأميني - لتشير الى نهاية البروتين او الببتيد الحاوي على مجموعة كربوكسيل حرة (COOH -) وبما ان السلسلة تبدأ من N الى C فان الطرف الكربوكسيلي يمثل النهاية.

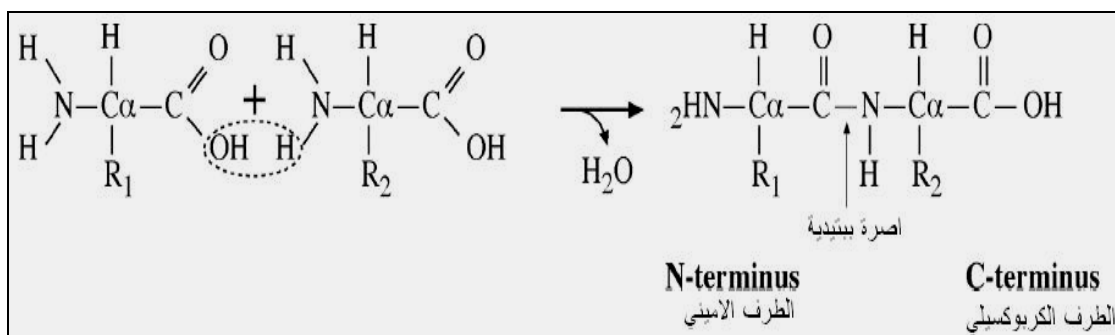
ويحتوي هذا الطرف على إشارات الاحتفاظ والاستبقاء Retention signal لغرض إفراز البروتينات، وأكثر الحوامض الامينية شيوعا التي تحتفظ نهاياتها الكربوكسيلية بالإشارات هي التوالي KDEL او HDEL وهذه تحتفظ بالبروتين في الشبكة الاندوبلازمية Endoplasmic reticulum retention signals وتمنع البروتين من الدخول الى المسار الإفرازي الا عند الضرورة.

ويتعرض الطرف الكربوكسيلي الى عدة تحويلات بعد عملية الترجمة وأكثر هذه التحويلات شيوعا هو إضافة الدهون التي تكون بمثابة جزيئة إرساء تساعد بعض البروتينات على اقتحام الأغشية الخلوية لترسو فيها دون احتوائها على دومين يمكنها من الدخول واستعراض الأغشية. ومن التحويلات الأخرى Prenylation التي تشمل Farnesylation وإضافة Geranylgeranyl - ، لتساعد في اقتحام الأغشية. وقد تكون بعض هذه التحويلات غير مباشرة على النهاية الكربوكسيلية ولكنها ان تمت فهي تكون على ثملات قريبة من الطرف الكربوكسيلي على وجه الخصوص. ومن البروتينات التي تحور على الطرف الكربوكسيلي

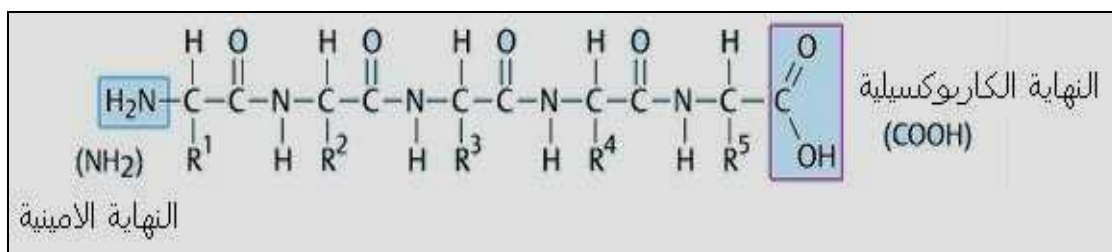
بروتينات G-proteins الصغيرة المرتبطة بالأغشية. ومن التحويرات الأخرى إضافة المراسي Glycosylphosphatidylinositol (GPI anchors) لتساعد البروتينات في الاقترام والرسو في الأغشية الخلوية. كما ان التحويرات التي تجري على هذا الطرف يمكن ان تحفز بعض الإنزيمات كما في إنزيم كثرة RNA (RNAP II).

الأواصر الببتيدية Peptide bonds

الأواصر التي تربط الحوامض الامينية وتسمى ايضا Amide bond وتكون بين ذرات الجزيئة الرئيسية لتكون سلاسل الببتيدات والبروتينات وتكوين هذه الأواصر سهلا نظرا لطبيعة الحوامض الامينية المتأينة وغير المتناظرة اذا انها تكون ثنائية القطب Dipolarized خاصة عندما تكون في محلول متعادل الحموضة (Zwitterion) اذ تكون مجموعة الأمين حاوية على البروتون والكربوكسيل منزوعة البروتون وهي صفة تعتمد على الرقم الهيدروجيني الذي توجد فيه الجزيئات وبذلك تمثل مجموعة الكربوكسيل مستقبلا جيدا للأواصر الهيدروجينية ما عدا في حالة البرولين فان مجموعة الأمين هي التي تكون معطية للهيدروجين ، وعند تكون الأصرة يرتبط الطرف الأميني من جزيئة مع الطرف الكربوكسيلي من جزيئة أخرى مع طرح جزيئة ماء (Dehydration reaction) ، كما موضح في التفاعل الآتي :

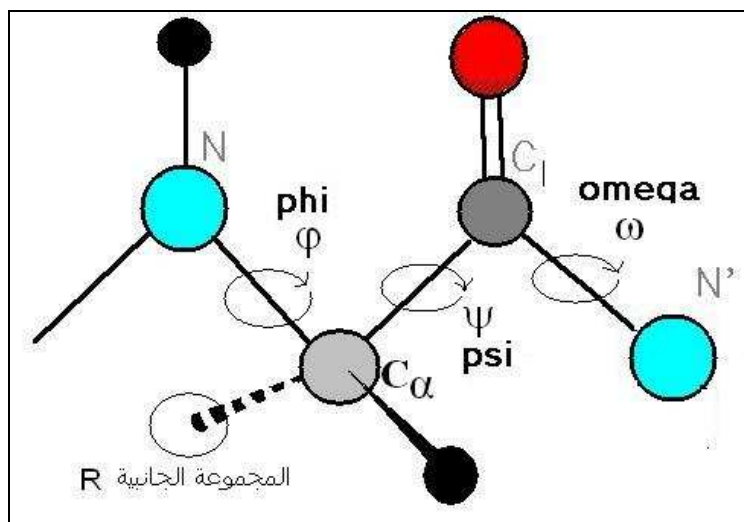


وتكون المسافة بين ذرتي الكربون ألفا للحامضين بحدود 380 بيكومتر، وبتكرار عملية الارتباط التي تتم على الرايبوزومات تتكون الببتيدات الموضحة في التركيب الآتي :



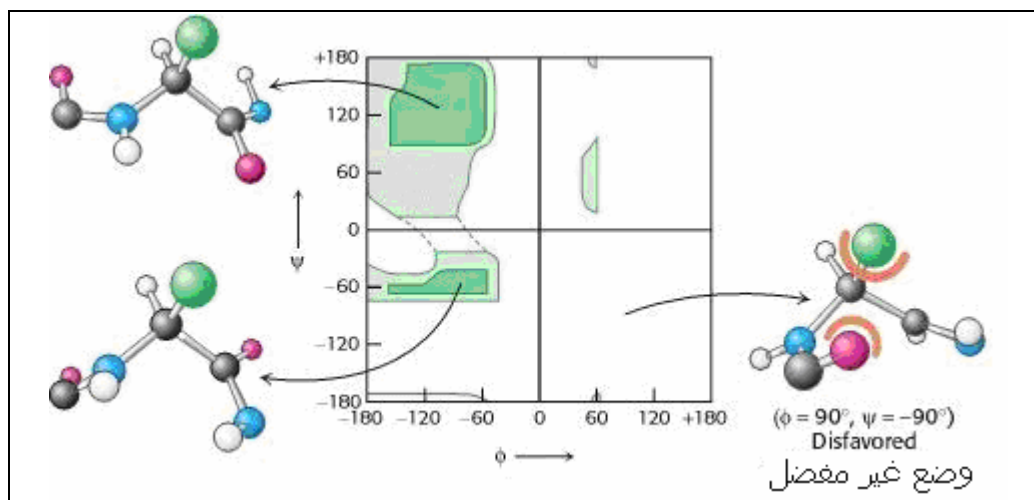
وتكوين الأصرة الببتيدية يحتاج الى إضافة طاقة حرة، وتكون هذه الأصرة ثابتة جدا. وعندما تتكون الأصرة التي تكون في مستوي تحتم على مكونات الحوامض الامينية توزيع يتحدد بالزاويا المزدوجة Dihedral angles وهي ϕ الواقعة بين النتروجين وكربون ألفا ، والزاوية

ψ الواقعة بين ذرة $C\alpha$ وكربون الكربوكسيل فهذه يمكن ان تختلف وتمثل درجات حرية للبروتين وهي التي تحدد الأبعاد الثلاثية للبروتينات فيما بعد. اذ انها تقييد بقياسات تسمح بتراكيب ثانوية محددة وليست مطلقة ومن ثم يترتب عليه تكوين التراكيب الثانوية وطوي البروتينات وهذه الزوايا موضحة في الشكل 70.3 :



الشكل 70.3 : زوايا ϕ و ψ في الآصرة الببتيدية

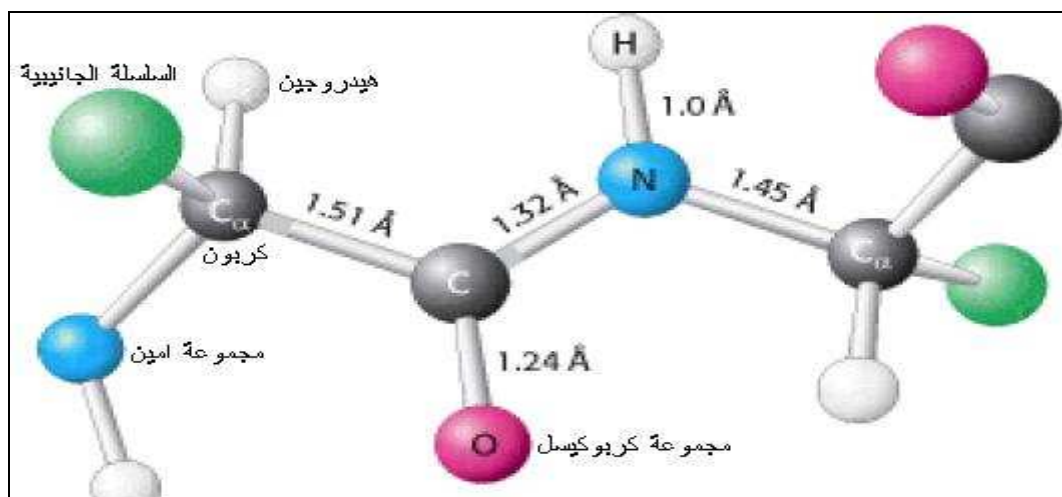
وتتراوح قيم الزوايا بين $+180^\circ$ الى -180° ، فالدوران باتجاه عقرب الساعة يعطي الزوايا قيم موجبة وبالعكس ولذلك تسمى هذه الزوايا في بعض الأحيان بزوايا الانفتال Torsion angles. ان توائم الزاويتين ليس مطلقا كما انه في بعض الأحيان لا توجد توافقية نتيجة لتصادم الذرات، وهذه العلاقات تحدد بمخطط راماجاندران Ramachandran plot الموضح في الشكل 71.3 :



شكل 71.3 : مخطط راماجاندران

والشكل يوضح قيم الزوايا ϕ و ψ التي يمكن ان توجد بين الذرات فيشير الشكل مثلاً الى انه عندما تكون ϕ قائمة و ψ قائمة بالاتجاه المعاكس (أي قيم سالبة) فهو وضع غير ملائم. ويشير الشكل ايضاً الى ان ثلاثة أرباع وحدة الدائرة (وحدة الدائرة بأربعها الأربعة في القياسات الرياضية) لا تكون ملائمة وان هناك قيم خاصة فقط هي التي تسمح بان يكون توزيع الذرات ملائماً كما موضح باللون الأخضر الغامق والأقل منها تفضيلاً بالأخضر الفاتح ، اذ ان القيم العامة تؤدي الى عمليات تصادم بين الذرات وتؤدي الى ابتعاد الذرات والتي تكون من الأساسيات المميزة لتوزيع الذرات المجسم وذلك لانه لا يمكن ان تكون الذرات في المكان نفسه وفي الوقت نفسه، وجزء من هذا التأثير تضيفه السلاسل الجانبية للحوامض الامينية من حيث صفات حجمها او مجاميعها الفعالة او غيرها من الصفات.

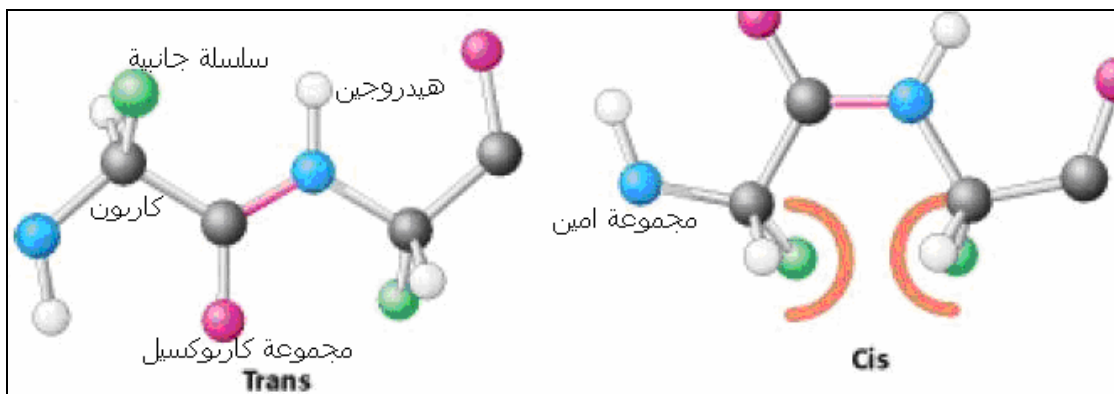
وبالعودة الى الحقائق المذكورة سابقاً وفحص هيكل البروتين تتضح عدة حقائق منها ان الأصرة الببتيدية هي مستوية بشكل أساس وعليه فعند ارتباط اثنين من الحوامض الامينية فان الست ذرات المكونة لهما ستكون في المستوى نفسه (افتراضاً) ، والحقيقة الأخرى ان الأصرة الببتيدية تمتلك بعض صفات الأواصر المزدوجة نتيجة اشتراك الكترونات بين الذرات $O=C-N$ والتي تمنع دوران الذرات حولها وهذا ما يؤدي الى استواء الأصرة اي تكوين مستوى الأصرة Peptide plane ، ونظراً للتركيب الصلب الذي تبديه الأصرة الببتيدية وثبوتيتها فهي تجبر الذرات المرتبطة بها على ان تقع في المستوى نفسه ولكن بتقييدات (كما ذكر آنفاً). وصفاً الأصرة المزدوجة التي تمتلكها يعبر عنها بطول الأصرة التي تربط بين الذرات فطول الأصرة بين $C-N$ هي 1.32 انكستروم في حين ان الأصرة المفردة (العادية) بين $C-N$ تبلغ 1.49 انكستروم ، اما في حالة وجود الأصرة المزدوجة بين العنصرين $C=N$ فتكون قيمتها 1.27 انكستروم والشكل التالي (72.3) يوضح أطوال الأواصر :



شكل 72.3 : أطوال الأواصر عند تكوين الأواصر الببتيدية

ان الأصرة الببتيدية غير مشحونة وبعد استقرار الذرات وفق قيم الزوايا الازدواجية يمكن ان تنشأ سلاسل من الحوامض الامينية التي يكون فيها توزيع الحوامض الامينية حول محور

السلسلة بالشكل المتخالف Trans وفيه تقع ذرتي $C\alpha$ على جانبي الأصرة الببتيدية، أو تكون بالوضع المتوافق Cis وفيه تقع ذرتي $C\alpha$ للحامضين المتجاورة على الجهة نفسها من مستوى الأصرة الببتيدية كما موضح في الشكل 73.3 :



شكل 73.3 : توزيع الحوامض الامينية على محور الأصرة الببتيدية

ويكون توزيع الحوامض الامينية في البروتينات من النوع المتخالف وذلك لتجنب التصادم الذي يحدث بين المجاميع المرتبطة بذرة الكربون الرئيسية ($C\alpha$) والذي يمكن ان يحصل من الوضع المتوافق بكثرة مقارنة بالوضع المتخالف، ما عدا ان الوضع المتوافق Cis يوجد في حالة ارتباط البرولين X-proline linkages ففي هذه الحالة يظهر تفضيل Cis لان نتروجين البرولين يرتبط الى ذرتي الكربون ويقلل من الأبعاد المجسمة.

القوى المساهمة في تركيب البروتين

يتضح مما سبق ان الخطوة الأولى في تكوين البروتينات هي تكوين الأواصر الببتيدية التي تربط الوحدات الأساسية (الحوامض الامينية) لتكوين المكونات المتمثلة بالببتيدات والبروتينات، وان تأثير الأواصر الببتيدية ومواصفاتها تضيف الكثير من التأثير في تركيب البروتينات، وفضلا عن ذلك فهناك بعض القوى الأخرى التي تساهم في تحديد التركيب الثانوي والثلاثي الجسم ونوعية تفاعلات البروتينات مع ما يحيطها من بيئة. ومن هذه القوى:

الأواصر الهيدروجينية

فالمعروف ان الببتيدات تحوي عدداً من معطيات البروتونات ومستقبلاتها سواء كان ذلك في السلسلة الرئيسية أو في السلاسل الجانبية (R) للحوامض الامينية. ومن جهة ثانية فان البيئة التي يوجد بها البروتين تحوي العديد من معطيات ومستقبلات الأواصر الهيدروجينية مثل الماء ولذلك فمن المتوقع ان تتكون الأواصر الهيدروجينية ليس فقط داخل السلسلة الببتيدية وانما تحدث مع البيئة المحيطة بها وهي البيئة المائية في معظم الأنظمة الحيوية.

وتمثل الأواصر الهيدروجينية نوعاً من التداخلات الكهربائية المستقرة وهي تشبه تداخلات المكونات ثنائية الأقطاب Dipole-dipole interactions والتي تشمل الهيدروجين من أحد الشمالات والأكسجين من ثمالة أخرى ويمكن ان تحصل بين ذرات السلسلة الرئيسية أو السلاسل الجانبية فالهيدروجين يكون في المجاميع المعطية مثل N-H ذات شحنة موجبة قليلة، في حين

يكون الأوكسجين هو المستقبل للأصرة الهيدروجينية والذي يكون مشحون بشحنة سالبة خفيفة مثل مجاميع $C=O$. ان نمط الارتباط بالأواصر الهيدروجينية يكون العامل السائد في تحديد الأنواع المختلفة من تراكيب البروتين الثانوية.

القوى الكارهة للماء

مثلما ذكر سابقا ان البروتينات تتكون من سلسلة من الحوامض الامينية التي قد تكون محبة للماء او كارهة له والتي تضيفها السلاسل الجانبية وهذه الصفات هي التي تعمل على تداخل السلاسل الجانبية مع البيئة المحيطة وهي التي تلعب الدور الأهم في إعطاء البروتينات الشكل الخاص بها، فمثلا في حالات طوي البروتينات الكروية يعكس التوازن بين الطاقات المتعكسة لتكوين الأواصر الهيدروجينية بين السلاسل الجانبية المحبة للماء والبيئة المائية المحيطة وبين التناظر بين البيئة المائية و السلاسل الجانبية الكارهة لها (اي للماء). ان كراهية الماء في السلاسل الجانبية تميل الى إبعادها من المناطق الخارجية للبروتين لتتصوي الى الداخل وبالتالي تكون هي القوى المؤثرة التي تحدد الأشكال المجسمة التي يمكن للبروتين ان ينطوي بها ويكون شكله المجسم العام.

القوى الكهربائية المستقرة Electrostatic forces

هذه القوى يمكن ان تكون على أشكال منها بين الشحنات Charge-Charge او تكون بين الشحنة ومجاميع ثنائية القطب Charge-dipole او تكون بين مجاميع ثنائية القطب مع بعضها Dipole-dipole. وقوى التداخل الكهربائي المستقرة تعد من القوى المهمة في تثبيت التركيب البروتيني ، وتحدث عند وجود زيادة من الشحنات السالبة في منطقة معينة من البروتين تتعادل مع شحنات موجبة من مناطق أخرى مؤدية الى تكوين جسور بين الثمالات ذات الشحنات المختلفة وهي يمكن ان تعمل بمدى طوله 15 انكستروم. وعندما تقترب بمسافة 3 انكستروم تتكون بينهم أصرة مثل الأواصر الهيدروجينية.

والتداخلات من النوع الأول وفي حالتها المثالية تشجع طوي البروتينات والتي تكون بين السلاسل الجانبية للحوامض الامينية مثل اللايسين (K) او الارجنين (R) وحامض الاسبارتيك (D) او حامض الكلوتامين. وفي الحالة الثانية يلاحظ صرف كميات من الطاقة لطوي البروتينات أثناء تداخل الشحنات مع جزيئة ثنائية القطب مثل الماء، وذلك لان العزم الثنائي القطب Dipole moment الذي يوجد في بعض المجاميع الجانبية القطبية للحوامض الامينية يؤثر في تداخلها مع الماء. وعليه يلاحظ ان الأغلبية العظمى من الحوامض الامينية الموجودة على سطوح البروتينات (ومنها الكروية) تكون حاوية على مجاميع مشحونة او قطبية.

قوى فاندرفال van der Waals forces

هناك قوى فاندرفال جاذبة وأخرى متافرة تسيطر في طبيعة البروتين الناتج بعد طويه ، فالجاذبة هي التي تحدث بين الذرات غير المشحونة وغير المرتبطة المتجاورة. والمتافرة هي التي تعمل في التداخلات التي تحدث عندما تصبح الذرات غير المشحونة وغير المترابطة (المذكورة توأ) قريبة جدا من بعضها ولكن لا تحت تكوين تركيب ثنائي القطب Dipole وتكون قوى فاندرفال (التي تساهم في ثبوت البروتين الكلي) تداخلات آنية بين الذرات عندما تصبح بمثابة مركبات ثنائية القطب عابرة والتي يمكن ان تنعكس وهذا التذبذب بين حالة ثنائية القطب يؤدي الى حصول قوى جاذبة. كما ان التناظر يكون ناتجا من تناظر الالكترونات التي

تحدث عندما تكون هناك سحابتان من الإلكترونات تبدأ بالتداخل. وبالرغم من أن قوى فاندرفال ضعيفة جدا مقارنة بالقوى الأخرى التي تتحكم في بنية البروتين إلا أن كثرتها الحاصلة في البروتينات الكبيرة تجعل منها قوى مهمة مؤثرة في طوي البروتينات.

قوى أخرى

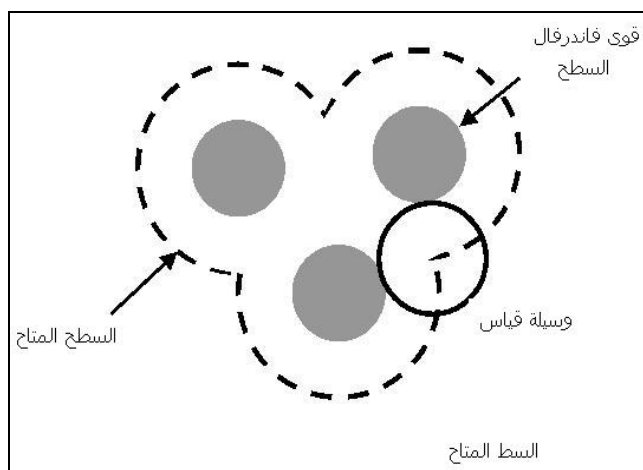
أن طوي البروتينات يمكن أن يشمل مساهمة أواصر أخرى منها الأواصر التساهمية مثل تكوين الأواصر الكبريتيدية المزدوجة $S=S$ التي تتكون بين ثملات السستين لتكوين العناقيد Clusters الخاصة بالمعادن العاملة كـ Prosthetic groups كما أن لقيم ψ و ϕ في جزيئة الحامض الأميني تأثير في فهم طوي البروتينات.

صفات البروتينات الأخرى

أن توالي الحوامض الأمينية ونوعيتها وارتباطاتها يعد الأساس في تحديد صفات البروتينات وهو الذي يضيف عددا من الصفات المهمة منها:

المساحة السطحية للبروتينات

تمثل المساحة السطحية المتاحة Accessible surface area والموضحة في الشكل 74.3 :

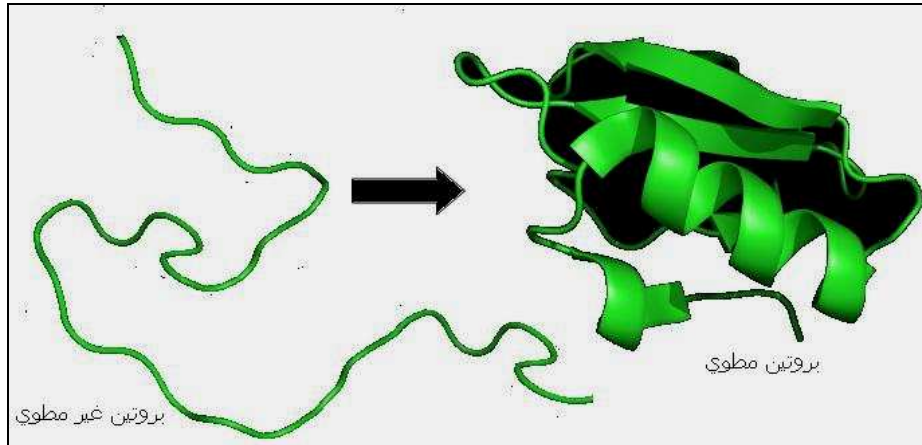


شكل 74.3 : المساحة السطحية المتاحة للبروتينات

المساحة السطحية للبروتين التي يمكن للمذيب الوصول إليها. وتحسب بالانكستروم المربع، وصفت عام 1971 من قبل Lee - Richards لذلك تسمى بعض الأحيان بـ Lee - Richards molecular surface وهناك عدة طرق لقياسها وبرامج خاصة بها. وتستعمل لقياس انتقال الطاقة الحرة اللازمة لتحريك الجزيئة الحيوية من المحيط المائي (المذيب) إلى مذيب غير قطبي مثل بيئة دهنية. والمساحة السطحية وبمشاركة بعض القوى مثل قوى فاندرفال وغيرها تحدد تركيب البروتين كما أنها تكون مهمة في التنبؤ لتحديد المواقع الفعالة والحواتم Epitopes للأجسام المضادة وكذلك التحويلات الحاصلة بعد الترجمة. وهناك عدداً من المؤشرات التي تعتمد عليها المساحة ومن أهمها حجم الكرية للجزيئة وتستعمل بكثرة في تصميم البرامج الخاصة بالتنبؤ بالتراكيب الثانوية للبروتينات.

تحويلات بعد الترجمة

تعد التحويلات بعد الترجمة Post-translational modifications من الأساسيات لإعطاء البروتين أو الببتيد المتعدد الصفة الوظيفية وتعد من الجوانب المهمة في دراسة المكنون البروتيني . فالعديد من البروتينات تحتاج الى عمليات ربط تساهمي لتصيح قادرة على أداء وظائفها وقد يكون ذلك قبل او بعد عمليات الطوي. ويكون هذا أكثر وضوحا في الخلايا حقيقية النواة التي تحصل فيها أكثر التحويلات في الشبكة الاندوبلازمية وأجسام كولجي Golgi apparatus ، ومن التحويلات فلق بعض أجزاء البروتين ، وتكوين الجسور الكبريتيدية الثنائية وإضافة بعض المجاميع مثل السكريات او الفوسفات او المثل او الاستيل وغيرها من المجاميع على ثملالات خاصة في البروتين، او ربط الببتيد المتعدد بمكوثرات صغيرة من الكربوهيدرات او الدهون، وإضافة Prosthetic groups لبروتينات الإنزيمات لإنتاج البروتين الناضج. والتحويلات المذكورة تؤدي الى تغير في صفات البروتينات منها تغير الحجم او تغير مواقع الكراهية للماء ، او تغير الهيئة الكلية للبروتين. وكل هذه تؤثر مباشرة في تداخلات البروتينات مع بعضها وكذلك في توجيه البروتينات الى المواقع التي تعمل بها ، و عملية الطوي وان كانت هي التوجه الصحيح في الأغلب الا ان فك الطوي يكون ضروريا، فمثلا يكون فك الطوي أنيا ضروريا" لعبور الأغشية الخلوية والانطلاق الى البيئة المحيطة، كما ان عملية فك الطوي يعد مرحلة أساسية لتحليل البروتينات التي أنهت مهمتها لتذهب الى الجسميات البروتينية Proteasome (مقبرة الخلية) لغرض تفكيكها، كما ان بعض البروتينات تحتاج الى عمليات طوي وفك سريع أثناء أداء مهماتها كما في بروتينات العضلات أثناء التمدد والتقلص والشكل التالي(75.3) يوضح الفرق بين بروتين مطوي وآخر غير مطوي



شكل 75.3 : الفرق بين بروتين قبل الطوي وبعده

طوي البروتينات

عملية طوي البروتينات عملية فيزيائية وحيوية تتم ليصبح الببتيد المتعدد ذو شكلا "مميزا" يلاءم وظيفته. وهناك العديد من البروتينات التي تشترك في نمط الطوي ولكن ليس من الضروري ان تشترك في توالي الحوامض الامينية. ووجود البروتينات مطوية بشكل صحيح مكونة تراكيب

ثلاثية الأبعاد يعد أساسا لوظيفة العديد من البروتينات وهي الحالة الطبيعية الغالبة، ولذلك فإن أي تغيير في طويها يؤثر في وظيفتها، فالبروتينات داخل البيئة الداخلية للخلايا والتي هي بيئة مائية بطبيعة الحال وعند حدوث أي تغيير عن البيئة المتعودة عليها سيؤدي إلى تغيير شكلها وذلك بإضعاف القوى التي تربط السلاسل.

وهناك عدة قوى تؤثر في عملية الطوي وهذه قد تكون فيزيائية فالملاحظ أن الأصرة بين ذرة الكربون الرئيسية في الحوامض الأمينية كربون ألفا ومجموعة الأمين وكذلك بين ذرة كربون ألفا ومجموعة الكربوكسيل هي أواصر مفردة لذلك فإن أي اثنين من وحدات الببتيد (الحوامض الأمينية) المتجاورة يمكن أن تدور حول هذه الأواصر متخذة ترتيبات مختلفة ومثل هذه الحركات تساعد في طوي البروتين بطرق مختلفة . ومن القوى الأخرى هي تكوين الأواصر الهيدروجينية الداخلية في الببتيد مما يوفر المزيد من الثبوت ، وهذه الأواصر تعتمد على البيئة التي يوجد فيها البروتين أو الببتيد، فالأواصر الهيدروجينية الداخلية أو المغلفة داخل مركز كاره للماء لها مشاركة أكبر في ثبوت البروتين مقارنة بالأواصر المعرضة للبيئة المائية الخارجية. وتشارك الثمالات القطبية وغير القطبية في طوي البروتينات. ويلاحظ في العموم انطمار الثمالات الكارهة للماء إلى الداخل لتكون بعيدة محمية من المذيبات المحيطة وفي العموم تكون البروتينات التي تأخذ صفة بروتينات مطوية يكون لها لب أو مركز كاره للماء والذي يتم تثبيته بالسلاسل الجانبية للحوامض الأمينية المشاركة، أما سطوحها الخارجية والمعرضة للمذيبات ففي العادة تحوي على الثمالات المشحونة وسلاسلها القطبية هي التي تتداخل مع الماء المحيط، ولكن هذا ليس كذلك مع البروتينات المكونة من وحدة واحدة .

تتأثر عملية الطوي أيضا بالمساحة السطحية المتاحة والتي تعد من أهم العوامل المشاركة إذ تحدد مدى تعرض البروتين للمذيبات المحيطة وفيما إذا كانت محبة أو كارهة للماء وهذه بدورها تتأثر بوجود الأملاح والايونات ودرجة الحرارة.

ومن جهة ثانية فإن عمليات الطوي تتأثر بالطاقة والثرموداينمك ، فمقارنة مكوثر مثل ببتيد غير مطوي والذي يكون حلزون عشوائي Random coil ويمكن أن يوجد على أي شكل من الأشكال ، فإن الانتروبي Entropy (مقياس لعشوائية المادة) يعاكس حالة الطوي ويجب أن يعوض بتدخلات تفضل الشكل المطوي، لذلك فالمكوثرات ذات المرونة العالية والتي يكون لها عدد كبير من الهياكل لا تنطوي بتركيب واحد محدد.

ويحصل بعض الأحيان أن يضطرب الطوي أي تحصل عملية مسخ نتيجة لعدة ظروف منها ارتفاع درجة الحرارة أو التجميد وكذلك يؤدي تطرف الأرقام الهيدروجينية إلى اضطراب طوي البروتينات وبعض الأحيان يكون المسخ قابل للرجوع ولكن في أحيان كثيرة يكون غير قابل للرجوع ودائمي. وفي الخلايا الحية يكون المسخ مؤذيا، والاستجابة للكروب أو الإجهاد تكون مسئولة عن إصلاح هذه الأضرار.

مسارات عمليات طوي البروتينات

عمليات طوي البروتينات لا يمكن ان تقتصر على تدخل العوامل الفيزيائية وعلاقتها بالطاقة فقط وانما هناك بروتينات الوصيفات Chaperones التي تشارك بشكل فاعل في الطوي. وفي اغلب الأحيان تكون عملية الطوي مرافقة لعملية الترجمة خاصة في البروتينات الكبيرة، وتخليق البروتينات على الرايبوزوم يبدأ بالطرف الأميني، والبروتينات الصغيرة تطوى بعد ان يترك الطرف الكربوكسيلي الرايبوزوم ، اما في البروتينات الكبيرة وخاصة المكونة من عدد من الدومينات فيبدأ طويها ولا يزال الطرف الكربوكسيلي عالقا بالرايبوزوم كما ذكر آنفا .

وعمليات الطوي تكون سريعة وتأخذ مقياس ملي ثانية، فمثلا في البكتريا فان البروتينات الصغيرة تطوي بحدود 10-50 ملي ثانية وان كان هذا محددا بمدى تحلل ATP وتوفر الطاقة اي انه في هذه الحالة يكون ATP بمثابة المؤقت في حين ان البروتينات الكبيرة (في البكتريا) تدخل دورة الوصيف لطويها .

والملاحظ ان مسارات طوي البروتينات الكبيرة (في العموم) معقدة ولا تزال بعض جوانبها غامضا، والملاحظ ايضا ان الطوي بمساعدة الوصيفات يكون ضروريا" جدا في بيئة داخل الخلايا التي تكون مزدحمة جدا بمختلف الجزيئات والمركبات ولا مجال لعمليات الطوي الذاتي وبذلك تمنع الوصيفات التكتل Aggregation ، فضلا عن ان الطوي بمساعدة الوصيفات يمنع عمليات الطوي الخاطئ الذي قد ينشأ من التعرض للحرارة او غيرها من الظروف داخل الخلية، وتساعد فيه بعض الإنزيمات مثل PDI Protein disulfide isomerase وإنزيم PPI peptidyl-prolyl cis-trans isomerase التي تساعد في تفاعلات المزامرة او المناظرة وتكون ضرورية في طوي عدد من البروتينات والملاحظ ان البروتينات التي تعمل فيها إنزيمات المزامرة Isomerization تأخذ وقت طويل لطويها قد يصل الى ساعات خاصة اذا كانت تحت ظروف خارج الأنظمة الحية (In vitro).

أنواع مسارات الطوي

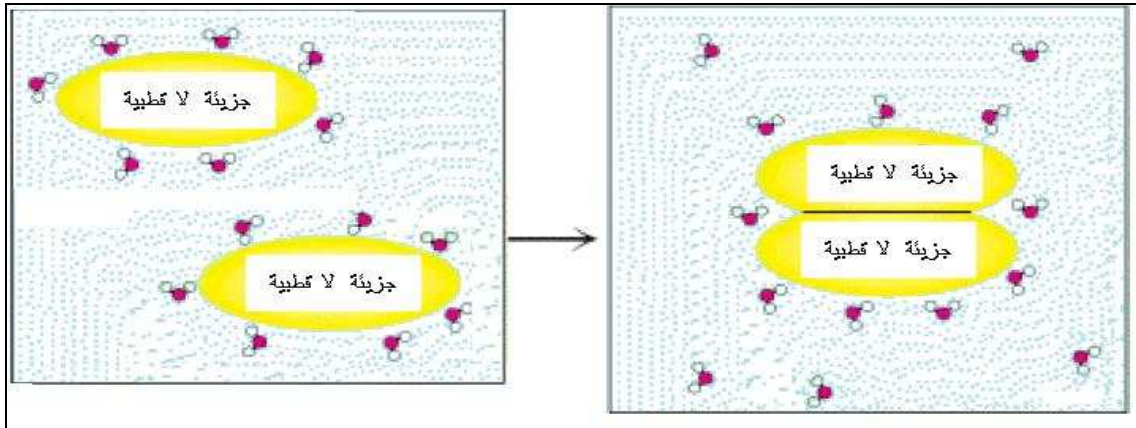
يوجد أكثر من مسار لطوي البروتينات ولكن تبقى حقيقة واضحة تتمثل بان توالي الحوامض الامينية في البروتين هي التي تحمل المعلومات الخاصة بطويه وتكوين الهيئة الطبيعية له وكذلك هي التي تحدد المسار الذي تسير فيه التواليات لتصل الى بروتينات بهيئة مفضلة ومريحة من ناحية الطاقة ولذلك لا بد ان تمر هذه في حالة وسطية هي Molten globule state ، ومن جهة ثانية فان طوي البروتينات يعتمد على البيئة المحيطة حيث وجد ان البروتينات المتشابهة تطوى بشكل مختلف (وليس اختلاف أساسي) اعتمادا على البيئة التي توجد فيها. وقد وضعت عدة موديلات لوصف عملية الطوي منها:

* **العملية الهرمية:** وتشير هذه الى انه في البداية يتكون التركيب الثانوي من حلزونات ألفا α -helixes وصفائح بيتا β -sheets (التي سيأتي ذكرها لاحقا) وبالتداخلات الواسعة بين هذه التراكيب تتكون التراكيب الثنائية ألفائقة وتستمر الأخيرة بالتداخل الى ان يطوى كل البروتين.

* **الطوي التلقائي او المسند:** مسار يعتمد على الانتروبي وتشارك فيها ببتيديات مفردة وتكون حالة الطوي بمثابة مصب للطاقة ألفائضة، فالطاقة الحرة العالية تعد مقياسا لعدم الثبات والتي تكون عالية في البروتينات غير المطوية، وبذلك تكون عملية طوي البروتينات وتكوين عدد قليل من البنى او الهيئات التي يكون فيها البروتين ملائمة لهذه الحالة. وفي هذا المسار يحصل

تقلص للبيبتيدات المتعددة لتكون بحالة مضغوطة Compact state وهي التي تسمى بـ Molten globule.

ويمكن توضيح ما ذكر أعلاه بان وجود البروتينات بحالة غير مطوية يؤدي الى اضطراب النظام اذ ان كل جزيئة تتخذ هيئة خاصة بها وبالتالي يكون مجموع الانتروبي لمجموعة الجزيئات عالياً وسالب القيمة وهذا يقل عند طوي البروتينات الى أشكال منتظمة ويكون لجزيئات الماء دور في عمليات الطوي من خلال الأواصر الهيدروجينية، فضلا عن وجود قوى أخرى مثل القوى الأيونية ، وفي هذه الحالة فان الجزيئات اللاقطبية لا يمكن ان تشارك في التداخلات الأيونية او الأواصر الهيدروجينية، ولذلك كان تداخل الجزيئات اللاقطبية مع الماء اقل تقصيلا من تداخل جزيئات الماء مع بعضها ولكن مع هذا فان جزيئات الماء الملامسة للجزيئات اللاقطبية تكون حولها ما يشبه القفص وتصبح مرتبة أكثر وأكثر من باقي جزيئات الماء الحرة وبالتالي تقلل من الانتروبي. وعند اقتراب جزئيتين لا قطبية من بعضها تنطلق جزيئات الماء التي تذهب لتتداخل بشكل حر مع باقي جزيئات الماء كما موضح في الشكل : 76.3



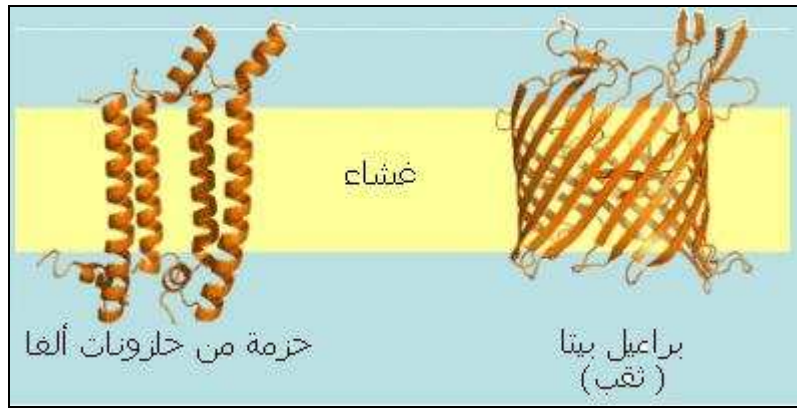
شكل 76.3 : تداخل الجزيئات اللاقطبية

وتميل الجزيئات اللاقطبية للتجمع داخل الماء لان الانتروبي للماء سيزداد بإطلاق جزيئات الماء وهذا ما يسمى بالتأثير الكاره للماء والذي يساعد في العديد من العمليات الحيوية. وداخل مجمع الجزيئات الكارهة للماء ترتبط الجزيئات مع بعضها لتزيد من ثبوت البروتينات الناتجة ويكون ذلك بمساعدة بعض الأواصر الضعيفة وقوى فاندرفال. ويمكن تلخيص الحالة بان عملية الطوي تحدث عندما يكون الحاصل الكلي للطاقة الحرة سالبا ويكون ذلك بالمجمع بين الانتروبي المتعلق بالتأثير الكاره للماء و Enthalpy (المحتوى الحراري للمادة) المرتبط بتكوين الأواصر الهيدروجينية وقوى فاندرفال التي تؤدي الى جعل قيم الطاقة الحرة سالبة. ويمكن ان تكون عملية الطوي شاملة للحالتين أعلاه.

طوي البروتينات الغشائية

مما تقدم اتضح ان البروتينات عموماً تطوى بحيث تكون الأجزاء الداخلية هي الكارهة للماء والأجزاء الخارجية هي المحبة للماء، ولكن هذه الحقيقة تنعكس بالنسبة للبروتينات الغشائية. ولتوضيح طوي البروتينات الغشائية لابد من تقسيمها :

البروتينات الغشائية الأصلية: وهي البروتينات التي يشفر لها بالجينوم الأصلي الطبيعي للخلايا وتكون مسئولة عن الفعاليات الفسلجية الأساسية. وهذه يتم تجميعها بعمليات معقدة ولكنها تشترك في انها تخلق برايبوزومات خاصة متصلة بالأغشية بواسطة بروتينات خاصة تسمى Translocon تقع في داخل الغشاء الخلوي، وهو يوفر نفق خاص لحقن البروتينات المخلقة حديثاً على الرايبوزومات الى داخل الغشاء. وبعد اكتمال تخليق هذه البروتينات تنفك الرايبوزومات من Translocon، وينطلق البروتين الى داخل الغشاء والذي عندها يطوى الى التركيب الثلاثي الأبعاد المعد لوظيفته. والشكل التالي (77.3) يوضح هذه الحالة.



شكل 77.3 : بروتينات غشائية مطوية

البروتينات الغشائية غير الأصلية: بروتينات تمثل بشكل عام السموم والبيبتيدات المضادة للميكروبات والتي تحشر في الأغشية المستهدفة دون المرور بمسار البروتينات الأصلية المذكور أعلاه. وفي العموم تكون عمليات الإقحام هي عمليات فيزيوكيميائية.

ان أهم الأسباب وراء ثبوت وتركيب البروتينات الغشائية هي الحرارية الديناميكية وذلك لان الكلفة الحرارية لنقل المركبات المشحونة او القطبية الى الطبقة الهيدروكربونية الثنائية المكونة للغشاء تكون عالية جداً وهذا يترتب عليه :

- ان معظم السلاسل الجانبية للحوامض الامينية للقطع التي تستعرض الأغشية يجب ان تكون غير قطبية مثل ألفالين، النين، الليوسين، الايزوليوسين او الفنيل-النين.
- اما المجاميع القطبية جداً في العمود الفقري للبيبتيدات المتعددة المستعرضة للأغشية فيجب ان تساهم في الأواصر الهيدروجينية لغرض تقليل كلفة الطاقة لنقلها الى الطبقة الداخلية الهيدروكربونية وهذه الأواصر الهيدروجينية يمكن ان توفرها أشرطة بيتا التي تكون بشكل تراكيب مغلقة β -barrel.

وبعد استقرار البروتين في الأغشية وبغض النظر عن الطريقة التي وصل بها، فإنه سيكون في أدنى الحالات من مستويات الطاقة الحرة ولكنها لا تكون متساوية على جميع الأصعدة فهناك فرق في الطاقة في حالة تداخل البروتين مع بروتين آخر وبين البروتين والطبقة الدهنية وهذه هي التي تحدد المستويات الدنيا من الطاقة وبالتالي تحدد تركيب البروتين داخل الغشاء ثنائي الطبقة وذلك لأن البروتين يكون غير ذائب في الطبقة الدهنية وعندما يكون غير مطويا" فإنه يؤدي إلى الحاجة إلى كلفة عالية من الطاقة عند تعرض الأصرة الببتيدية إلى الطبقة الهيدروكربونية، ومن جهة ثانية فإن هذه البروتينات تكون غير ذائبة في الماء نظرا لطبقته (الخارجية) الكارهة للماء لذا كان لابد من طويها بشكل ملائم للبيئة التي توجد فيها. ولذا كانت عمليات التنبؤ لتحديد تركيب البروتينات الغشائية ليست بالعملية السهلة وإنما تحتاج إلى تفاصيل أكثر من التي ذكرت وخاصة المتعلقة بحالة التداخل مع الطبقة الدهنية الكارهة للماء.

الطوي الخاطئ والمسوخ للبروتينات

في بعض الحالات لا يتم طوي البروتينات بشكل صحيح وهذا ينتج من عدة أسباب ، لعل في مقدمتها تغيرات الطاقة. فتغيرات الطاقة المرافقة لعمليات طوي البروتينات تكون واطئة وتتراوح بين 1-15 كيلوسعرة ١ مول وهي مكافئة لعدد قليل من الأواصر الهيدروجينية، وأي تغير فيها يمكن أن يؤدي إلى مسخ البروتينات Denaturation. والفشل في طوي البروتين إلى الشكل المطلوب يرافقه إنتاج بروتينات غير فعالة وربما كانت سامة أو مؤذية.

ومن الملاحظ أن عملية المسخ لا تكون موازية لعملية فك طوي البروتينات بشكل كامل والعشوائية في البنية التي يوجد فيها البروتين فبعض البروتينات تكون فعالة رغم وجود حالة طوي جزئي فيها. وتميل البروتينات الممسوخة إلى تكوين عدة تداخلات مع الماء للتعويض عن التداخلات الموجودة في الحالة الطبيعية، وتكون بمثابة مجمع متباين من التراكيب التي هي في حالة تحولات مستمرة والبعض منها يكون سريعا.

ومن أهم الأسباب المؤثرة في طوي البروتينات هو انحراف درجة الحرارة عن الحد الطبيعي سواء كان ارتفاع أو انخفاض عن المدى الذي تعيش فيه الخلايا ويؤدي إلى إنتاج بروتينات غير ثابتة بالحرارة. ولذلك ففي البكتريا المحبة للحرارة والتي تنمو بدرجات حرارة أعلى من 100°م فهذه تحتاج أن تكون بروتيناتها والمجمعات البروتينية فيها ثابتة بدرجة حرارة أعلى من حرارة النمو ومثل هذه المجاميع تحمي بروتيناتها بواسطة مجموعة الوصيفات التي يطلق عليها بروتينات الصدمة الحرارية Heat shock proteins التي تساعد البروتينات أن تبقى مطوية وكذلك تساعد في طويها عند انفراط الأخير وهناك بعض البروتينات التي لا يمكن أن تطوى حتى في الحالات الطبيعية إلا بمساعدة مثل هذه الوصيفات.

ومن الأسباب الأخرى التي تؤدي إلى فك طوي البروتينات وبالتالي مسخها وجود التراكيز الملحية العالية وكذلك الأرقام الهيدروجينية المتطرفة والقوى الميكانيكية ووجود الماسخات الكيميائية.

وتوجد مجموعة من البروتينات المتخصصة لإصلاح الاختلال في عمليات الطوي وهي الوصيفات التي تصنف وفق أوزانها الجزيئية ومنها ما يسمى Chaperonin وتسمى بعض الأحيان بروتينات الصدمة الحرارية نظرا لاكتشافها أول مرة تحت ظروف الإجهاد الحراري ، هذه المجموعة تميز البروتينات الممسوخة وترتبط إليها بتداخلات بروتين – بروتين ويمكن أن

تعكس حالة المسخ ، وعندما يكون الضرر اكبر من قابلية الإصلاح تقوم بتفكيكها. لذا فالجينات المسؤولة عن تخليق الوصيفات يتم تنظيمها بظروف الإجهاد وغالبا في خلايا بدائية النواة يكون بتأثير العوامل المسؤولة عن الإجهاد Stress-response factors.

وعلى العموم فان الوصيفات لا تكون الحاجة لها محصورة بحالات الإجهاد وتصحيح البروتينات المدمرة، فهناك كميات منها تنتج في الخلايا التي هي ليست تحت الإجهاد ، فمنها ما يرتبط بالبروتينات النانئة من الرايبوزومات لتساعد في طيها بالشكل الصحيح، لذا تكون الوصيفات مهمة للبروتينات وهي في حالتها الطبيعية، فضلا عن اشتراكها في نقل البروتينات الجديدة الى مكانها الخاصة في الخلية ، واضطراب الوصيفات ووظائفها يؤدي الى العديد من الأمراض .

بعض الأمراض المتعلقة بطوي البروتينات

هناك عدة أمراض لها علاقة بحالة عدم الطوي الصحيح والذي ينتج عنه التكتل وعدم ثبوت البروتينات، ومثل هذه تحدث نتيجة الطفرات او تغير البيئة المحيطة بالبروتينات، ولذلك فان فهم آلية طوي البروتينات يمكن ان يؤدي الى تصميم الأدوية لتصحيح هذه الأمراض .

ومثل هذه الأمراض يطلق عليها الأمراض البروتينية Proteopathy diseases ويكون وراء فقدان الوظيفة هو اضطراب الطوي ومن هنا جاءت فكرة العلاج باستبدال البروتينات Protein replacement therapy كما في استعمال الوصيفات الصيدلانية Pharmaceutical chaperones لغرض إعادة طوي البروتينات المطفرة (على وجه الخصوص) وإعادتها الى وظائفها الطبيعية.

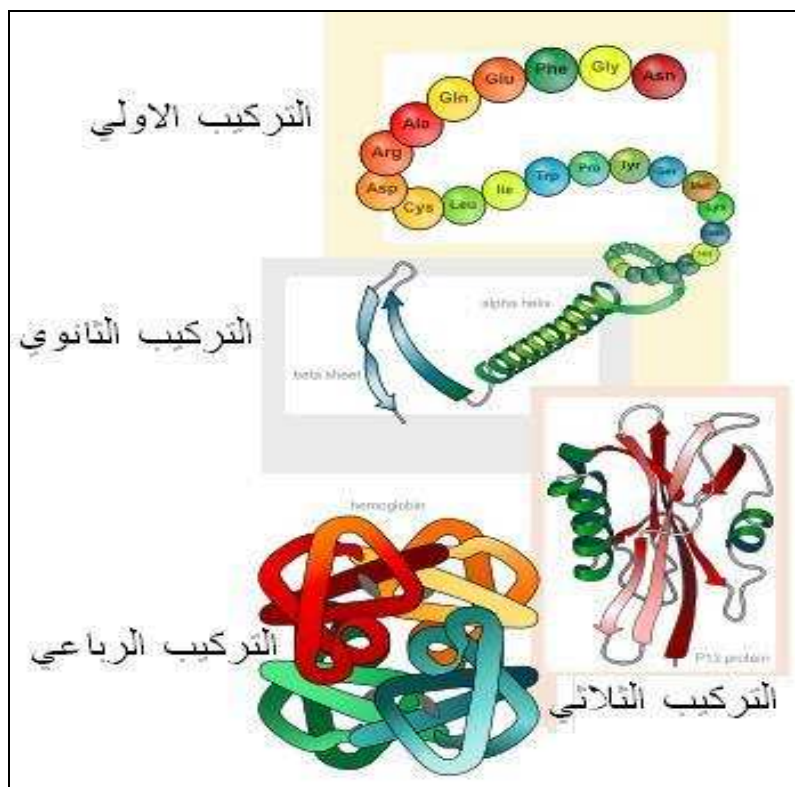
بعض البروتينات عندما يضطرب طويها تؤدي الى تكوين تكتلات غير ذائبة وممرضة Aggregated protein disease مثل البريون Prion وهو بروتين يوجد في الدماغ بالحالة الطبيعية PrPC يتحول الى الحالة غير الطبيعية PrPSC والآخر يتكاثر مؤديا الى تكوين تجمعات ويسبب ظهور الأمراض مثل جنون البقر وغيره من الأمراض التي تصيب الدماغ . ومن الأمراض الأخرى ذات العلاقة بتكتلات البروتين داخل الخلية مرض الباركنسون Parkinson ومرض Huntington والأمراض المذكورة تتعلق غالبا بتقدم العمر. وكذلك مجموعة الأمراض Amyloid-related Alzheimer's disease ، وأمراض أخرى لها علاقة بالتوارث العائلي، ومنها يتضح ان طوي البروتينات الخاطئ له علاقة واضحة بالأمراض العصبية الانحلالية Neurodegenerative diseases.

ولكن طوي البروتين الخاطئ لا يتعلق فقط بالأنسجة والأعضاء العصبية فهو يمكن ان يشارك في بعض أنواع الحساسية. فضلا عن مشاركته المهمة في مرض التحوصل الليفي Cystic fibrosis الذي ينتج عن خطأ في طوي أحد البروتينات الغشائية المسمى Cystic fibrosis transmembrane conductance regulator (CFTR) والبروتين يعمل كقناة لايونات الكلوريد وعطبه ينتج عن طفرة تؤدي الى حذف الحامض الأميني الفنيل – النين في الموقع 508 في CFTR وتسبب عدم انتظام طويه .

تراكيب البروتين ومستوياتها

التركيب البروتينية يمكن ان ترتب هرمياً بأربع مستويات اعتماداً على زيادة التعقيد وهي:

التركيب الأولي، التركيب الثانوي، والتركيب الثلاثي المجسم، والتركيب الرباعي. والمستويات موضحة في الشكل 78.3 :



شكل 78.3 : مستويات تركيب البروتينات

وفي العموم تترتب البروتينات على شكلين أو صنفين عامة من التراكيب هي: البروتينات الكروية Globular proteins والليفية Fibrous proteins. فالكروية هي مطوية وملفوفة بشكل منضغط والتراكيب الليفية تكون على شكل خيوط أو أشكال متطاوله، وذلك لوجود مناطق معينة في الببتيد تحتم الوضع المجسم المحدد الذي يتأثر بتوالي الحوامض الامينية في التركيب الأولي.

وتتم في العادة دراسة مستويات تراكيب البروتينات باستعمال الأشعة السينية ومطياف الرنين النووي المغناطيسي وقياس التداخل القطبي المزدوج Dual polarization interferometry. وتساهم عدد من القوى في تثبيت التراكيب ابتداءً من التركيب الثانوي الى التركيب الرباعي، وعمليات التثبيت تتم بقوى لا تساهمية وتتمثل بالأواصر الهيدروجينية وتداخلات الكهربية المستقرة وقوى فاندرفال السالفة الذكر.

التركيب الأولي Primary Structure

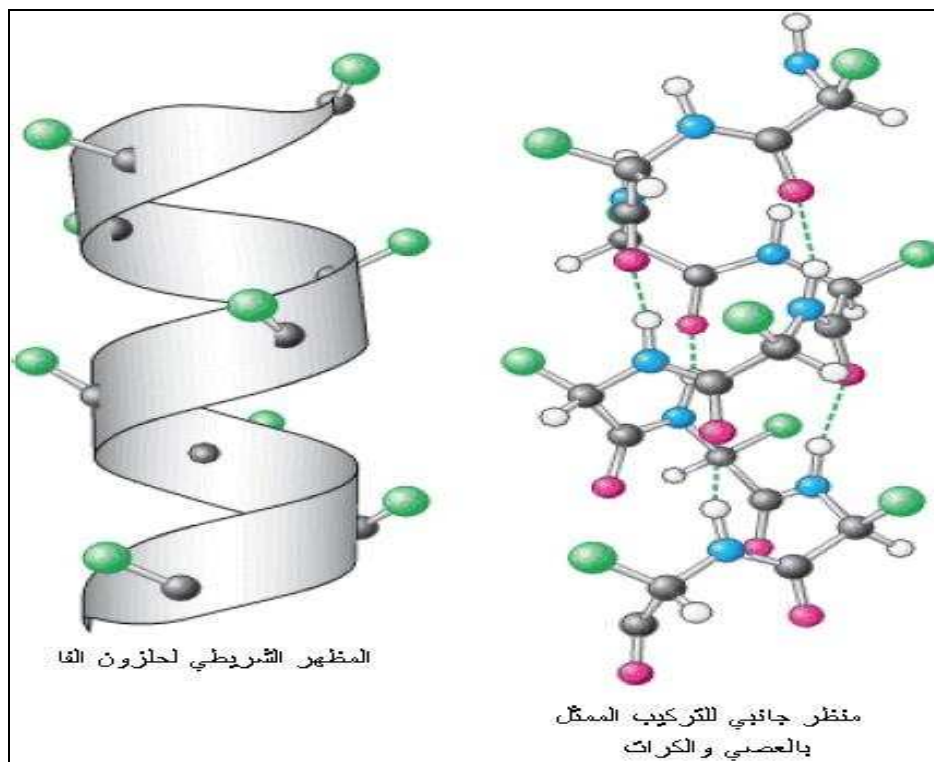
ويعني به توالي الحوامض الامينية باتجاه واحد، وتربط الحوامض الامينية مع بعضها بالآصرة الببتيدية ذات الطبيعة التساهمية على الرايبوزوم أثناء عملية الترجمة اي أثناء تخليق البروتين، ويعد مستوى التركيب هذا أبسط المستويات الموجودة في البروتينات.

التركيب الثانوي secondary Structure

ويعرف ايضا بالهيئة الموقعية Local conformation للآصرة الببتيدية ويتصف بدرجة عالية من التنظيم والترتيب والتكرار. ويتمثل التركيب الثانوي بحلزونات ألفا α -helixes وصفائح بيتا β -sheets وقد اقترحت في بداية الخمسينات من القرن الماضي ثم تلاها تحديد تراكيب أخرى مثل لفات بيتا β -turns وعروة اوميكا Ω -loop لتعمل في إعطاء البروتين التركيب النهائي.

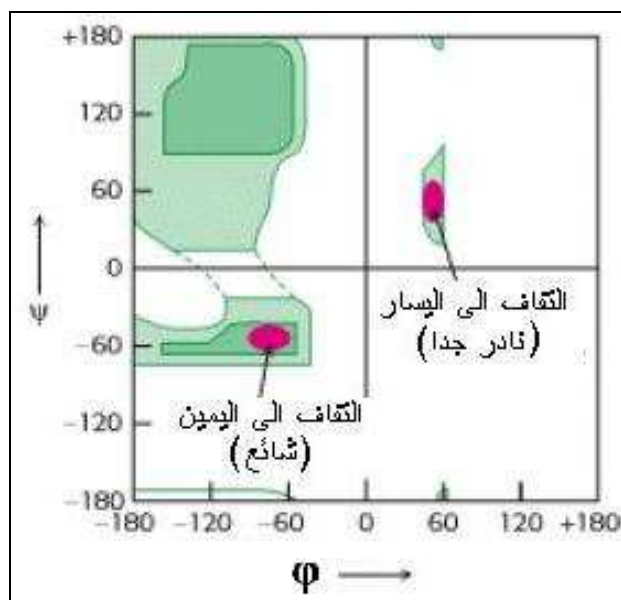
حلزونات ألفا

تركيب ينتج من التواء السلسلة الببتيدية المفردة، والتوزيع الأساسي للحلزونات يشبه المفك الميكانيكي Cork screw ويختلف محتوى البروتينات من الحلزونات ، فالبعض منها يخلو من هذه التراكيب في حين يكون البعض الآخر مكون بشكل كلي من حلزونات ألفا. حلزون ألفا تركيب شبه عصوي يمثل الجزء الداخلي منه السلسلة الرئيسية او العمود الفقري وتبرز السلاسل الجانبية الى الخارج كما موضح في الشكل 79.3 :



شكل 79.3 : تركيب حلزون ألفا

وتكون كل لفة في الحلزون مكونة من 3.6 حامض أميني والمسافة بين اللفات 5.4 انكستروم. ان اتجاه لوي الحلزون يمكن ان يكون الى اليمين باتجاه عقرب الساعة او الى اليسار بعكس اتجاه عقرب الساعة وفقا لمخطط راماجاندران الموضح في الشكل 81.3 :

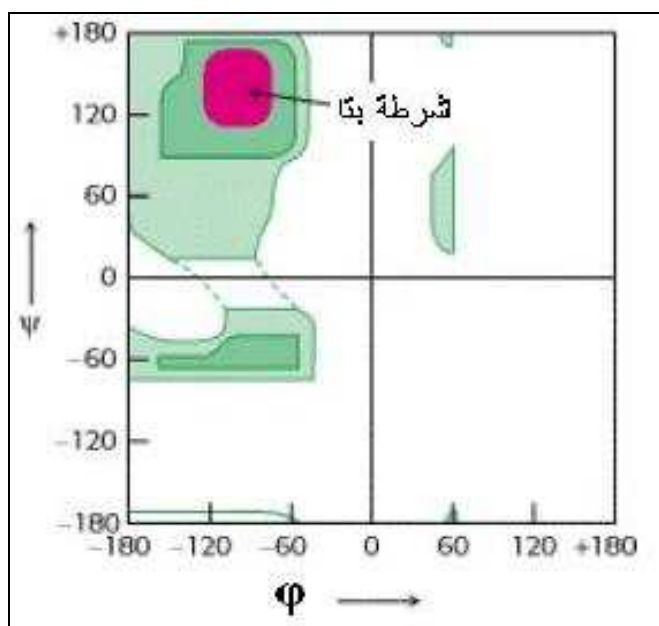


ان كلا الاتجاهين ممكنا ولكن في البروتينات فان الأغلبية العظمى من الحزونات تلتف الى اليمين لانها مفضلة من ناحية الطاقة وذلك بتقليل التصادم بين مجاميع السلسلة الرئيسية والسلاسل الجانبية. ويكون معدل زاوية ϕ حوالي 60° والزاوية ψ حوالي 45°. وتميل

الشمالات الكارهة للماء ان تكون الى الداخل والمحبة للماء الى الخارج ، وبناءا على الحسابات أعلاه فان كل ثالث ثمانية في التوالي تميل ان تكون كارهة للماء، وتميل حلزونات ألفا الى احتواء كل من الالنين والكلوتامين والليوسين والمثيانونين ولكن ليس الكلايسين والبرولين او التايروسين. وبما ان الأواصر ألفاعلة بشكل رئيس في هذا المستوى من التركيب هي الأواصر الهيدروجينية لذلك تكون أكثر ثبوتا عند قلة الماء.

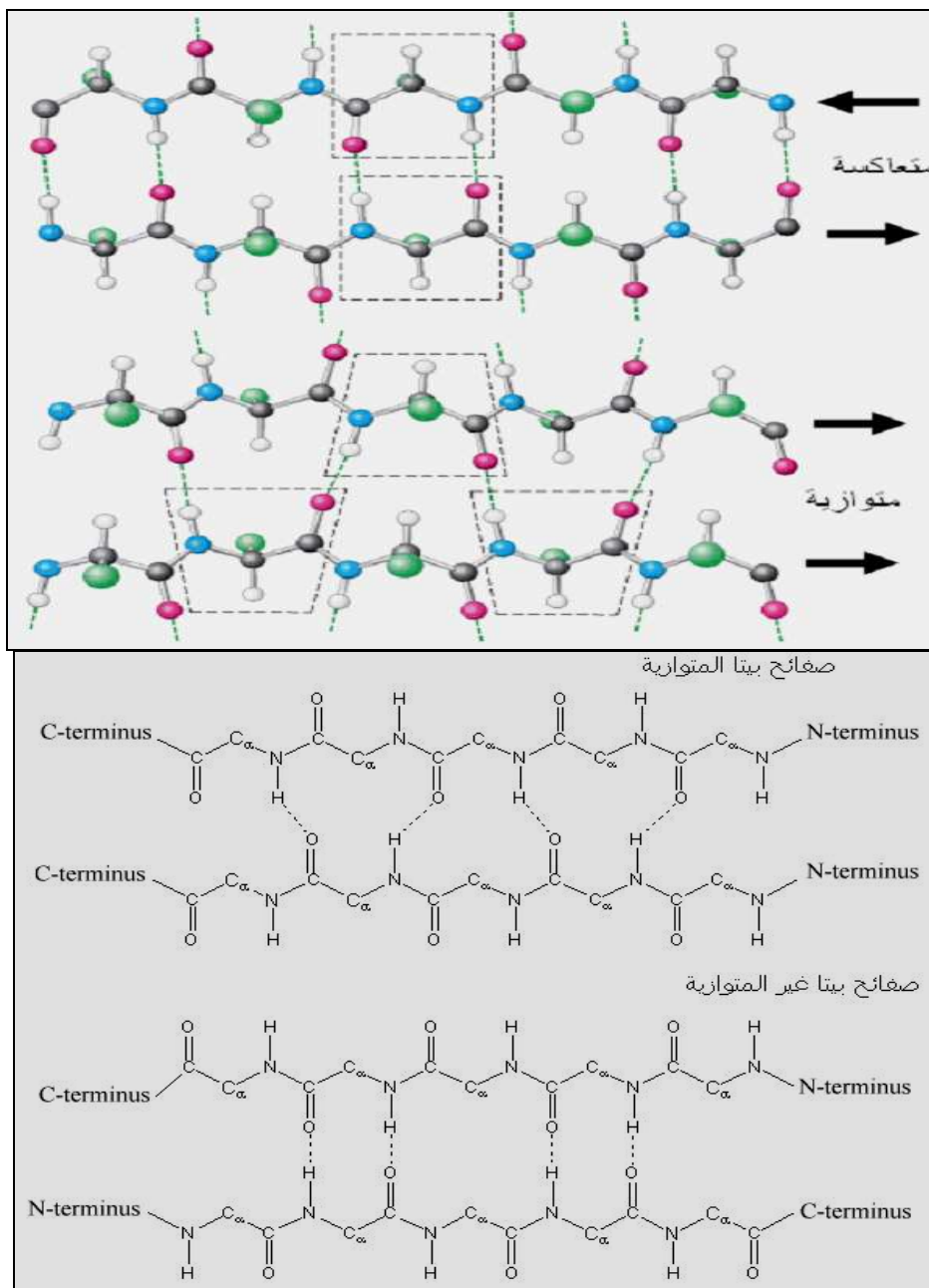
صفائح بيتا β -sheets

تراكيب ممتدة مكونة من أشرطة بيتا β -strands وهي سلاسل ببتيدية متعددة ممتدة وغير ملتوية كما في حلزونات ألفا ولذا يكون البعد بين الحوامض الامينية على طول الشريط حوالي 3.5 انكستروم مقارنة بـ 1.5 انكستروم في حالة الحلزونات وتكون السلاسل الجانبية موزعة الى أعلى وأسفل الشريط وقيم الزوايا ϕ و ψ للأشرطة موزعة على مخطط راماجاندران كما موضح في الشكل 82.3 :



شكل 82.3 : ملائمة تكوين أشرطة بيتا على مخطط راماجاندران

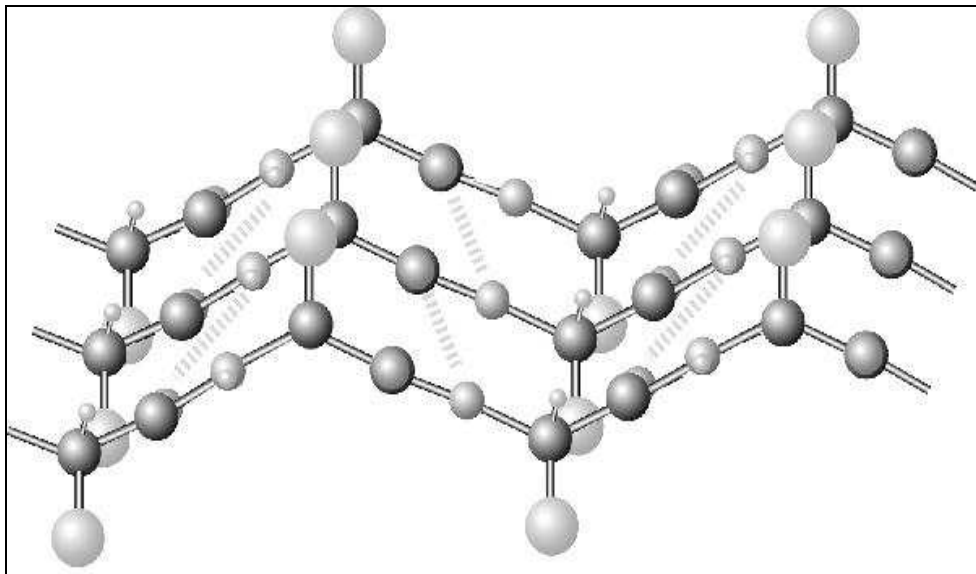
وتميل الأشرطة الى إظهار نمط متبادل من المناطق المحبة للماء والأخرى الكارهة له، وعلى العموم فان الأشرطة المظمورة تكون كارهة للماء وتتكون الصفائح من ارتباط شريطين او أكثر من أشرطة بيتا بواسطة الأواصر الهيدروجينية، فالأشرطة في صفائح بيتا اما ان تكون متوازية او متعكسة كما في الشكل 83.3 :



شكل 83.3 : انتظام الأشرطة المكونة لصفائح بيتا

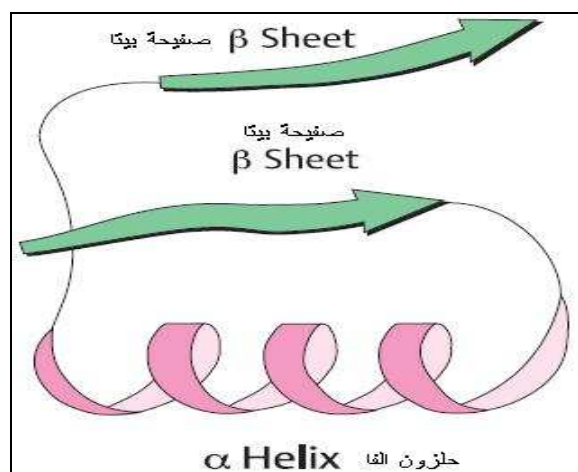
ويظهر من الشكل انه في حالة التوازي (الجزء الأسفل) فان الأصرة الهيدروجينية تربط كل حامض أميني من أحد الأشرطة باثنين من الحوامض الامينية في الشريط الثاني اي تكون مجاميع الأمين في الجهة نفسها ويكون الاتجاه من الطرف الأميني الى الطرف الكربوكسيلي. اما في الحالة المتعاكسة فان الأواصر الهيدروجينية تربط كل مجموعة CO من حامض أميني في أحد الأشرطة مع مجموعة أمين -NH لأحد الحوامض الامينية في الشريط الآخر ، ويمكن

ان تكون الصفائح خليط من الاثنين. ويمكن ان ترتبط عدة أشرطة والعدد الأمثل بين 4-5 ولكن قد تصل الى 10 لتكوين صفائح بيتا ويوضح الشكل التالي (84.3) المظهر العام لصفائح بيتا المتوازية:



شكل 84.3 : المظهر العام للأشرطة في صفائح بيتا

وبعض الأحيان تكون صفائح بيتا ملتوية بعض الشيء. وتكون صفائح بيتا من التراكيب المهمة في بعض البروتينات ، فالبروتينات الرابطة للحوامض الدهنية تكون تقريبا مكونة من صفائح بيتا. والشكل التالي (85.3) يوضح علاقة صفائح بيتا مع التراكيب الأخرى في البروتين



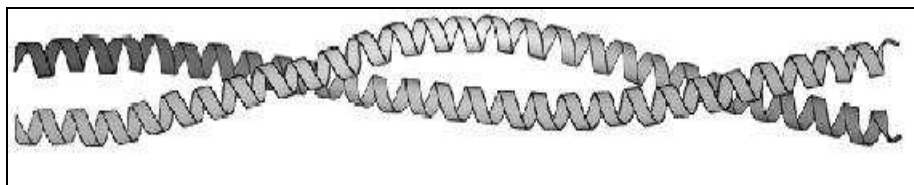
شكل 85.3 : ارتباط صفائح بيتا مع حلزونات ألفا

التركييب الأخرى

هناك تراكييب موضعية أخرى لا تعود الى حلزونات ألفا او صفائح بيتا تساهم في تحديد التركيب الثانوي للبروتينات، وتكون الحاجة اليها ماسة وذلك عند تغير اتجاه تركيب البروتينات كما في البروتينات الكروية الذي يحتاج الى عكس الاتجاهات وهذه الانعكاسات تتصف بالعموم بكونها استدارات حادة ومنها لفات بيتا β -turns وكذلك عروات اوميكا وانحناءات دبوس الشعر . وهذه التراكييب ليس لها نظام مرتب او تركيب متكرر وتراكييبها صلبة. واذا كانت غير مرتبة بشكل كامل فتعرف باللفات العشوائية . واللفات والعروات تقع في الغالب على سطوح البروتينات لان ثمالاتها تميل ان تكون من النوع المشحون وثمالات قطبية لذلك فهي تشارك في التداخل بين البروتينات او تداخل البروتينات مع غيرها من الجزيئات وتنشأ من تكوين الأواصر الهيدروجينية بين المواقع المختلفة وتكون مهمة من الناحية الوظيفية لانها توجد في العادة في المواقع الفعالة من البروتين. وتكون متغايرة من وجهة النظر التطورية اذ تعاني المناطق بشكل كبير من طفرات الحذف Deletion mutants وطفرات الإقحام Insertion mutants.

التركييب الثانوية الفائقة Supersecondary structure

تراكييب تتكون من تداخل مكونات التركيب الثانوي بعد ارتباطها بتداخلات مثل الأواصر الهيدروجينية واستعمال اللفات او الاستدارات او العروات لتكون شكل صلب او منضغط Compact shape. ومنها اللفات الملتوية Coiled coils وهي تراكييب خاصة بحلزونات ألفا التي تتصف بتجمع اثنين او أكثر من الحلزونات الملتفة على بعضها كما موضح في الشكل 86.3 :



شكل 86.3 : اللفات الملتوية

والحلزونات التي تكون هذا التراكييب الملتفة على بعضها لها نمط خاص من صفة الكراهية للماء، فتكون الثمالات الكارهة والمحبة للماء بنمط خاص يتكرر كل 7 ثمالات ، اذ تكون خمس منها كارهة للماء واثنين محبة للماء، أي يكون الطابع الغالب عليها هو كراهية الماء. وقد تظهر في هذا المستوى تراكييب ثانوية واضحة مرتبة بشكل دومينات ذات وظائف خاصة او قطيقات تركيبية مثل helix-turn-helix. وقد تشارك التراكييب الثانوية الفائقة في التنظيم كما في البكتريا التي تنظم عمليات الانتساخ ببعض من هذه البروتينات ، وكذلك الحال مع الأحياء الحقيقية النواة فمثلا زمام الليوسين Leucine zipper الذي تركيبه (helix-turn-helix) او أصبع الزنك تساهم في عمليات تنظيم الانتساخ .

ميل الحوامض الامينية للتراكيب الثانوية

ليس كل الحوامض الامينية تميل للدخول الى التراكيب الثانوية المذكورة أعلاه وذلك لأسباب عدة. وتكرار الحوامض الامينية وميلها للدخول في التراكيب الثانوية موضحة في الجدول 9.3

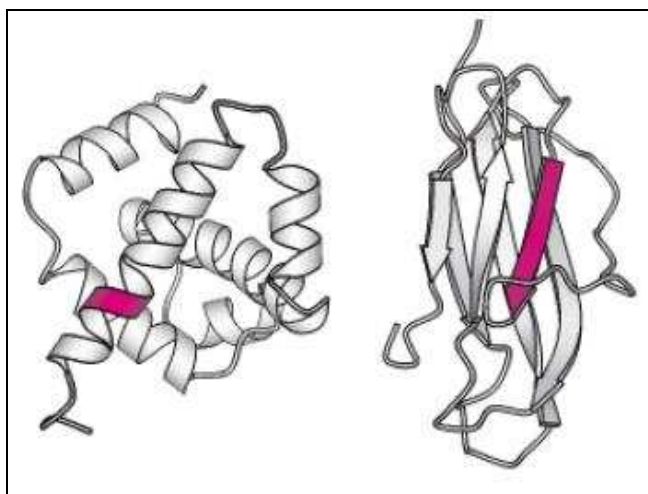
جدول 9.3 : ميل الحوامض الامينية للدخول في التراكيب البروتينية

الاستدارات	صفائح بيتا	حلزونات ألفا	الحامض الاميني
0.78	0.90	1.29	Ala
0.80	0.74	1.11	Cys
0.59	1.02	1.30	Leu
0.39	0.97	1.47	Met
1.00	0.75	1.44	Glu
0.97	0.80	1.27	Gln
0.69	1.08	1.22	His
0.96	0.77	1.23	Lys
0.47	1.49	0.91	Val
0.51	1.45	0.97	Ile
0.58	1.32	1.07	Phe
1.05	1.25	0.72	Tyr
0.75	1.14	0.99	Trp
1.03	1.21	0.82	Thr
1.64	0.92	0.56	Gly
1.33	0.95	0.82	Ser
1.41	0.72	1.04	Asp
1.28	0.76	0.90	Asn
1.91	0.64	0.52	Pro
0.88	0.99	0.96	Arg

فمثلا الحوامض الامينية الالانين والميثايونين والليوسين وحمض الكلوتاميك والسستئين والكلوتامين واللايسين والهستدين تكون أكثر ميلا للدخول في تراكيب حلزونات ألفا، ولكن دخول كل من الكلايسين والبرولين يؤدي الى تخريب الحلزون وخاصة البرولين الذي تركيبه Pyrrolidine based imino acid (HN=) وتركيبه الحلقي يحدد الحركة بشكل كبير حول الأصرة الببتيدية ويحدد قيمة الزاوية ϕ بقيمة قريبة من (-60°) التي يوجد فيها وبذلك فهو يتداخل مع امتداد الحلزون واضطراب الحلزون سيكون مؤثرا في الطويات الإضافية للسلسلة

الرئيسية للبتيد عندما يطوى ليكون البروتينات الكروية . ومن جهة ثانية يلاحظ من الجدول ان الفالين والايزوليوسين وغيرها (الموضحة في الجدول) تميل ان تكون من صفائح بيتا. والتفسيرات لها ان الهيئة الأصلية للتراكيب الثانوية هي حلزونات ألفا وان التفرعات عند الذرة β -carbon كما في الفالين والثريونين والايزوليوسين تميل الى ضعفة الحلزونات بسبب التصادم بين السلاسل الجانبية، وفي هذا المجال فان حجم السلسلة الجانبية يميل للمساهمة في ميل الحوامض الامينية. فضلا عن الصفات الأخرى للسلاسل الجانبية فمثلا السيرين والاسبارتات والاسبارجين يمكن ان تخرب الحلزونات لان سلاسلها الجانبية تحوي على واهبات ومستلمات للأواصر الهيدروجينية بشكل قريب جدا للسلسلة الرئيسة وبذا تتنافس مع المجاميع CO و NH للسلسلة الرئيسة. اما الكلايسين فيكون ملائما لجميع التراكيب تقريبا .

ومن العودة الى الجدول يلاحظ ان تفضيل الحوامض الامينية للدخول في أحد التراكيب لا يكون كاملا، فمثلا الكلوتامات وهي إحدى الثمالات المكونة للحلزونات بقوة وتصل قيمها الى 1.44 وهذا يكون ضعف ميلها للدخول الى صفائح بيتا البالغ 0.75 وهذا يشير الى تدخل عوامل أخرى. وقد لوحظ ان بعض التواليات الخماسية او السداسية تميل ان تكون تركيب معين في بروتين يختلف عنه في بروتين الآخر كما موضح في الشكل 87.3 :



شكل 87.3 : اختلاف قطع من تواليات معينة في إعطاء تراكيب ثانوية مختلفة في بروتينات مختلفة

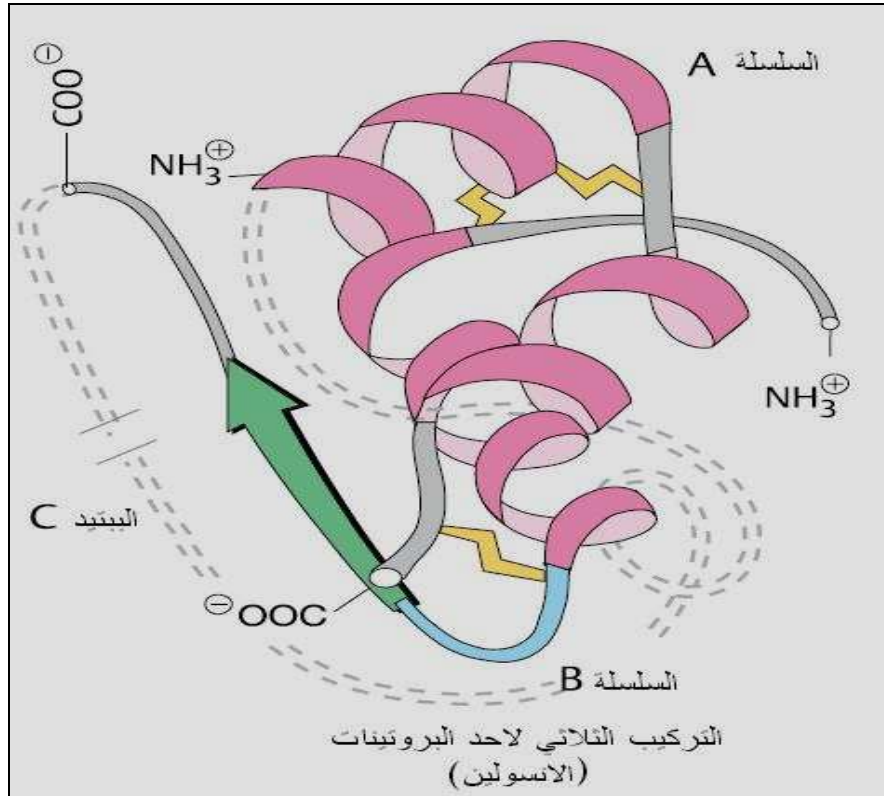
ومن هذا يستنتج ان بعض الحوامض الامينية لا تحدد التركيب الثانوي، ولذلك يلجأ الى التركيب الثلاثي والتداخلات التي تحصل بين الثمالات البعيدة في التركيب الثانوي ويمكن ان تكون الحد الفاصل في تحديد التركيب الثانوي على الأقل لبعض قطع البروتين.

التركيب الثلاثي Tertiary structure

التركيب الثلاثي المجسم يمثل مستوى أعقد من المستوى الذي سبقه ويدرس عادة باستعمال الأشعة السينية ومطياف الرنين النووي المغناطيسي. وفي هذا التركيب يصبح هناك توزيعاً

مكانيا" اي تركيب مجسم للتراكيب الثانوية اي ترتب العلاقة المكانية بالنسبة للتراكيب الثانوية الواحدة مع الأخرى.

وبصورة عامة تكون البروتينات الفعالة هي التي لها تركيب ثلاثي مجسم والذي يتحدد بتوالي الحوامض الامينية في التركيب الأولي وبالرغم من التطور الحاصل في مختلف النواحي العلمية والعملية الا انه ما تزال هناك فجوة بين ربط علاقة التوالي ووظيفة البروتين ، والحقيقة ان هناك قلة من البروتينات التي تم تحديد تركيبها الثلاثي المجسم وعليه فان تحديد التركيب الثلاثي من توالي الحوامض الامينية لا تزال مسألة مشرعة الأبواب. والشكل التالي (88.3) يوضح التركيب المجسم لأحد البروتينات (الأنسولين):



شكل 88.3 : التركيب المجسم (الثلاثي) لبروتين الأنسولين

وينتج التركيب الثلاثي من تداخلات غير متخصصة للتراكيب الثانوية وهي تداخلات لا تساهمة تعود الى التراكيب الثانوية وهي التي تحدد عمليات الطوي الكلي للبروتين. ومن التداخلات المساهمة في تثبيت التراكيب الثلاثية هي الأواصر الهيدروجينية والتداخلات الأيونية والجسور الكبريتيدية الثنائية وتداخلات القوى الكهربائية المستقرة وقوى فاندرفال. ولعل أهم هذه القوى هي الجسور الكبريتيدية فان تكوينها يقلص من الانتروبي مقارنة بالحالة غير المطوية، ومن الجدير بالذكر فان هذه الجسور نادرة في البروتينات السائتوبلازمية نظرا لكونها بيئة مختزلة جدا كما ذكر سابقا في وصف الأنظمة الحيوية.

ويظهر التركيب الثلاثي الأبعاد عند تعبئة وترتيب التراكيب الثانوية ويمكن ان تظهر بأشكال مختلفة منها البروتينات الكروية التي تكون متألفة مع الماء. وكذلك تظهر في البروتينات الغشائية التي توجد في المنطقة الدهنية من الغشاء وعندها تثبت بتداخلات كارهة للماء مع الدهون.

التركيب الرباعي Quaternary Structure

يمثل أعقد مستويات التراكيب البروتينية . وهو التركيب الناتج من تداخل عدد من ببتيدات متعددة في معقد بروتيني تتداخل فيما بينها بتداخلات لا تساهمية وكل ببتيد متعدد يمثل وحدة ثانوية او Monomer والكثير من البروتينات ذات التركيب الرباعي تحوي على اثنين او أكثر من الببتيدات المتعددة (الوحدات الثانوية) ترتبط مع بعضها بقوى لا تساهمية تشبه تلك التي تثبت التراكيب الثلاثية ومنها الجسور الكبريتيدية الثنائية.

ولذلك فهو يمثل التوزيع المكاني للوحدات الثانوية المتداخلة مع بعضها وأبسط أنواع التراكيب الرباعية هو المزدوج Dimer الذي يتكون من وحدتين قد تكون متشابهة كما في البروتينات التي لها القابلية للارتباط بـ DNA، ويليه في التعقيد المكوثر الثلاثي Trimer المكون من ثلاث وحدات ثانوية . والمجمعات الأكبر هي Multimer. وقد تكون الوحدات الثانوية متشابهة Homodimer او مختلفة Heterodimer وتسري التسمية على بقية أعداد الوحدات، ويطلق مصطلح قليل التعدد Oligo- على الوحدات حتى وان كانت بحدود أربع او خمس وحدات. ومن الأمثلة على البروتينات الفعالة وهي في حالة التركيب الرباعي هو الهيموغلوبين Haemoglobin الذي يحمل الأوكسجين في الدم فهو مكون من وحدتي ألفا ووحدتي بيتا مرتبة بشكل رباعي $\alpha_2\beta_2$ ولذلك فهو يمثل أحد البروتينات قليلة التعدد المتباينة Hetero-oligomer ويكون ترتيب الوحدات في الأغلب له علاقة بمحاور التناظر.

ويمكن للتراكيب الرباعية ان تتذبذب بين تراكيب متشابهة لأداء وظائف حيوية معينة وهو ما يسمى بتحول البنية الكيماوية وتكون حالة الانتقال هي التغيرات البنيوية ، وفي بعض البروتينات يكون الترتيب الرباعي الصحيح ضمانا لأداء وظيفة سليمة للبروتين. ولكن من الجدير بالذكر ان ليس كل البروتينات لها تركيب رباعي ووحدات ثانوية كي تكون فعالة وانما البعض منها يكون مكونة من وحدة واحدة.

بعض التراكيب الخاصة في البروتينات

القطيفات Motifs

تحتوي البروتينات على بعض التواليات الخاصة تسمى القطيفات التي يمكن ان تتكرر في توالي البروتين عدة مرات وتكون في بعض الأحيان بمثابة طمغة مميزة للبروتينات ذات العلاقة وعادة تكون مكونة من توالي حوامض أمينية محددة كما في EPIYA التي توجد في سموم CagA للبكتريا *Helicobacter pylori* . او تكون بشكل قطيفات تركيبية مكونة من عناصر قليلة مثل Helix-turn-helix (ثلاث عناصر تركيبية) وفضلا عن ذلك تضم العروات بأطوال مختلفة وبدون تركيب متخصص مما يجعل هذه القطيفات مكان تراخي ضروري لجلب العناصر التركيبية الأخرى في الفراغ والتي لا يشفر لها بواسطة تواليات DNA متجاوزة في الجين قريبة من بعضها البعض. وتستعمل بكثرة في برامج الحاسوب الخاصة بالتنبؤ بتركيب او تعريف البروتين (كما سيأتي ذكره في الفصول القادمة).

الانتينات Inteins

من التواليات الأخرى هي الانتينات وتمثل قطع من البروتين تستطيع استئصال نفسها وإعادة لحم أو وصل الأجزاء البقية التي يطلق عليها Exteins بأصرة ببتيديّة ، ولذلك يطلق عليها بعض الأحيان انترونات البروتين. تحوي عدد من الجينات قطع تشفر للانتينات في مواقع مختلفة منها ولذلك يطلق على هذه القطع بالعناصر الوراثية الأنانية Selfish genetic elements ولكن يكون من الأفضل تسميتها بالعناصر المتطفلة، وذلك لان العناصر الأنانية تتنافس مع الجينات الأخرى لأداء بعض الوظائف الخاصة، في حين ان المتطفلة تكون بدون وظائف. ويعتقد ان الانتينات تتوسط في وصل بعض البروتينات بعد ترجمتها من mRNA.

والبروتينات الحاوية على الانتينات الأولية تترتب بوجود الاكستين N-extein على الطرف الأميني يليه الأنتين ثم C-extein على الطرف الكربوكسيلي ، وبعد قطع القطعة الوسطية ينتج الاكستين ، وقد اكتشفت عام 1987 ومن ثم وجد انها توجد في أشكال الحياة الثلاثة البدائية النواة وحقيقيات النواة والاركية وكذلك في الفيروسات.

وهذه العناصر البروتينية تكون فعالة جدا في وصل وخياطة البروتينات لذلك وجدت لها تطبيقات مهمة في التقنيات الحيوية، فيوجد أكثر من 200 منها معروف في الوقت الحاضر أحجامها بين 100-800 حامض أميني وقد هندست لأغراض خاصة مثل الاشتراك في تخليق البروتينات أو تعليم قطع من البروتين لغرض دراسة البروتينات بطريقة الرنين النووي المغناطيسي. واستغلت أيضا في النواحي الصيدلانية حيث ان البروتينات الحاوية على قطع الأنتين سوف لا تقوم بوظيفتها وذلك لان وجودها يؤدي الى اضطراب تركيب البروتين وبذا تستعمل بعض الأدوية لمنع استئصال الأنتين وبذلك يضمن عدم أداء وظيفته خاصة بالنسبة للبروتينات المؤذية. كما انها تستعمل لغرض التعبير البديل Allotropic expression ويقصد به التعبير عن جينات الجينوم التي يعبر عنها في العادة في الماييتوكوندرية. والتعبير البديل أو المهندس وراثيا أقتراح ان يكون أحد الوسائل المستقبلية للعلاج الجيني لبعض الأمراض التي لها علاقة بالميتوكوندرية. فكراهية الماء لبعض البروتينات تعيق توريدها الى الماييتوكوندرية لذلك فإقحام الانتينات غير الكارهة للماء سيسهل عملية التوريد والدخول الى الماييتوكوندرية، ومن ثم يتم قطع الأنتين داخل الماييتوكوندرية ليعود البروتين الى حالته الطبيعية.

وهناك نظام خاص لتسميتها يعتمد على الاسم العلمي للكائن الذي يوجد فيه الأنتين والجزء له علاقة بالجين أو الاكستين ، فمثلا الأنتين الموجود في *Thermoplasma acidophilum* والذي له علاقة بإنزيم ATPase الموجود في الفجوة يسمى Tac VMA، فالحروف الثلاثة الأولى كافية لتحديد الكائن ولكن بعض الأحيان تتم إضافة حرف آخر أو رقم لتحديد السلالة وإذا كان هناك أكثر من أنتين يشفر بأحد الجينات فيعطى الأنتين ملحق رقمي ابتداء من 5' الى 3' مثل Msm dnaB-1. وقطعة الجين التي تشفر للأنتين تعطى الاسم نفسه الذي يكون بأحرف كبيرة Pfu IR R-1 والجين المقابل له يكون بأحرف صغيرة مائلة. وقد تكون الانتينات من النوع المنشطر أي ان أجزاءها يشفر لها بأكثر من جين.

ببتيدات الإشارة Signal peptide

ببتيدات تقع في الطرف الأميني للبروتين تساعد في توجيه البروتينات الى المسار الإفرازي. وتخلق هذه على الشبكة الاندوبلازمية الخشنة في الخلايا حقيقية النواة. وهناك البعض منها يوجه البروتينات الى الماييتوكوندريا مثل mtTP (Mitochondria targeting peptide) ، في حين يقوم cpTP (Chloroplast targeting peptide) في النباتات بإدخال البروتينات الى داخل البلاستيدات الخضراء. وفي الخلايا بدائية النواة فهي تساعد في تصدير البروتينات الى الخارج عبر الأغشية الخلوية .

الدومينات Domains

هذه الأجزاء التركيبية هي جزء من البروتين الكلي الذي يثبت نفسه ويطوى بشكل مستقل عن باقي توالي البروتين، ويمكن ان يظهر المتشابه منها في عدد من البروتينات، ويمكن ان تسمى بشكل منفصل وهي تتفرد بالوظائف الحيوية للبروتينات التي تنتمي اليها مثل دومين ربط الكالسيوم Calmodulin ، ونظرا لأنها تنتج من جينات متضاعفة او أجزاء من الجين التي يمكن ان تتضاعف وتتحرك على الجينوم كما يتضح من دراسات التطور، فهذا يعني انها يمكن ان تثبت بشكل منفصل عن باقي البروتين ويمكن ان تقايس او تنقل من بروتين الى آخر لتكوين الكايمرات Chimeras لأداء وظائف جديدة ، وهي الآلية التي تستعمل في العديد من البروتينات.

مواقع البروتينات

تؤدي البروتينات وظائفها في المكان المخصص لها بعد التخليق وإجراء التحويرات اللازمة عليها، فالبعض يبقى في الساييتوبلازم او السائل الخلوي والآخر يقتصص في التراكيب الغشائية وأخرى تفرز الى الوسط المحيط . كما ان بعض البروتينات تحتاجها الخلايا طوال حياتها في حين هناك بروتينات تحتاجها الخلايا تحت ظروف محددة، كما تختلف الكميات اللازمة في كل حالة.

ومكان وجود البروتينات يكون مقنناً فمثلا المستضدات التي تحفز الجهاز المناعي تكون موجودة على سطوح الخلايا مثل البكتريا وتكون مرشحة لإنتاج اللقاحات الوقائية Vaccines، في حين تكون للبروتينات المطمورة في الأغشية دوراً كبيراً في التأقلم البيئي مثل مكونات الأغشية المكونة من الأنظمة الثنائية Two component systems. اما البروتينات المفروزة فهي التي تترك الخلايا بشكل كامل وتكون عادة سامة للخلايا المنتجة لها، او تستعمل لقتل الخلايا المضيفة فتقوم الخلايا البكتيرية مثل بحقن المضيف بها مما يؤدي في بعض الحالات الى قتل المضيف بسرعة . ولكل من البروتينات توالي خاص به يؤهله ان يوجد في المكان الصحيح فمثلا البروتينات الغشائية تسود فيها الثمالات الكارهة للماء، في حين تحوي البروتينات المفروزة على إشارات خاصة تسمى توالي الإشارة Signal sequence او ببتيدي الإشارة التي توجد عند الطرف الاميني للبروتين، وإشارات الإفراز هذه هي التي تسحب البروتين الى الخارج بنظام الإفراز والعملية تعتمد على:

- البروتين المرشح للإفراز الى خارج الخلايا
- ماكينة الإفراز

وتزال أو تفصل الإشارات المذكورة من البروتين الأولي Pre-protein أثناء عملية الانتقال عبر الغشاء الخلوي. وهناك العديد من الإشارات المرتبطة بالطرف الأميني التي تم تشخيصها في البكتيريا المسؤولة عن قيادة البروتينات إلى مسارات الإفراز المختلفة سواء في البكتيريا السالبة أو الموجبة لصبغة كرام . وأكثر ببتيديات الإشارة هي التي تقود البروتينات إلى نظام الإفراز المسمى Sec-dependent الذي يعرف أيضا بالنوع الثاني Type II ، وأشارت دراسات الجينومات المقارنة إلى أن هذا النظام ثابت في كل البكتريات تقريبا، فالبروتينات التي تحمل هذه الإشارة تميز من قبل وصيقات متخصصة تحملها إلى ثقب النقل في الأغشية وتنقل البروتينات خلالها بشكل غير مطوي (أي ممسوخ). والإشارات الخاصة بالنوع الثاني من أنظمة النقل تتكون من تواليات قصيرة من الحوامض الأمينية 7-15 ثمانية بشكل دومين مكون من حلزونات ألفا، والحوامض الأمينية المشاركة تكون ذات شحنة موجبة مثل الأرجنين واللايسين، وتفصل قطعة حلزون ألفا هذه بعدد قليل من الثمالات والتي تكون أكثر تفضيلا هي المكونة من الالنين والسيرين عند موقع الانفلاق الذي يعمل عليه إنزيم Signal peptidase I وهذا النمط يكون مختلف قليلا عن البروتينات الدهنية التي تفرز بالنظام نفسه . والنظام الآخر للإفراز الذي يوجد في البكتيريا الموجبة والسالبة لصبغة كرام يعرف بنظام الأرجنين المزدوج أو التوام Twin arginine ويسمى أيضا Tat secretion system، ويعتمد المسار على نظام وماكنة إفراز مختلفة عن المذكورة أعلاه ولكن إشارة الإفراز تكون شبيهة بتلك الخاصة بالنظام أعلاه ولكنها أقصر قليلا وتحتوي على ثمالتين متجاورة من الأرجنين، وهذا النظام ينقل البروتينات التي تم طويها بشكل كامل في السائل الخلوي.

وهناك أنظمة نقل أخرى تستخدم لنقل القليل من البروتينات منها

ABC transport system ويعرف أيضا بنظام النقل من النوع الأول أو TISS وهذه توجد في البكتيريا السالبة لصبغة كرام بشكل رئيس الحاوية على غشائين الأول الغشاء الخلوي أو الداخلي والغشاء الخارجي الذي يكون خارج الجدار الخلوي، وتحتوي على فسحة محيطية Periplasm بين الغشاء الداخلي والجدار الخلوي. وتعتمد على ثلاث مكونات:

- ناقل يحوي على مواقع ارتباط للـ ATP وجاءت التسمية من ATP-Binding-Cassette.
- بروتين مندمج في الغشاء Fusion protein الذي يعمل حلقة وصل بين الغشاء الداخلي والخارجي.
- بروتين ثلاثي الوحدات Trimeric protein في الغشاء الخارجي يعمل على وصل الخلايا بمحيطها الخارجي.

وهذا النظام يكون متخصصا "بالمادة التي ينقلها وتكون جيناتها متجاورة في الجينوم ، وفي هذا النظام يكون الطرف الكربوكسيلي هو المحدد لعملية الإفراز وليس الطرف الأميني ولا ينفلق توالي الإشارة أثناء الإفراز ويستعمل النظام لإفراز عدد من السموم وبعض الإنزيمات الخارجية مثل البروتينات واللايبيزات وبعض البكتيريوسينات .

وتتملك بعض البكتريات نظام آخر متخصص للإفراز هو النوع الرابع Type IV ويسمى أيضا T4SS الذي يكون تركيبه يشبه شفرة السكين ينتهي من الخلية ويشبه الشعيرات Pili

وبواسطته يفرز البروتين الى الخارج او يحقن في الخلية المضيفة كما في حقن سم CagA من قبل البكتريا *Helicobacter pylori* في خلايا المعدة للمضيف. ونظام الإفراز الثالث **Type III (T3SS)** يوجد في بعض *Proteobacteria* ويوجد في الأنواع الممرضة من البكتريا المعوية *Enterobacteriaceae* وبعض الممرضات النباتية. اي في عدد قليل ويتكون من تركيب ابري الشكل يخترق الغشائين وينتج الى الخارج على سطح البكتريا ويسمح النظام بنقل بعض البروتينات من الساييتوبلازم الى الخارج مباشرة او الى الخلايا المضيفة مباشرة. ويتألف النظام من 20 مكون تركيبية ويحتاج الى بروتينات تنظيم إضافية ووصيفات كي يكون فعالا، كما ان بعض مكوناته تشبه مكونات الاسواط. ووجد ان جينات هذا النظام توجد في جزر الجينوم *Genomic islands* والبروتينات التي تفرز بهذا النظام موجودة ضمن هذه الجزر وهذا التواجد القريب على الجينوم يعد من الأساسيات للتنبؤ بالجينات المسؤولة عن البروتين المفروز.

تصنيف البروتينات

تصنف البروتينات لغرض تسهيل دراستها وهو من المزايا المهمة كتوجه عام للإعلان عن الجينومات، وتعتمد عملية التصنيف على عدد من المؤشرات مثل:

- التركيب
- الفعالية
- الدور الأيضي

وهذه تسهل عملية التعرف على البروتينات التي يكون من الصعب التعرف عليها بالاعتماد على عمليات الاصطاف *Alignment* فضلا عن انها تساعد في إدامة قواعد البيانات وذلك بتطوير عمليات الإعلان وإيضاح الأخطاء في التهميش والتوضيح، وفي العادة يرافق عملية التحديث التزويد بوسائل مهمة لاسترداد *Retrieval* البيانات اللازمة من بين الكم الهائل من البيانات، كما ان تصنيف البروتينات الى عوائل تعكس عوائل الجينات المسؤولة عنها والمهمة في دراسة الجينومات المقارنة *Comparative genomics* وإيجاد العلاقات التطورية. وهناك أكثر من طريقة لعلمية تصنيف البروتينات منها:

- التصنيف الهرمي لعوائل البروتينات الكبيرة *Superfamily* والتي تصنف الى عوائل، وفي هذه الحالة يعتمد التصنيف على تشابه التواليات ليعطي عنايد شاملة وغير متداخلة بشكل ترتيب هرمي الذي يعكس علاقاتها التطورية، والطريقة تسمح بضم البروتينات الجديدة التي تظهر تشابها "كاملا" للتواليات وبالتالي يشمل التشابه الأطر العامة للدومينات المشتركة من حيث التركيب او العدد.
- بعض العوائل تصنف على أساس التشابه في الدومينات والقطيفات.

وفي جميع الأحوال فان علاقة البروتين مع العائلة هي ليست علاقة واحد الى واحد فهي يمكن ان تكون واحد الى عدة أي ان العائلة يمكن ان تشمل العديد من البروتينات. وتصنف البروتينات الى عوائل كبيرة ثم عوائل يساعد في تحديد وظائف البروتينات غير معروفة الوظيفة او ما تسمى بروتينات افتراضية.

بروتينات خاصة

هناك بعض البروتينات حازت على الاهتمام اما لبساطة تركيبها واستعمالها كموديولات دراسية، او لانها تميزت بخاصية معينة توضح جانب معين من الدراسات، وسيتم التطرق الى البعض منها التي استعملت في تصميم البرامج للتنبؤ ببروتينات أكثر تعقيدا ويجب ان تذكر ان تحديد التوالي الأولي يكون حلقة الوصل بين المعلومات الموجودة في DNA والتركيب الثلاثي المؤدي للوظائف وهو ما تعتمد عليه العديد من البرامج.

• **البروتينات الخارجية:** وتمتاز باحتوائها على الجسور الكبريتيدية الثنائية المكونة من السستئين والتي تكاد تخلو منها البروتينات داخل الخلايا.

• **البروتينات المعقدة:** وتمثل البروتينات التي ترتبط مع جزيئات من هويات أخرى مثل الكربوهيدرات ، وتركيبها هذا ينتج بعد عمليات الترجمة ضمن تحويلات بعد الترجمة ، وتؤثر في التخصص الوظيفي للبروتينات الناتجة، ويطلق على البروتينات المرتبطة بالكربوهيدرات مصطلح البروتينات السكرية Glycoproteins وتوجد منها نوعان N-linked مثل ارتباط السكر الى Amide للمجموعة الجانبية الموجودة في الاسبارجين، او تكون O-linked وفيها ترتبط مجموعة الكربوهيدرات الى مجموعة الكربوكسيل لكل من السيرين او الثريونين او لمجموعة الهيدروكسيل للحامض الأميني المحور Hydroxyl lysine. وهناك العديد من البروتينات السكرية مثل تلك الموجودة على سطوح كريات الدم الحمر والاختلاف بينها يكون بالجزء السكري وهي التي تحدد مجاميع الدم التي تزيد عن 100 مجموعة تعود الاختلافات فيها الى الجزء السكري وأهمها A, B, O والتي تخصص بفعالية بعض الجينات المسؤولة عن ادماج السكر وكذلك في حالة البروتينات السكرية المفروزة .

وتتداخل البروتينات مع الدهون بواسطة تداخلات لا تساهمية لتكون البروتينات الدهنية Lipoproteins والتي تتعلق وظيفتها بخرن ونقل الدهون والكوليسترول في الجسم.

• **البروتينات الكروية:** تراكيب بروتينية تكون بشكل كريات منضغطة وتكون ثمالاتها الكارهة للماء الى الداخل والمحبة للماء والقطبية الى الخارج لذلك تكون ذاتية ومحاطة بجزيئات الماء، ومن الأمثلة الشائعة الإنزيمات و Myoglobins ومحرركات الخلايا Cytokines والهرمونات البروتينية.

• **البروتينات المكونة من حلزونات ألفا:** من أفضل الأمثلة عليها بروتين الفرتين Ferritin الخازن للحديد ، و 75% من البروتين مكون من حلزونات ألفا، والحلزونات فيه يمكن ان تلتف مع بعضها مكونة تركيب ثابت يصل طوله الى 1000 انكستروم. وصفائر حلزونات ألفا توجد في بروتينات العضلات والبروتينات الليفية وكرياتين الشعر والقرون والأظافر، حيث يعمل حبل الحلزونات في مجال الوظائف الميكانيكية بعد ان يكون حزم من الألياف القوية. وكذلك الحال مع الهيكل الخلوي الذي يمثل دعائم ومنصات الخلية الداخلية التي تكون غنية بمثل هذه الخيوط.

• **البروتينات الغشائية:** تختلف عن باقي البروتينات بكونها توجد في محيط دهني ولغرض ان تثبت في مكانها لابد من ان تتوزع ثمالاتها الكارهة للماء الى الخارج والمحبة للماء الى

الداخل. وفي معظم البروتينات التي تستعرض الأغشية تكون تراكيبها الخارجية من حلزونات ألفا (كما مر ذكره) .

أما بالنسبة للبروتينات التي تقع خارج الغشاء الخلوي وغير المفروزة مثل البروتينات الواقعة في الفسحة المحيطة في البكتريا فيلاحظ ان نمط ترتيبها يختلف ، فهي تتكون من أشرطة بيتا، وبعض العروات التي تربط القطع ، والعروات هنا تقع في الطور المائي لذا تكون من الثمالات المحبة للماء ، ولكن بعض الأحيان تقع هذه العروات في مناطق تلاقي الطور الدهني للغشاء مع الطور المائي خارج الغشاء لذلك كانت من النوع ثنائية الصفات Amphipathic أي توجد فيها ثمالات قطبية تواجه الطور المائي وأخرى كارهة للماء تكون على الجانب الدهني. ان هذه الثمالات ثنائية الصفات يمكن ان تكون بشكل حلزونات تمتلك صفة الدورية لكل 3 او 4 ثمالات (أي التناوب بين ثمالات كارهة للماء وأخرى محبة للماء).

عوامل الانتساخ Transcription factors

بروتينات خلوية تعد الأساس في عمليات التنظيم وذلك بارتباطها الى مواقع محددة على DNA، وبالتالي تحدد حركة او انتساخ المعلومات الوراثية من DNA وإنتاج mRNA أي ان لها القابلية للارتباط بـ DNA ، وقد تؤدي المهمة لوحدها او بمساعدة مجموعة أخرى من البروتينات وذلك بتكوين المعقدات ويمكن ان تؤدي الى تنشيط تخليق mRNA او إيقافه بواسطة توجيه الإنزيم RNAP الى المنطقة المطلوبة. ونظرا لكونها من أساسيات التنظيم لذلك يتوقع ان توجد في كل الكائنات الحية. وتختلف أعدادها في الكائنات ولكن الصفة العامة ان عددها يزداد بازدياد حجم الجينوم ، وفي الجينومات الكبيرة يوجد أكثر من عامل انتساخ ١ جين. وفي الإنسان يوجد حوالي 2600 بروتين تحوي على دومينات للارتباط بـ DNA واغلبها من عوامل الانتساخ ، ولذلك فهي تشكل اكبر عائلة بروتينية في الإنسان اذ ان حوالي 10% من الجينوم البشري يشفر لهذه العوامل. كما ان الجينات تنجح بمواقع واضحة للارتباط بعوامل الانتساخ ويكون التعبير الكفوء للجينات محتاجا لتعاون تأثير عدد من عوامل الانتساخ.

التركيب العام لعوامل الانتساخ

تمتاز معظم عوامل الانتساخ بتركيب أساسي الذي يظهر في الشكل التالي (89.3) الذي يوضح تركيب البروتين وطرفيه الأميني والكربوكسيلي



شكل 89.3: التركيب العام لعوامل الانتساخ

ويحوي البروتين على ثلاث مناطق رئيسة هي:

DNA-binding domain = DBD الدومين الذي يرتبط الى DNA

Signal-sensing domain = SSD الدومين الذي يتحسس الإشارات

Transactivation domain = TAD الدومين المنشط

ان ترتيب وعدد هذه الدومينات يمكن ان يختلف في أنواع مختلفة من عوامل الانتساخ وبعض الأحيان يؤدي دومين واحد وظيفتي التحسس (SSD) والتنشيط (TAD). ان الدومين الذي يرتبط الى DNA (DBD) هو الذي يرتبط الى مواقع معينة من DNA مثل المشجعات او الممهدات وتوجد قواعد بيانات خاصة تضم العديد من عوائل هذا الدومين ، ويسمى التوالي الذي ترتبط اليه عناصر الاستجابة Response elements وهي تواليات من DNA ، ويشمل الارتباط قوى كهربائية مستقرة والأواصر الهيدروجينية وقوى فاندرفال. وعلى العموم فان التداخلات تختلف في قوة ارتباطها ، ولذلك يرتبط عامل الانتساخ بعدة مناطق من التوالي القريبة من بعضها جدا بقوى متفاوتة من التداخلات، فمثلا TBP أحد عوامل الانتساخ يمكن ان يرتبط الى عدد من التواليات المشابهة للـ TATA box ، ولكن التخصص في الارتباط يتم بالاعتماد على عوامل أخرى تعود للـ DBD نفسه من خلال حصول الازدواجية Dimerization لاثنين من العوامل .

اما TAD الذي يكون مكانا لارتباط بروتينات أخرى مثل مساعدات تنظيم الانتساخ Transcription coregulators ويسمى عادة بدومين التنشيط . ويمثل دومين تحسس الإشارات جزءا " غير ضروري لكل عوامل الانتساخ ، وعند وجوده يتحسس الإشارات الخارجية وينقلها الى باقي معقد الانتساخ مؤديا الى رفع او خفض التنظيم للتعبير الجيني، ومن الجدير بالذكر ان كل من DBD و SSD يمكن ان يقعا على بروتينات منفصلة التي ترتبط الى معقد الانتساخ لتنظيم التعبير الجيني.

وفيما يخص TADs فقد أطلقت عليها تسميات اعتمادا على مكوناتها من الحوامض الامينية التي قد تكون أساسية للفعالية او تكون الأكثر وفرة في دومين TADs فضلا عن اختلاف عدد الحوامض الامينية في الدومين. فمثلا عملية التنشيط بعامل الانتساخ Ga14 يتم بواسطة الحوامض الامينية الحامضية لذلك يسمى دومين التنشيط الحامضي ، بينما عامل الانتساخ Gcn4 فتسود فيه الحوامض الامينية الكارهة للماء لذلك يسمى دومين التنشيط الكارهة للماء Hydrophobic activation domain. وهناك برامج حاسوب خاصة يمكن ان تتعرف على هذه الدومينات.

تصنيف عوامل الانتساخ

هناك أكثر من وسيلة لتصنيف عوامل الانتساخ منها التصنيف على أساس الآلية التي تعمل بها، او الوظيفة التنظيمية او تشابه التوالي.

اولا : تصنيف عوامل الانتساخ بالاعتماد على الآلية منها ما يشترك في تكوين معقدات قبل البدء مثل TFIIB ، TFIID ، TFIIE ، TFIIIF ، TFIIF ، TFIIF . وهذه تكون شائعة وتتداخل مع منطقة لب الممهد المحيطة بموقع بدء الانتساخ لكل جينات الصنف الثاني Class II، وكذلك تشمل عوامل الانتساخ اليسارية وتتمثل بالبروتينات التي ترتبط الى يسار موقع البدء لتحفز او تكبح الانتساخ وهذه مرادفة لعوامل الانتساخ الخاصة. وتختلف المجموعة الأخيرة فيما بينها اعتمادا على تواليات التعرف Recognition sequences الموجودة قريبة من الجين وهي كثيرة منها أصبع الزنك و Basic-Zipper وغيرها.

ثانيا : التصنيف الثاني المعتمد على الوظيفة فتقسم عوامل الانتساخ الى:

- عوامل الانتساخ الدائمة Constitutively active TFs وتوجد في كل الخلايا وعلى طول الوقت.
 - العوامل الاشتراطية Conditionally active TFs وهذه تحتاج الى تنشيط وتكثر تحت ظروف خاصة.
 - العوامل الخاصة بالتطور Developmental active TFs وتكون هذه في خلايا خاصة والتعبير عنها يكون تحت سيطرة صارمة وعند بدء التعبير عنها وإنتاجها سوف لا تحتاج الى تنشيط إضافي.
 - العوامل المعتمدة على الإشارات وهذه تحتاج الى إشارات خارجية للتنشيط ومنها المستلمات النووية التي تحتاج الى رباط خارجية Paracrine-dependent او داخلية Endocrine للتنشيط. والبعض الآخر غير النووي يحتاج الى رباط من داخل الخلية Autocrine dependent مثل P53 والمستلمات النووية اليتيمة وغيرها.
 - العوامل المقيمة في النواة وهذه تقيم في النواة بغض النظر عن حالتها من التنشيط.
 - العوامل السايروبلازمية الكامنة Latent cytoplasmic factors وتوجد في السايروبلازم بصورة غير نشطة وعند تنشيطها تنتقل الى النواة.
- ثالثا :** التصنيف الثالث للعوامل يعتمد على تركيبها وتشابه تواليات الحوامض الامينية فيها وبالتالي تشابه التركيب الثلاثي للدومينات التي ترتبط الى DNA، وهذه تقسم الى العديد من الأصناف والأخيرة تقسم الى عوائل كبيرة والتي تقسم بدورها الى عوائل متعددة.
- آليات عمل عوامل الانتساخ**
- ترتبط عوامل الانتساخ اما الى مناطق المشجعات او الممهديات في DNA المجاورة للجينات التي تنظمها اعتمادا على عامل الانتساخ وبالتالي يكون التنظيم اما عاليا او اوطئا، ويكون ذلك بعدة أشكال:
- تثبيت او غلق ارتباط الإنزيم RNAP الى DNA.
 - المساعدة في أستلة او إزالة الاستيل من الهستونات ويكون ذلك اما بشكل مباشر او من خلال تجنيد بروتينات أخرى للمساعدة في هذه العملية.
 - ومن البروتينات العاملة في هذا المجال إنزيم أستلة الهستونات Histone (HAT) acetyltransferases الذي يقوم بإضافة مجموعة الاستيل الى الهستونات وبذا يساعد على إضعاف الترابط بين DNA والهستونات مما يجعل DNA متاحا لعملية الانتساخ وزيادتها. اما الإنزيم الآخر Histone deacetyltransferase (HDAC) فيقوم بإزالة مجموعة الاستيل من الهستونات مما يؤدي الى تقوية الترابط بين DNA والهستونات وبالتالي جعل DNA غير متاح لعملية الانتساخ او يقلل من عملية الانتساخ اي يكون هناك تنظيم سلبي.
- وظائف عوامل الانتساخ**
- عوامل الانتساخ يمكن ان تقرأ وتفسر البصمة في DNA ، فارتباطها وعدم ارتباطها يؤثر في الوظائف الخلوية . وتشارك عوامل الانتساخ في الكائنات متعددة الخلايا في عمليات التطور والاستجابة للمؤثرات الخارجية. والعوامل الفعالة في هذا المجال تعمل على فتح وغلق الجينات المطلوبة والتي بالتالي تؤثر في التغيرات الحاصلة في المظهر الخارجي للخلايا، او الفعاليات

التي تحتاجها الخلايا لتقرير مصيرها النهائي الذي يحصل عند تخصص الخلايا ومنها عائلة Hox و SRY في الإنسان الواقعة في منطقة تحديد الجنس الذي يعمل بشكل مهم في تحديد الجنس.

فضلا عن ما ذكر فانها تشارك في مهام أخرى مثل حالة التواصل بين الخلايا. فالخلايا تتواصل بإطلاق جزيئات تعمل على بدء سلسلة من التفاعلات بعد استلامها للإشارات ، ويكون التنفيذ النهائي لسلسلة نقل الإشارات من مهام عوامل الانتساخ التي تحفز أو تثبط العمليات للـ DNA. وتشارك العوامل في :

تنظيم الفعاليات في حقيقيات النواة بواسطة عوامل الانتساخ

في هذه المجموعة من الخلايا هناك مجموعة من عوامل الانتساخ العامة General transcription factors (GTFs) الضرورية لحدوث الانتساخ، وعدد منها لا يرتبط الى DNA وانما يكون أحد مكونات المعقدات الأولية للبدء التي تتداخل مع RNAP مباشرة. ومنها TFIID ، TTFIIB ، TFIIA التي تقع مواقعها الى يسار الجين الذي تنظمه. وإضافة الى عوامل الانتساخ العامة هناك عوامل أخرى تنظم الجينات بشكل تفاضلي وذلك بالارتباط الى المشجعات من DNA المجاورة للجينات المراد تنظيمها. والأخيرة تكون مهمة لتحديد التعبير الذي يحصل في خلايا محددة وفي الوقت المناسب والكميات المناسبة اعتمادا على التغير في حاجة الكائن لها.

• الاستجابة للظروف الخارجية

تعمل عوامل الانتساخ في الاستجابة للظروف الخارجية وتنظم التعبير الجيني على ضوء الإشارات القادمة ومنها تنظيم الجينات المسؤولة عن تخليق بروتينات الصدمة الحرارية، حيث يقوم عامل الصدمة الحرارية Heat shock factor (HSF) الذي يكون مهما لتنشيط الجينات اللازمة عند ارتفاع درجات الحرارة. وكذلك العامل الآخر Hypoxia-inducible factor (HIF) الذي ينشط الجينات الضرورية لعيش الخلايا في بيئات قليلة الأوكسجين، ومثال آخر العامل Sterol regulatory element binding protein (SREBP) الذي يساعد في الحفاظ على مستويات الدهون في الخلايا.

• السيطرة على دورة الخلية

هناك بعض عوامل الانتساخ الخاصة بجينات السرطان Oncogenes والجينات الكابحة للأورام Tumor suppressor genes تساعد في تنظيم دورة الخلية اذ تحدد حجم الخلية وغيرها من المؤشرات لتحديد متى يجب ان تنقسم الخلية الى خليتين بنوية ، ومنها c myc والذي له دور مهم في نمو الخلايا والاستماتة.

تنظيم الخلايا لعوامل الانتساخ

من المعروف ان الخلايا تحتاج الى مستويات مختلفة من التنظيم وعوامل الانتساخ (التي لها وظيفة تنظيمية) وعوامل الانتساخ تحتاج الى عمليات تنظيم هي الاخرى وهذه تتم بعوامل انتساخ أخرى. وتنظيم عوامل الانتساخ يمكن ان يتم بعدة آليات:

- تخليق عوامل الانتساخ كغيرها من البروتينات التي تنتسخ من جينات على الكروموسومات الى RNA الذي يترجم الى بروتينات عوامل الانتساخ وأي خطوة من هذه الخطوات تكون عرضة للتنظيم الذي يؤثر في إنتاجها وفعاليتها.
- بعض الأحيان تنظم عوامل الانتساخ ذاتيا مثل استعمال عمليات وعروات التنظيم الرجعي السلبي فيعمل عامل الانتساخ كابحا لنفسه ويقل إنتاجه بحيث يبقى ضمن الحدود المطلوبة.
- عوامل الانتساخ كغيرها من البروتينات (في الخلايا حقيقية النواة) تنتسخ في النواة ثم تترجم في السايكوبلازم. وبعض عوامل الانتساخ الموجودة في النواة لها إشارات خاصة تبقىها في النواة، وهذه تعد مهمة في عملية تنظيم عوامل الانتساخ لأنها تؤدي وظيفتها في النواة، لذلك وجب نقلها من السايكوبلازم بعد تخليقها على الرايبوزومات وهذا يحتاج الى ربطها برابط خاصة ثم يتم نقلها الى النواة. ولذلك فإن هذه العملية تساعد في التنظيم بمساعدة الدومين الحساس للإشارة SSD باستعمال طرق مختلفة، ومنها الارتباط الى الربائط Ligands وهذا الارتباط ليس هو الوحيد الذي يؤدي الى انتقال وحركة عوامل الانتساخ في الخلية. والارتباط الى الربيطات يمكن ان يؤثر في عامل الانتساخ وفيما اذا سيكون قادرا للارتباط الى DNA او عوامل مساعدة أخرى.

- الفسفرة لبعض عوامل الانتساخ تكون لازمة لتصبح قادرة على الارتباط الى DNA .
- تتداخل عوامل الانتساخ مع بعضها لتكوين المزدوجات او ترتبط الى بروتينات منظمة مرافقة وبذا تصبح قادرة على أداء وظائفها .

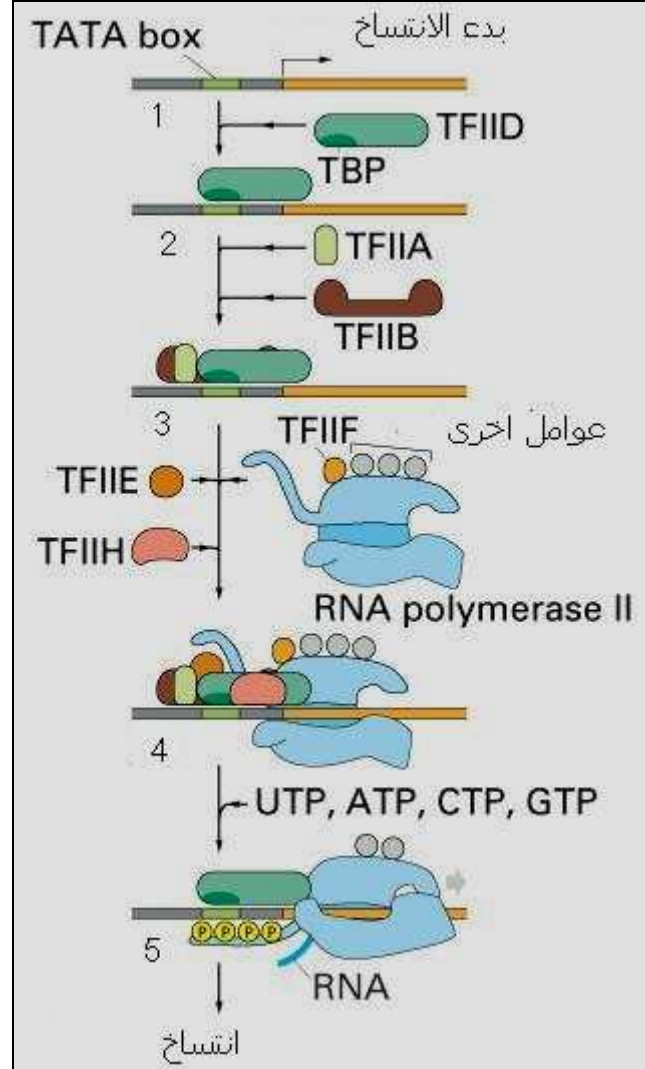
الوصول الى مواقع الارتباط الى DNA

في حقيقيات النواة ينظم DNA مع الهستونات لتكوين الجسيمات النووية بعد التفاف 147 قاعدة منه حول مركز ثماني من الهستونات. و DNA في هذه الحالة يكون غير متاحا لعدد من عوامل الانتساخ، وهذا يعني ان هناك عوامل أخرى قادرة للوصول الى DNA في الجسيمات النووية وهذه تسمى العوامل الرائدة Pioneer factors . وفي العموم فإن الجسيم النووي يجب ان يتغير بطريقة ما مثل إعادة تشكيل الكروماتين، او فك التفاف الجسيم النووي بالتغيير الحراري مثلا ليؤدي الى إتاحة مؤقتة لعوامل الانتساخ بالدخول والارتباط الى الأماكن المخصصة لها، وفي هذا الموضع يحتاج عامل الانتساخ الى التنافس مع الهستونات او البروتينات اللاهستونية لذا يمكن ان تحصل حالة تضاد في تنظيم الجين نفسه. ومن الجدير بالذكر ان معظم عوامل الانتساخ لا تعمل لوحدها وانما تحتاج الى عدد من البروتينات والعوامل المساعدة الأخرى لتسمح بتجمع المعقدات الأولية مع RNAP كما موضح في الشكل (90.3) .

الأهمية السريرية لعوامل الانتساخ

لعوامل الانتساخ أهمية سريرية كبيرة وذلك لان:

- الطفرات التي تحصل فيها ترتبط بعدد من الأمراض.
 - يمكن ان تكون العوامل أهدافا دوائية.
- ومن الأمراض المتعلقة بعوامل الانتساخ هو السرطان بشكل رئيس وذلك لتأثيرها في جينات السرطان او الجينات الكابحة للأورام.



شكل 90.3 : تداخل عوامل الانتساخ مع البروتينات الأخرى للقيام بعملية الانتساخ

تحويلات بعد الترجمة وأداء البروتينات للوظائف

يبدأ في بعض الأحيان أداء الوظيفة لبعض البروتينات مباشرة بعد تخليقها ولكن في معظم الأحيان تحتاج البروتينات إلى عدد من التحويلات التي يطلق عليها تحويلات بعد الترجمة التي تكون أساسية لأداء الوظائف أي أن البروتين المخلق تَوّاً يكون طليعة لبروتينات أخرى ومن هذه التحويلات:

- طوي البروتين المشار إليه سابقاً
- إضافة العوامل المرافقة Cofactors
- إضافة السكريات
- إضافة الدهون

وبعض البروتينات تحتاج الى دمج وحدات بروتينية مع بعضها (كما في حالة إنتاج التركيب الرباعي) لتكون معقدات فعالة. ومن التحويلات الأخرى تكوين الجسور الكبريتيدية الثنائية والفسفرة وغيرها.

التنظيم Regulation

يعد تنظيم الفعاليات في الأنظمة الحيوية من أعقد المواضيع نظرا لان الفعاليات داخل الخلية تحصل بشبكات متداخلة مع بعضها ولكن سيقسم الموضوع الى عدة جوانب يتم فيها توضيح الآليات الخاصة به وان كان هناك تداخل كبير ثم بعد ذلك قد يسهل الربط بينها مع الأخذ بنظر الاعتبار ان كل كائن حي يمكن ان يستعمل طريقة مختلفة في التنظيم. وبما ان الفعاليات الحيوية لها مسارات يكمل احدهما الآخر ومنها الانتساخ تليه الترجمة، ويعد الأول الأكثر أهمية وتكون عمليات التنظيم على مستويات مختلفة ويمكن ان تشمل التنشيط الرجعي او Downstream effect فضلا عن وجود بعض الجينات التي لا تحتاج الى تنظيم وتقوم بفعاليتها طول حياة الخلية ولكن قد تكون بمعدلات مختلفة وهي جينات الإدامة Housekeeping genes وكلها تكون ضمن شبكات تنظيم معقدة.

مستويات تنظيم الانتساخ

يحدث تنظيم لعملية الانتساخ اما بمستوى عام وغير متخصص او يكون بمستويات متخصصة تشمل جين واحد او عدد قليل من الجينات ولعل أهم مرحلة في تنظيم الانتساخ هي عملية البدء وذلك يوفر الطاقة والجهد على الخلايا ، وبعد هذه المرحلة تعمل بقية المراحل بمعدلات ثابتة تقريبا، اذ لا فائدة من بدء عملية نسخ أحد الجينات لا تحتاجه الخلية وإلغاءه بعد مرحلة نظرا لعدم الحاجة اليه. ومن جانب آخر فان بعض جوانب تنظيم الانتساخ يمكن التنبؤ بها من تحديد تواليات الجينوم. وهناك عدة مستويات للتنظيم تشمل التنظيم العام والخاص الذي تعمل به جزيئات متخصصة جدا وقليلة ، وستدرس هذه في نوعي الخلايا بدائية وحقيقية النواة.

المستوى الأول: موقع الجين يحدد مدى إمكانية انتساخه ، ففي الخلايا حقيقية النواة تكون بعض الجينات غير متاحة للانتساخ عند وقوعها في مناطق الإسكات من الكروماتين المتباين، ولغرض انتساخها يتم تغيير طبيعة المنطقة للسماح لإنزيم كثرة RNA (RNAP) للدخول لانتساخها.

اما في بدائية النواة فان الجينات المراد انتساخها بمستويات عالية تكون في العادة قرب منطقة التضاعف ومنها جينات rRNA.

المستوى الثاني: التغيرات في الهيئة التركيبية للـ DNA وهذه تنظم بالهستونات والبروتينات الشبيهة بالهستونات (في حالة البكتريا) فهناك بعض مناطق DNA تكون ملتفة بشدة حول الهستونات وبذا تكون اقل عرضة للانتساخ. ففي البكتريا *E. coli* هناك بروتين IHF شبيه بالهستونات تنتشر نسخ من تواليه بألاف النسخ في جينوم البكتريا اعتمادا على ظروف التتمية ويظهر في أطلس الكروماتين ، ويلاحظ فيه ان المناطق ذات التعبير العالي تحوي القليل من IHF وهذه العملية يطلق عليها الإسكات الكروماتيني Chromatin silencing ، وهي واضحة جدا في الخلايا حقيقية النواة ولكنها تهمل في كثير من الأحيان في بدائية النواة.

المستوى الثالث: وفرة العامل سيكما في الخلايا بدائية النواة اذ ان العامل سيكما يساعد في فتح أشرطة DNA المزدوجة فهو يرتبط الى منطقتين من الممهد 10- و 35- (كما ذكر سابقا) وقد

تتغير هذه المواضع بالنسبة لعوامل سيكما أخرى، والارتباط الموضح يولد عزم لفك التواء الأشرطة ومن ثم فتحها لذلك تعد وفرة العامل سيكما أحد آليات تنظيم بدء الانتساخ ، وتحصل عمليات التنظيم هذه عند انتساخ الجينات الخاصة بمعالجة الاجهادات والاستجابة لها مثل سيكما 54 وتلك المتعلقة بعملية التبوغ Sporulation التي تنظم بوفرة العامل سيكما.

المستوى الرابع: وهو خاص ايضا بالخلايا بدائية النواة ويتمثل باختيار العامل سيكما. فالعامل ينتسخ من جينات خاصة به ويحتاج الى عامل سيكما خاص به لتخليقه وهذا ما يشير الى وجود شبكات متداخلة لعمليات التنظيم بمختلف آلياتها الموجبة والسالبة والشبكات هذه محكومة بعدد من بروتينات الإشارة التي تستشعر الظروف الداخلية والخارجية للخلية اي ان الخلايا تكون مهياة بشكل مثالي للاستجابة بمختلف الظروف وعلى ضوءها يتم إنتاج العامل سيكما المطلوب والتوالي الى يسار الممهد هو الذي يحدد اي نوع من عوامل سيكما يمكنها الارتباط.

والعامل سيكما المذكور أعلاه والذي دارت حوله المناقشة هو سيكما 70 الخاص بعمليات الايض تحت الظروف الاعتيادية وهو اقل ثبوتا من بقية العوامل وتوجد الآلاف من المواقع التي يرتبط بها في جينوم البكتريا مثل *E. coli* ، بينما تقل المواقع الى مستوى المئات لعوامل سيكما الأخرى. وقد تكون بعض هذه المواقع متشابهة في الأنواع القريبة من بعضها جدا وذلك لأنه لا يوجد ممهد واحد يعمل في كل أنواع البكتريا.

وفيما يخص المستوى الثالث والرابع الخاص بالبكتريا يمكن القول ان طبيعة وعدد عوامل سيكما تختلف بين أنواع البكتريا وان كان هناك نوعا من الثبوت ضمن الجنس الواحد، لذلك كانت هناك صعوبات في التعبير عن الجينات الغريبة ضمن مجال الهندسة الوراثية. والإجماع في هذا المجال يشير الى ان مواقع ارتباط العامل سيكما في *E. coli* كموديل دراسي هي ثبوت وجود TA في بداية box -10 المعروف TATAAT او ما يسمى بـ Pribnow box ، والمعروف ان ترادف البيورينات بعد البريميدونات يؤدي الى التقليل من ثبوت الحلزونات وبالتالي يسهل فتحها وذلك لان هذا الترادف يحتاج الى طاقة قليلة لفتحها.

ومما يلاحظ في الخلايا بدائية النواة ان ليس كل جين له ممهد خاص به نظرا لوجود الاوبرونات التي تعمل على انتساخ الجينات ذات العلاقة الوظيفية في مسار ايضى محدد ولكن هناك بعض البكتريات التي تنتج رسائل متعددة وتعطي بروتينات تقوم بوظائف ليس بينها اي ارتباط.

المستوى الخامس: ويخص ايضا العامل سيكما الذي يتأثر بعدد من البروتينات ويمكن ان تؤثر بشكل مختلف فهي اما تشجع عملية النسخ او تغلقها وهذا المستوى يكون في مجال التنظيم الخاص جدا.

المستوى السادس: المستوى يخص ثبوت جزيئات mRNA وقد يكون ضمن التنظيم العام او الخاص ويصعب الجزم بالحالة لان التأثير يعتمد على الجينات المعنية. والجزيئات القصيرة العمر تنتج بروتينات قليلة مقارنة بالجزيئات طويلة العمر. وفي الحالتين فان الجزيئات القصيرة او طويلة العمر سوف تتفكك والا أصبح إنتاج البروتينات بلا حدود وهي حالة غير مرغوب فيها. وطول عمر الجزيئات يكون بمدى الثواني الى دقائق فمثلا معدل عمر الجزيئات في *E. coli* حوالي دقيقتين ويمكن ان يطول بتأثير بعض التراكييب التي تمنع التفكك، كما ان بعض الجزيئات غير المشفرة ncRNA (التي سيأتي ذكرها لاحقا) يمكن ان ترتبط الى mRNA

مؤدية الى إعاقة او تشجيع ترجمتها. وفي حالة الاوبرونات التي تعطي رسائل متعددة من mRNA، فان الفواصل بين الجينات يمكن ان تكون تراكيب تمنع RNAP من الاستمرار وتقليل كميات mRNA للمناطق التي تقع على يمينها مقارنة بالجينات الواقعة الى يسارها، فضلا عن مشاركة جزيئات ncRNA في كفاءة ترجمة الرسائل الناتجة. وعلى العموم فان ثبوت جزيئات mRNA يؤثر في عمر الجزيئات ويعتمد على تراكيب حلقية التي تكون بشكل عروات عند الطرف 3' من الجزيئة التي يمكن ان تقضم بالإنزيم RNase ويفككها.

المستوى السابع: يخص الممهد ولكن من جوانب أخرى وهي طريقة عمل العوامل من حيث الموقع، فالعناصر المنظمة في اغلب بدائيات النواة تعمل في الموقع Cis وذلك بالارتباط بمواقع خاصة Specific-Cis –acting قريبة من الممهد كما في حالة 10- و 35- المذكورة أعلاه ، في حين في الخلايا حقيقية النواة قد تكون العوامل المؤثرة بعيدة جدا عن الممهد كما في المشجعات اي تكون ضمن Trans يمكن ان تؤدي العوامل البعيدة الى إسكات الجين. ومن جهة ثانية فان التفاضات أشرطة DNA حول الهستونات يكون له تأثير في التنظيم اذ توفر التقوس في الأشرطة لتسهيل تداخلها مع RNAP.

وضمن هذا المستوى تدرج المنظمات الايجابية اي عوامل الانتساخ ، والجينات التي تخلقها توجد في مواقع مختلفة وبعيدة عن الجينات التي تنظمها اي انها ضمن التأثير البعيد Trans، فضلا عن العوامل السلبية والمثبطات التي تعمل ايضا عن بعد، وهناك المئات من عوامل الانتساخ العاملة في استجابات الخلايا وتكون ضمن شبكات خلوية مختلفة .

تنظيم الترجمة

العملية التي تلي عملية الانتساخ والتي تبدأ في الأحياء بدائية النواة قبل انتهاء مرحلة الانتساخ، اما في حقيقية النواة فتكون منفصلة وتتم في أجزاء مختلفة في الخلية (كما مر ذكره في موقع آخر).

في الخلايا بدائية النواة والنمط المذكور أعلاه يكون تنظيم الترجمة بوفرة جزيئات tRNA وتحميلها بالحوامض الامينية المطلوبة، وعملية التحميل ليس من السهولة التنبؤ بها نظرا لانها تؤثر بوفرة الحوامض الامينية المعتمدة على الحالة الايضية للخلية.

ونوع جزيئات tRNA تكون محددة بتواليات الجينوم ويمكن استنباطها منه، اما أعدادها فيعتمد على مواقع جيناتها التي تكون قرب نقطة التضاعف في الخلايا بدائية النواة تشير الى وجود أعداد كبيرة منها ، كما ان وجود بعضها ضمن اوبرونات rRNA يشير الى الاحتياج الكبير لها. ومن جهة ثانية تتدخل مسألة استعمال الشفرات في توفير هذه الجزيئات. لذلك يكون التنظيم على طرفي نقيض، فالبروتينات المنتجة بشكل كبير تستعمل الحوامض الامينية التي شفراتها تغطي بوفرة من tRNA الخاصة بها، في حين ان البروتينات الحاوية على حوامض امينية التي تكون جزيئات tRNA الخاصة بها غير شائعة فتكون عملية ترجمتها بطيئة. وعليه فان التفاضل في استعمال الشفرات الوراثية يؤثر في عملية الترجمة وبالتالي التعبير البروتيني وهذا ما يوضح الفشل عند إدخال جينات من أصول مختلفة او مصنعة الى الخلايا لغرض التعبير عنها. لذلك ولأجل حصول تعبير مثالي ومن ورائه الحصول على كميات جيدة من البروتينات

يجب اختيار الشفرات المفضلة لدى المضيف. وهناك جداول وقواعد بيانات خاصة بهذه المسألة على الأقل بالنسبة للنموذج الدراسي *E. coli*. وعند حصول انتقال أفقي للجينات يلاحظ ان العامل الأكثر تأثيرا في قبول الجينات الغريبة هو نوعية الشفرات للحوامض الامينية متعددة الشفرات فالجينات الملائمة والمفيدة للكائن المضيف او على الأقل غير المؤذية هي التي تبقى وتنتشر في مجتمع المضيف.

جوانب التنظيم

بعد ذكر مستويات تنظيم الانتساخ ومشاركة بعض الجزيئات فيها، هناك جوانب أخرى قد لا تخص الانتساخ وهذا يمكن ملاحظته من دراسة الجزيئات المشاركة.

التنظيم بـ DNA

يحتوي DNA بالإضافة الى الجينات المشفرة على مخزون كبير من التواليات الخاصة بالتنظيم، فهو يشفر للعديد من جزيئات RNA غير القابلة للترجمة، كما ان بعض الفيروسات تبدأ موادها الوراثية من RNA لتكوين البروتينات. اذن يلعب DNA دوراً أكثر من الدور المزدوج وهي الحفاظ على المعلومات الوراثية والتشفير للبروتينات، وانما ينظم إنتاج البروتينات وتقنينها وفق حاجة الخلية، وهذا ما يلاحظ في الخلايا حقيقية النواة التي تحوي على الكمية نفسها من DNA في الكائن الواحد ولكنها تختلف من حيث المكون البروتيني، فضلا عن الاختلاف في نوعية البروتينات الناتجة تحت الظروف المختلفة والتي تكون أكثر وضوحا في الأحياء وحيدات الخلايا.

ويؤثر DNA في التعبير الجيني عند تغير طوبوغرافيته وهذا النوع من السيطرة يقع ضمن التأثير العام ويكون غير متخصصا" وبذا يكون ملائما لكبح انتساخ عدد من الجينات كما في الكروماتين المتباين، وهذا يسمح بالتعبير عن جينات في أنسجة خاصة مثلا في المراحل الجنينية، وتلغى فعاليتها بعد الحياة الجنينية. ويساهم DNA في تطور الأحياء، فالأخطاء التي تحصل من قبل إنزيم الكوثر DNAP والتي لا يتم إصلاحها تعبر الى الخلايا الجديدة وفي حالة كونها مفيدة فإنها ستكون مصدرا للتطور، اما غير المفيدة فتعطي إشارات للخلايا للتعامل معها. كما ان قابلية DNA الخاص بالخلية الى تقبل قطع جديدة من كائنات وخلايا أخرى تمثل ايضا رافدا مهما لعمليات التطور، لذلك يعد DNA ومن ورائه الجينوم تركيب غير ثابت وانما هو في حالة تطور مستمر وان كان غير ملحوظ نظرا لطول المدة التي يستغرقها. ومشاركة DNA في التنظيم هو سماحه لبعض البروتينات بالارتباط بتواليات خاصة فيه للتعبير ومنها عوامل الانتساخ ويتم التحكم بهذه عن طريق فتح او غلق التعبير الذي يعتمد على عدة ظروف أهمها تواليات DNA.

مشاركة RNA في التنظيم

تشارك جزيئات RNA بشكل كبير في تنظيم الفعاليات الحيوية وبطرق مختلفة وطرق دراستها والكشف عن فعاليتها ونسق وجودها في الخلايا والتقنيات المستعملة في هذا المجال تقع ضمن مجال يطلق عليه دراسة مكنون RNA هو RNAomics ويستهدف هذا المجال من الدراسة تحديد تواليات جزيئات RNA ويكون هذا محددا للجزيئات التي تكون موجودة بكميات كبيرة في الخلية ومميزة عن غيرها في الطول والثبوت مثل rRNA و tRNA ويشمل مجال الدراسة

هذا الكلونة وإنتاج cDNA الناتجة من جزيئات RNA بعد تحديد تواليها لتحديد نسق وجودها في الخلايا للاستفادة منها وهي الأكثر استعمالاً، كما تستهدف الدراسات الكشف عن الجزيئات المنظمة باستعمال طرق التهجين وهي الأخرى من الطرق المستعملة على نطاق واسع، ويضم المجال دراسات Genomic SELEX وتهدف هذه تحديد تواليات جزيئات RNA الخاصة التي ترتبط ببروتينات خاصة وإنتاجها خارج الأنظمة الحية باعتماد تواليات الجينوم كأساس والطريقة تطبيقها محدود.

ومن أهم الآليات التي يشارك فيها RNA في عمليات التنظيم:

تفكك جزيئات mRNA: عملية تخليق البروتينات يمكن أن تنظم بمدى حياة جزيئات mRNA التي تكون قصيرة وتتكك بسرعة بعد أداء مهامها، ففي الخلايا بدائية النواة تكون أعمارها بين عدة ثواني إلى قليل من الدقائق، أما في الخلايا حقيقية النواة فتند في عدة دقائق إلى حوالي 10 ساعات كما في β -globulin mRNA. وهناك عدة آليات مستقلة تسيطر على انقلاب وتكك mRNA ومنها :

- عملية إزالة ذيل الأدينين تليها عملية التفكك التي يتوسط فيها معقد الجسيم الخارجي Exosome complex، فالجزيئة تحوي بحدود 200 قاعدة تكون في حالة قصر مستمر بتأثير أنواع من إنزيمات إزالة الأدينين Deadenylases التي تعمل بالاتجاه من 3' إلى 5' وعندما يصل طول الجزيئة إلى حد حرج تبدأ بالتفكك، وبعد إزالة الأدينين فإن الجسيم الخارجي وهو معقد بروتيني كبير يحوي على Exoribonucleases يقوم بتفكيك الجزيئة من النهاية 3' ويعمل المعقد البروتيني المذكور أيضاً في تفكيك RNA الناتجة من الطفرات التافهة وبذا يمنع تخليق بروتين مبدور أو يحوي على أخطاء. إن عملية تقصير ذيل الأدينين تختلف من جزيئة إلى أخرى وتعتمد على الجزيئات المرتبطة بالـ RNA التي يمكن أن تزيد أو تقلل من إزالة الأدينين.

- التفكك بإنزيمات قطع داخلية خاصة.

- يمكن أن يحصل الفلق استجابة للارتباط إلى جزيئات RNA صغيرة مثل siRNA و miRNA (التي سيأتي ذكرها لاحقاً).

- إزالة الغطاء Cap وهو من الجزيئات يوفر فرصة لعمليات هضم إضافية بالاتجاه من 5' إلى 3' بواسطة إنزيمات القطع الخارجية.

جزيئات RNA غير المشفرة

الدراسات الحديثة مثل دراسة مقارنة جينومات عدد من الكائنات التي استهدفت مشاركة RNA في الفعاليات الحيوية بعيداً عن كونه حلقة الوصل بين المكون الوراثي إلى ما يترجم عنه من بروتينات وفعاليات أوضحت أن هناك نمطاً من الثبوت في تواليات الانترونات والمناطق بين الجينات يصل إلى 3-10 مرات أكثر مقارنة بالمناطق المشفرة وهذا الثبوت يشمل مناطق أو قطع Blocks مكونة من عدة مئات من النيوكلوไทيدات في جينومات الكائنات التي شملتها الدراسات وهي الإنسان والفيضان وفقرات أخرى مختلفة. والضغط الانتخابي الذي أدى إلى ثبوت هذه التواليات غير معروف ولكن أحد التفسيرات أنها تكون جزءاً من الشبكات التنظيمية الحيوية ذات الأنماط متعددة التداخل تؤدي إلى جعل حاصل التغيرات المعوضة لكل المكونات

مساوية للصفر ولكن هذه المناطق قد تعاني من التغيرات والانفراج نوعاً ما مما يشير الى وجود انتخاب ايجابي للتغيير في هذه التواليات التي لها علاقة بالاختلاف المظهري. وبناءاً عليه فان هناك الكثير من جزيئات RNA غير المشفرة Non-coding RNA وهو المصطلح العام الذي يضم كل الجزيئات عدا mRNA ومنها tRNA و rRNA والجزيئات المنظمة (Riboregulators) والتي بمشاركة عدد من البروتينات يمكن ان تقوم بشتى العمليات من التنظيم. فعدد نسخ mRNA التي يمكن ان تنتج ما يزيد عن عدد الجينات نظراً لان كل جين يمكن ان ينتج عدداً مختلفاً منها (وتم مناقشة هذا في موقع آخر ضمن موضوع الخياطة البديلة واقتصاد الجين) ، وقد أشارت الدراسات الرياضية ومعادلات الاشتقاق للعلاقة بين التنظيم لعدد من الجينات ان العلاقة غير خطية مع عدد الجينات في الشبكات الخلوية، وذلك لان الجينات التي تنتج نسخ باستعمال الخياطة البديلة او الجينات الجديدة تحتاج الى تنظيم خاص بها وهذا يعطي زيادة خطية، ولكن المنظمات نفسها تحتاج الى تنظيم وهذا يعني ان الكل يندمج في شبكات ودوائر متداخلة ولذا ظهرت العلاقة اللاخطية . ولذلك فان وجود جزيئات مختلفة من RNA يعكس الوظائف المتعددة المسؤولة عنها. والجزيئات غير المشفرة تتاطب بها وظائف عدة منها العمل في مجال الخياطة البديلة والاشتراك في عملية الترجمة وتنظيم التعبير الجيني، وتؤدي في الغالب الى تقليل التعبير الجيني لذلك تسمى العملية إسكات الجين Gene silencing.

أصول الجزيئات غير المشفرة

هناك أكثر من مصدر للجزيئات غير المشفرة، فالبعض منها ينتج من الانترونات الموجودة في الجين كما في تأثير الخياطة البديلة في إنتاج MicroRNA (miRNA) كنسخ أولية ثم تعالج لتستعمل فيما بعد. فالانترونات التي تشكل في بعض الأحيان ما بين 95-97% من الجينات المشفرة للبروتينات وكون Intronic RNA يتفكك بعد فصله من المناطق المشفرة تصور غير صحيح اذ ان هذه الانترونات تكون فعالة وتعطي وتغذي شبكات التنظيم الخلوي بالوفير من المعلومات . ويمكن ان تشتق الجزيئات غير المشفرة من المناطق بين الجينات ومن التواليات المكررة وبعض الجزيئات تشتق ايضا من الاكسونات.

ويمكن تقسيم الجزيئات الى مجموعة ذات التأثير البعيد Trans-acting وأخرى Cis-acting ، اما المجموعة الثانية فتكون عاملة بشكل Cis-acting، وتحدث هذه من قبل الجزيئات المضمورة في mRNA التي ستترجم الى بروتينات في مناطق 5'-UTRs وتؤثر في التعبير بطرق مختلفة، مثل Riboswitch (الذي يمثل جزءاً من جزيئة mRNA التي ترتبط الى جزيئات صغيرة) ويمكن ان يرتبط بقطعة هدف صغيرة في الجين المستهدف مؤثراً في فعاليته. وهناك العديد من التراكيب RNA المضمورة في جزيئات أخرى تؤثر في عملية التنظيم بطرق مختلفة.

الآليات العامة لتأثير الجزيئات غير المشفرة

ان اشتراك جزيئات مهمة مثل الانترونات في عمليات التنظيم تشير الى وجود خط مواز للتنظيم الذي تقوم به الجينات المشفرة (بما تنتجه من وسائل تنظيم) ومسؤولية الجينات المنظمة. فالجينات في الخلايا حقيقية النواة تنتج بروتينات هي نفسها تكون بمثابة مصدراً لإشارات RNA التي يمكن ان تتواصل مع بعض الجينات او منتجاتها بشكل مستقل لأداء وظائف خلوية

وبهذا يمكن القول ان ظهور العمليات المتوازية في الأنظمة الحيوية تكون أساسية لعمليات التطور في الأحياء المعقدة، اذ ان التعقيد في الكائنات الحية يفرض بزوغ عدة جينات تشفر لإشارات RNA كمستوى عالي في شبكات التنظيم. وهذا يمكن توضيحه بان الأحياء حقيقية النواة المعقدة جداً تمتلك عدد من الانترونات الكبيرة مقارنة بأقرانها من الأحياء البسيطة من جهة، ومن الجهة الثانية فقد ازداد اكتشاف أعداد الجزيئات غير المشفرة بشكل كبير جداً والتي يمكن ان تعادل ما يقرب من نصف مكنون النسخ الناتج في بعض الأحياء مثل اللبائن.

وفي العموم يكون تأثير الجزيئات غير المشفرة بشكل غير مباشر وذلك بارتباطها الى بروتينات مكونة Ribonucleoprotein complex (RNP) والتي تكون تأثيراتها عند مستوى بعد الانتساخ او مرحلة الترجمة وهذا المعقد او الماكنة تكاد تكون عامة للأحياء مع وجود بعض الاختلافات ، اما الطريق المباشر فهو عمليات ازدواج القواعد المتكاملة وفعاليات الجزيئات غير المشفرة تكون بعد الانتساخ او أثناء الانتساخ وفيما يلي بعض هذه الآليات او الفعاليات:

- بعض الجزيئات المشتقة من الأشرطة النقيضة الطبيعية Natural antisense transcripts (NAT) ترتبط الى عوامل الانتساخ او الى DNA مباشرة مؤدية الى اضطراب وظائفها.

- البعض من NAT يمكن ان تثبط الانتساخ وذلك بتحويل الهستونات في منطقة الممهد.
- بعض الجزيئات ترتبط الى البروتينات العاملة في عملية الانتساخ مؤدية الى إبطال وظائفها.

وقد تكون الجزيئات المشتقة من الأشرطة النقيضة هي موضعية Cis-sense وهي نسخ تنشأ من المنطقة الجينومية نفسها التي تضم الجين المستهدف الذي تعمل عليه ولكن تكون باتجاه معاكس وتستطيع ان تكون مزدوج منضبط مع أهدافها وهذه يمكن ان يعبر عنها في المناطق المظموغة لكروموسومات الفقرات وتلعب دوراً مهماً في تركيب الكروماتين. اما النوع الآخر فهو Trans-antisense وهي جزيئات قصيرة تنتسخ من مواقع مميزة عن أهدافها mRNA وتكون مزدوجات غير منضبطة مع المناطق المكمل لها في الجزيئات المستهدفة ومن أمثلتها miRNA و siRNA.

اما العمليات الحاصلة بعد الانتساخ فهي الأخرى تشكل مجالا واسعا للجزيئات غير المشفرة لممارسة وظائفها. فالخطوات الخاصة بهذا الموضوع ذكرت في مكان آخر، ولكن بالإجمال تشمل عمليات الخياطة وإنتاج نسخ ناضجة للتصدير الى الساييتوبلازم. وعملية ورود الجزيئات mRNA الى الساييتوبلازم يمكن ان تتعرض الى التنشيط او التثبيط او التفكك اي تكون ثبوتيتها غير مضمونة وكل هذه تقع تحت السيطرة في مرحلة ما بعد الانتساخ . وعلى العموم فان كل جزيئة من RNA تصدر من النواة الى الساييتوبلازم يجب ان تمر خلال الغشاء النووي عن طريق معقد الثقب النووي Nuclear pore complexes (NPC) وهنا تختلف المسارات التي تصدر بها جزيئات RNA المختلفة ، وفي العادة تصدر بواسطة المصدرات Exportins (Karyopherins) ، وجزيئات mRNA تصدر بالاعتماد على مجموعة من البروتينات التي ترتبط بها ، و NPC يميز ويصدر فقط جزيئات mRNA المعالجة تماما والكاملة النضج وقد يكون هذا التمييز معتمدا على وجود الغطاء وذيل الأدينين التي ترتبط اليها بروتينات التصدير.

والجزيئات المصدرة من النواة تذهب الى مواقع محددة بتوجه من تواليات محددة في 3'UTRs و 5'UTRs وغيرها من العوامل البروتينية وجزيئات ncRNA الصغيرة التي يمكن ان تمنع جزيئات mRNA المصدرة وهذا يتفق مع حقيقة ان ليس كل جزيئات mRNA يمكن ترجمتها وبذا يكون عدم التكافؤ واضحا بين مكنون النسخ وما متوقع منه من المكنون البروتيني. وهناك العديد من جزيئات mRNA غير المترجمة تتجمع لتكون مجمع mRNPs مكونة ما يسمى بـ P bodies هذا التجمع يتداخل مع آليات تفكيك هذه المجمعات ومعالجتها ، ومصير الجزيئات المتجمعة يكون مختلفا فهي اما يزال منها الغطاء ليسهل تفكيكها او تعود بعد حصول بعض التغيرات الى مسار الترجمة ، او تبقى في هذه الأجسام وبالرغم من عدم وضوح جوانب هذه الأجسام الا انه من المسلم به انها تؤثر في التعبير الجيني.

ومن الآليات الأخرى التي تساهم فيها الجزيئات غير المشفرة هو إحباطها للتعبير الجيني بتداخلها مع الكروماتين كما هو الحال في تجميع الكروماتين المتباين في الفطريات والنباتات. كما انها يمكن ان ترتبط الى جزيئات RNA أخرى تساعد في فلقتها وكذلك يمكن ان تكبح عملية ترجمة المشفر منها (كما ذكر أعلاه) وهنا يمكن ان تكون بمثابة وسائل دفاعية بالنسبة للنباتات والحيوانات ضد الإصابات الفيروسية لعملها ضمن جهاز المناعة المعتمد على RNA (RNA-based immune system). كما ان هناك آلية الإسكات باستراتيجيات الأشرطة النقيضة للـ mRNA والمستعملة بكثرة في الآونة الأخيرة ، اذ يتم إدخال قطع من DNA او RNA التي تكون مكملة للـ mRNA المشفر للبروتين وتنهجن معه مكونة مزدوجات غير مألوفة في سايتوبلازم الخلايا وبذا تكون عرضة للتدمير من قبل إنزيمات الحالة للـ RNA .

أنواع الجزيئات غير المشفرة

أوضح مما سبق ان الجزيئات غير المشفرة لها أصول مختلفة وآليات عمل مختلفة كما ان هناك بعض الالتباس في التسميات وربما كان ذلك لاختلاف الأزمان المكتشفة فيها فضلا عن عدم وجود نظام لتسميتها واختلاف الأحياء التي اكتشفت فيها (جدول 6.3) ، ولكن هناك توافق على بعض الجزيئات ويشمل هذا الجزء بعض الجزيئات الواضحة المعالم والمتفق عليها.

Micro RNA (miRNA)

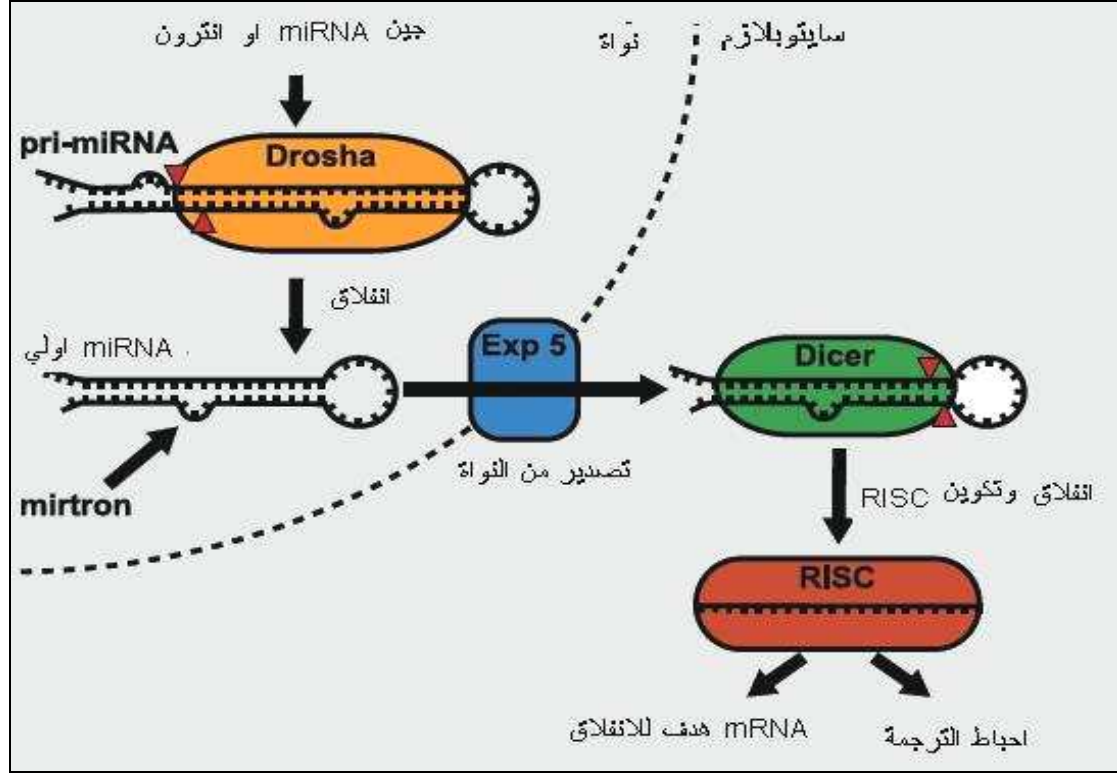
جزيئات الناضج منها تتراوح أطواله بين 20-22 نيوكلووتيد اكتشفت في عام 1993 ووجد ان هناك المئات منها في الحيوانات والنباتات، وتشتق من الانترونات وبعض الأحيان من الاكسونات او UTRs وتخلق بواسطة الإنزيم RNAP II ما عدا بعض جزيئاتها الموجودة في Alu sequences في الإنسان التي يتم انتساخها بواسطة الإنزيم RNAP III. وهناك حوالي 1000 جين لهذه الجزيئات في الجينوم البشري . وتداخل هذه الجزيئات يطلق عليه في العموم RNA interference (RNAi)

يتم تخليق جزيئات miRNA كنسخ أولية Pre-miRNA المكونة من حوالي 70 نيوكلووتيد مكونة تركيب يشبه ماشة الشعر يتم فلقتها بواسطة RNA III Exonucleases الموجود في الأحياء الراقية الذي يوجد منه نوعان Drosha المتخصص بفلق النسخ الطويلة مثل stem-loop الموجودة في النسخ الأولية للـ miRNA الحاوية تراكيب مزدوجة تصدر بواسطة المصدرات من النواة ثم تعالج بواسطة RNase III المسمى Dicer الذي يميز المناطق

المزدوجة في الجزيئات الأولية ويقطع الشريطين وبذا يزال تركيب ماشة الشعر المزدوج لتكون النتيجة النهائية قطع 20-22 نيوكلوئيد ترتبط الى أحد بروتينات Argonaut مكونة المعقد البروتيني RNA-induced silencing complex (RISC) اما الباقي من الجزيئة فيفكك. ويقوم المعقد الأخير بمسح جزيئات mRNA للبحث عن ما يكمله من القواعد النتروجينية ، والقواعد التي يبحث عنها المعقد و miRNA والموجودة في 3'UTRs للـ mRNA تشمل مكررات من 7-8 من التواليات ، وهناك مناطق أخرى مكونة من 6 قواعد يمكن ان تشارك في الارتباط ، والذي لا يشترط فيها التكامل التام ولكن عندما يكون هناك عدم تكامل تام في النهاية 5' تحتاج العملية الى تكامل تام على الطرف الثاني 3' وتساهم الطاقة الحرة في ثبوت المزدوجات المكونة من mRNA: miRNA ويرتبط المعقد بها ويبدأ بتفكيكها نتيجة الفعاليات الإنزيمية له المتمثلة ببعض بروتينات Argonaut (AGO) وبمساعدة ايونات المغنيسيوم وقد يكون ذلك بفلق ذيل الأدنين. ان جزيئات miRNA المحملة في المعقد RISC تحور التعبير الجيني بتخفيض معدل الترجمة وذلك اما بفلق جزيئات mRNA او تثبيط الترجمة. والحالة الأولى (الفلق) تحتاج الى التطابق التام اما عند عدم حصول الأخير فتعمل الآلية الثانية وهي كبح الترجمة والتي تمارس فعاليتها اما بإحباط عملية بدء الترجمة او إحباط عملية الاستطالة للبتيدات ولو انه في الآونة الأخيرة أمكن إثبات ان miRNA يمكن ان تساعد في ترجمة بعض mRNA المستهدفة. ومن الجدير بالذكر ان جزيئات miRNA يمكن ان تستهدف مئات من جزيئات mRNA اي تكون شبكة معقدة من التنظيم ولكن هناك ارتباط وثيق بين هذه الشبكات التي تؤثر فيها الجزيئة الواحدة. ومجمل العمليات موضحة في الشكل 91.3 .

وتتعرض هذه الجزيئات للتثبيط خلال مساراتها بواسطة الفيروسات التي تنتج بعض البروتينات لتعزيز الإصابة، فضلا عن ان بعض الفيروسات تنتج ايضا miRNA.

ان إنتاج نسخ miRNA من انترونات الجينات المضيفة يشير الى التعبير المترافق مما يشير بدوره الى وجود علاقة وظيفية وثيقة بين الجين المضيف و miRNA الموجودة في انتروناته. ويكشف عن هذه الجزيئات بطريقة وصمة نورذن Northern blot التي تحور لزيادة حساسيتها، وتستعمل المصفوفات الدقيقة لتحديد نسق التعبير عنها.



شكل 91.3 : ملخص لمسار فعالية جزيئات miRNA في إحباط فعالية mRNA

Small interfering RNAs (siRNAs) أو Short interfering RNAs

اكتشفت الجزيئات كجزء من عمليات إسكات الجين التي تحصل بعد الانتساخ-Post transcriptional gene silencing (PIGS) في النباتات وكانت مسؤولة عن إسكات الجينات عند تحوير النباتات وراثياً وما يساعدها في ذلك انها تنتشر بسهولة بين خلايا النبات من خلال Plasmodesmata (الممرات الموجودة بين الخلايا النباتية عبر الجدران الخلوية)، وتستطيع الانتشار الى مسافات أبعد عن طريق الأنسجة الوعائية لذلك تعد من العناصر المتحركة ولكن ليس بحرية تامة فضلا عن انها لا تتجمع في المناطق البعيدة عن إنتاجه ، تتكون من 20-25 نيوكلوไทيد بشكل أشرطة مزدوجة وتحمل اثنين من النيوكلوไทيد المفرد على طرف 3' وذلك ناتجا من فعل إنزيم القطع Dicer (المذكور آنفا) عند تأثيره في أشرطة RNA المزدوجة. وأغلب هذه الجزيئات التي تم تحديد تواليها كانت تحوي معدلات متوسطة من GC (53-55%) وهذا ما يؤدي الى عدم ثبوتها، وقد حددت القواعد النتروجينية في بعض المواقع لتكون أساسا في وضع برامج الحاسوب الخاصة للتعرف عليها.

وتحمل جزيئات siRNA في RISC لتشارك miRNA في بعض فعاليتها وتعتمد عملية التحميل على تناظر الجزيئات وثبوت الطرف 5' منها وهي الأخرى كانت من الأسس المستعملة في وضع البرامج الخاصة بالتعرف عليها فضلا عن حساب الطاقة الحرة على طرفي الجزيئة. وأشارت عمليات تحديد توالي هذه الجزيئات الى ان الطرف 5' يحوي على اليوراسيل (فضلا

عن تحديد مواقع أخرى). المهم ان وجود اليوراسيل في الموقع الأول يسهل عملية تحميل هذه الجزيئات في RISC فضلا عن احتوائها على نسب عالية من AT يسهل تحميلها ولكنها تقلل من الثبوت .

وتوجد أنواع من هذه الجزيئات اعتمادا على أصول اشتقاقها منها Repeat-associated siRNA (rasiRNA) وتنشأ من المكررات الموجودة في الجينوم مثل العناصر القافزة، وهذه توجد في كل من الأشرطة الحساسة وغير الحساسة للعناصر القافزة. وتشارك في إحباط العناصر القافزة وجزيئات RNA التي تتكامل معها في القواعد وتشارك في عمليات الإسكات على مستوى بدء الانتساخ وكذلك عند الانتساخ من خلال تحويل الكروماتين وبصورة خاصة في النباتات.

اما trans-acting siRNA (ta-siRNA) فهي صنف من siRNA التي تستهدف إسكات جينات مختلفة وليس الجين الذي ينتجها وتختلف عن cis-acting siRNA ، وتستحث هذه الجزيئات في النباتات تحت ظروف الإجهاد مثل زيادة الملوحة وبدا تعد كمنظمات ثانوية للتعبير الجيني في النباتات.

وهناك نوع آخر مشتق من الفيروسات Virus-derived siRNA ينتج عن الإصابة بالفيروسات ووجد ان النباتات الخالية من ظاهرة الإسكات الجيني الذي يحصل بعد الانتساخ (PTGs) او يكون فيها هذا النظام معطوب تكون حساسة جدا للإصابة بالفيروسات.

وتمارس siRNA فعاليتها بفلج جزيئات mRNA التي ترتبط اليها بعد إيجاد مكملاتها من القواعد بفعالية إنزيمية اذ تكون في الغالب محملة في RISC لذلك فهي تعد أحد مسارات تداخل RNAi المؤثرة في التعبير الجيني وقد أنتجت منها تواليات تركيبة لاستعمالها لأغراض معينة في اللبائن اذ تقوم بالإطاحة بالجين Knocking down الذي يحمل تواليات مكملتها لها والتركيبية تدخل عن طريق الإصابة Transfection ولكن عمليات الإطاحة بالنسبة للتركيبية منها تكون عابرة، ولهذه الأغراض وغيرها حظيت مثل هذه الجزيئات بالاهتمام مما أدى الى تطوير المصفوفات الدقيقة للكشف عنها.

(Short hairpin RNA) shRNA

جزيئات تشبه siRNA الى حد كبير وهي مزدوجة ولكن أحد طرفيها يكون بشكل تركيب ماشة الشعر ، تنتج على نطاق تجاري للأغراض العلاجية والمشكلة انه ليست كل الخلايا يمكن ان تتقبل هذه الجزيئات بمعدلات عالية والمحاولات جارية من قبل بعض الشركات لإنتاج جزيئات ذات صلاحية طويلة وكفاءة أعلى للأغراض العلاجية .

(Small nucleolar RNA) snoRNAs

وهي الجزيئات التي تشق من الجينات الموجودة في النوية .

(Piwi-interfering RNAs) piRNAs

جزيئات صغيرة بطول 25-29 نيوكلو تيد تعمل في إسكات الجينات وترتبط بعوائل خاصة من البروتينات ويتم إنتاجها في الخلايا الجنسية والملاحظ انها أطول قليلا من miRNA و siRNA . تتجمع جيناتها بشكل عناقيد في مناطق محددة من الجينوم وميزتها انه ينتسخ شريط واحد منها والدراسات حولها مستمرة.

كما سجلت بعض الجزيئات المشتقة من الجينات الكاذبة التي تقوم عمليات التعبير عن مثيلاتها من الجينات المشفرة للبروتينات ويمكن للقائمة الخاصة بالجزيئات غير المشفرة ان تطول ويساعد في ذلك تطور التقنيات الحديثة والحساسة التي تساعد في الكشف عنها.

مجمّل لوظائف الجزيئات غير المشفرة

أعداد الجزيئات المكتشفة في ازدياد وفي جينوم الإنسان عددها غير معروف ولكن تشير دراسات مكنون النسخ والمعلوماتية الحيوية الى وجود الآلاف منها، ونظرا لعدم توفر المعلومات حول وظائف العديد منها فالبعض قد تكون غير فعالة . ولكن ما متوفر من المعلومات يشير الى بعض الوظائف التي يمكن إجمال بعضها في التالي فضلا عن ما ذكر في مواضعه الخاصة فان تأثيراتها قد يكون متباينة ومنها:

- تساعد في السيطرة على حركة الجينات القافزة أثناء مراحل انتساخها وبعد الانتساخ.
- تساهم في ثبوت بعض جزيئات RNA.
- بعض الجزيئات مثل miRNA وجد ان لها تأثيرا "مضادا" للفيروسات، وهذا التأثير مسجل في النباتات بشكل أوثق اذ تساهم الجزيئات الصغيرة في إجهاض الإصابات الفيروسية بالمناعة المعتمدة على RNA.
- تساعد بعض الجزيئات مثل snRNA في أثناء خياطة RNA وإزالة الانترونات وكذلك تساهم في الخياطة البديلة، فضلا عن مساعدة بعضها في تحويل كفاءة الانتساخ والترجمة.
- المساعدة في عمليات الاستجابات الطبيعية لعوامل الإجهاد الحيوي واللاحيوي خاصة الاستجابات السريعة وذلك لان مسارات تخليقها قصيرة وليست كما هو الحال مع البروتينات مما يوفر الطاقة للخلايا.
- تعد أساسية في ثبوت الجينوم.
- تساعد في تطور الجينومات.
- يمكن ان يضخم دورها في الإسكات كما في مهاجمة miRNA للـ mRNA وتفكيكها واستعمال النواتج كقوالب لتخليق أشرطة مزدوجة تستعمل في إسكات جينات أخرى كما يحصل في النباتات بشكل رئيس.
- تساعد في نمو الخلايا او أخذها الى مسار الاستماتة عند عدم ملائمة الظروف.
- نظرا لقابليتها في إسكات الجينات كانت وسيلة ملائمة لإحباط جينات السرطان. وتستطيع بعض الجزيئات الإطاحة بالجينات بشكل متخصص ومنتخب في خطوط الخلايا او الموديلات الحيوانية وهناك العديد من جينات السرطان او الجينات الكابحة للأورام يمكن ان تخفض فعاليتها او تنظم بـ siRNA ولكن هناك عدة مساوئ لتدخلات RNA لا زالت تحتاج الى جهود للتغلب عليها ، فجزيئات siRNA المخلقة كيميائيا تكون قصيرة العمر وتؤدي الى الإحباط الجيني العابر، كما ان هناك صعوبة في إيصال الجزيئات الى أهدافها. ولذلك فالمحاولات جرت وتجري لإدخال siRNA الى جسم الإنسان بطرق ذات انتخابية عالية وبكفاءة عالية.
- تؤثر في تركيب الكروماتين وبالتالي التعبير الجيني وتشارك في الانقسام الخلوي والاختزالي وجوانب أخرى من الدائميكيات الكروموسومية.

- بعض جزيئات siRNA يمكن ان تحفز الاستجابة المناعية بشكل يعتمد على الخلايا.
- الحماية او الدفاع مثل piRNAs تكون معقدات وتقوم بإسكات القافزات الارتدادية من العناصر الوراثية في الخلايا الجنسية خاصة تلك التي لها علاقة بتخليق النطف. وجزئية CRISPR التي توجد في العديد من البكتيريا والاركية وتوجد في مناطق متكررة تفصل بينها فواصل متساوية الطول ويعتقد انها تشتق من العاثيات وتقوم بحماية الخلايا من الإصابة.
- تساهم العناصر الموجودة منها في نهايات الكروموسوم بالحفاظ على تركيبه، كما ان بعضها يقوم بتنشيط كروموسوم الجنس X .
- عدم توازن الجزيئات غير المشفرة من حيث المحتوى او حدوث تغيرات فيها نتيجة الطفرات يؤدي الى حدوث عدد من الأمراض مثل متلازمة Prader-Willi التي تتصف بنهم للأكل وقلة في التعلم وكذلك تساعد في ظاهرة التوحد Autism والزهايمر وفقدان السمع وغيرها.
- المشاركة في عملية التنظيم وهي أكثر الوظائف وضوحا ودراسة . وتقسم الآليات في ذلك اما عن بعد Trans-acting كما في حالة miRNA في الخلايا حقيقية النواة، والجزئية الواحدة يمكن ان تقلص التعبير الجيني لآلاف الجينات. ويكون في معظم الأحيان بالتكامل مع جزيئات mRNA في المناطق غير المشفرة وعادة تكون في 3'-UTRs والبعض الآخر من الجزيئات يتداخل مع الإنزيم RNAP او بعض عوامل الاستطالة. والبعض يعمل تحت ظروف الإجهاد مثل OxySRNA الذي يعمل تحت الإجهاد التأكسدي و B₂RNA العامل تحت إجهاد الحرارة. وهناك العديد من الجزيئات التي تعمل عن بعد وبآليات مختلفة (وقد ذكرت بعض التفاصيل الخاصة بهذه الجزيئات آنفا) .
- اما المجموعة الثانية فتكون عاملة بشكل Cis-acting، وتحدث هذه من قبل الجزيئات المضمورة في mRNA التي ستترجم الى بروتينات وتكون مضمورة في مناطق 5'-UTRs وتؤثر في التعبير بطرق مختلفة ، مثل Riboswitch الذي يمكن ان يرتبط بقطعة هدف صغيرة في الجين المستهدف مؤثرا في فعاليته. وهناك العديد من تراكيب RNA المضمورة في جزيئات أخرى تؤثر في عملية التنظيم بطرق مختلفة.
- اضطرابات الجزيئات غير المشفرة**
- سجلت العديد من الجزيئات غير المشفرة ذات العلاقة بالأمراض وفي مقدمتها سرطانات بعض الأعضاء وكذلك بعض الأمراض النفسية مثل التوحد Autism وانفصام الشخصية Schizophrenia وغيرها. وقد وجد ان بعض الطفرات المسببة للأمراض ليس لها علاقة بالطفرات الحاصلة في الاكسونات مما يشير الى ان الطفرات المسببة للأمراض تقع في المناطق المجاورة، فحصول طفرة واحدة في الانترون تؤدي الى توليد بعض الأمراض المسجلة في الحيوانات. ومن الملاحظات وجد ان عدم انتظام فعاليات miRNA يؤدي الى توليد السرطانات ولكن في بعض الحالات الخاصة يمكن ان تحور تطور السرطانات وتعمل بمثابة جينات كابحة للأورام .

الجزيئات غير المشفرة في الخلايا بدائية النواة

ازداد الاهتمام بمثل هذه الجزيئات في الخلايا بدائية النواة، وتقوم sRNAs (Small RNAs) بتنظيم التعبير الجيني عند مستوى بعد الانتساخ بآلية الازدواج مع جزيئات mRNA وبعض

الأحيان توفر فرصة لتنشيط الترجمة. وقد وجد باستعمال الدراسات الكمية ان التنظيم بالجزئيات الصغيرة هو الأفضل عند الاستجابات السريعة للإشارات الخارجية. وأشارت الدراسات الى ان عمر المعقد sRNA-mRNA ونسبة معدل إنتاجه يحدد بمستوى البروتينات المستهدفة مما يشير الى توفير sRNA عمليات تعبير جيني دقيق، والجزئيات في حالة بدائية النواة تتشابه بالأطر العامة مع حقيقية النواة ولكنها تختلف في التفاصيل وآلية التأثير.

وقد سجل وجود أكثر من 50-80 نوع من الجزئيات الصغيرة في *E. coli* أطوالها تتراوح بين 50-400 نيوكلويتيد وأكثرها ثابتة من حيث الطول أثناء التطور والوظائف التي تؤديها ويكون تدخلها مع mRNA بمساعدة الوصيف 9 Hf .

وفضلا عن آلية ازدواج القواعد التي تمارسها للتنظيم بعد مرحلة الانتساخ فانها يمكن ان تمارس دورها التنظيمي عند الانتساخ وذلك بالتداخل مع البروتينات العاملة وكذلك DNA، ولذلك فهي تكون ضمن شبكات التنظيم.

وعمليات التنظيم بالجزئيات غير المشفرة من RNA يلاءم الاستجابات السريعة في بدائية النواة مثل البكتيريا وكذلك الرجوع عن حالة الاستجابة بسرعة. وتفضل الاستجابة باستعمال sRNA خاصة عند الاجهادات العابرة لان إنتاجها أسرع من إنتاج البروتين.

مشاركة البروتينات في التنظيم

تشارك البروتينات بشكل فاعل في تنظيم الفعاليات الحيوية وذلك بكونها إنزيمات او عوامل انتساخ او كونها بروتينات تركيبية. فبصفتها كإنزيمات يلاحظ ان العديد من جزيئات RNA المنظمة مثل miRNA تحتاج الى RNAP في بعض الحالات.

اما من وجه كونها مكونات تركيبية فأفضل الأمثلة عليها هو الكروماتين المكون من ارتباط الهستونات وبروتينات لاهستونية مع DNA، ومن الأمور المهمة في الانتساخ هو حصول تحويرات على الهستونات ومن ورائها الجسيمات النووية Nucleosomes، وتحور الهستونات بأكثر من طريقة وتؤثر سلبا او إيجابا في التعبير الجيني اعتمادا على الثمالات من الحوامض الامينية المحورة ومن التحويرات:

- أستلة الهستونات بواسطة إنزيم أستلة الهستونات وتؤدي العملية الى تغير هيئة الكروماتين وفي الغالب يوفر فرصة اكبر لوصول RNAP وعوامل الانتساخ الى منطقة الممهد من DNA.
- مثيلة الهستونات وتتم بإنزيمات المثيلة Methylases وفي العادة تؤدي الى التقليل من الانتساخ بعدة طرق، ونمط المثيلة يكون متوازنا بعد انقسام الخلايا لذلك يكون لها أثرا مهما في تمايز وتخصص الخلايا، فضلا عن ان عملية المثيلة تمثل هذه الوضعية يكون لها دور كبير في الذاكرة اللاجينية Epigenetic memory.
- حدوث تغيرات في تركيب الكروماتين يمكن ان تسهل انتساخ الجينات.

ومن البروتينات المشاركة هي عوامل الانتساخ التي تستطيع الارتباط الى DNA او يمكن ان تعمل كمعامل مرافقة لأنزيم الكوثرية ومعقداته DNA .

وتوجد عوامل انتساخ عامة GTF (المذكورة سابقا) وهي التي يحتاجها الإنزيم RNAP II للارتباط بـ DNA وعملية البدء وتخليق RNA وليس كل عوامل الانتساخ العامة قادرة على الارتباط الى DNA ولكنها تدخل بمثابة أجزاء مهمة من المعقدات البروتينية الكبيرة التي تتداخل مع DNA ، والتي ترتبط فيها بصورة مباشرة عند توالي الممهد أو Cis-regulatory elements والأخيرة تواليات من DNA الحاوية على مواقع لارتباط عوامل الانتساخ.

والجزيئات البروتينية القادرة للارتباط بـ DNA تؤثر في عمليات الانتساخ بعدة طرق فهي يمكن ان تغلق أو تنشط الارتباط وذلك بمشاركتها في عمليات الأسئلة أو إزالتها، أو تزود الإنزيم بالمنشطات المرافقة Co-activator أو الكابحات المرافقة. فالتعبير في حقيقيات النواة مثل اللبائن يحتاج الى عوامل انتساخ مختلفة وهي بدورها تحتاج الى بروتينات لتنظيمها وبالنتيجة تكون الشبكة المسيطرة معقدة جدا ويكون هناك تنوع في المكان والزمان التي تظهر فيه أنماط مختلفة من الانتساخ، وتعجز بعض الأحيان برامج الحاسوب عن التنبؤ بها وتحتاج الى عمليات رياضية وإحصائية معقدة لتفسيرها.

ومن البروتينات الأخرى العاملة في مجال التنظيم هي عوائل Argonautes (Ago proteins) توجد بعوائل مختلفة وتشارك مع RNA في عمليات التنظيم وقد درست في النباتات اذ تكون اللاعب الرئيس في المسارات التنظيمية التي تشارك فيها جزيئات RNA الصغيرة والموديل الدراسي النباتي *Arabidopsis* (رشاد الصخر) يحوي على حوالي 10 عوائل من هذه البروتينات.

الوراثة اللاجينية Epigenetics

حقق دراسة يتناول التغييرات الكيماوية لكل من المواد الوراثية DNA والهستونات التي تضفي مرونة على الجينومات ومشاركة RNA التي تؤثر في الأنماط المظهرية دون المساس بتواليات DNA. فالجينات يعبر عنها بعد اختيارها وارتباط عوامل الانتساخ بها وفقاً لما تمليه الظروف الخارجية التي تحرك الشبكات الخلوية للاستجابة واستعمال تواليات معينة من الجينوم. وارتباط عوامل الانتساخ الى توالي معين يؤدي الى سلسلة من الأحداث لا تتعلق بتسلسل القواعد النتروجينية ولكن يتعلق مثلاً بتركيب الكروماتين الذي يساعد في تجميع معقدات الانتساخ، وهذه لا يمكن ان تقوم بعملية الانتساخ لوحدها فضلا عن ان عمليات الانتساخ يمكن إعادة برمجتها ببعض التحويرات التي تبقى حتى عند نقل هذه التواليات المحورة الى سايتوبلازم آخر وهو ما يسمى بذاكرة الخلية . ويمكن لهذه الجوانب (أي غير المتعلقة بتواليات DNA) ان تشارك في إحداث السرطانات والعديد من الأمراض الوراثية مثل Angelman syndrome و Prader-Willi syndrome وغيرها فضلا عن مشاركتها في تطور الجينومات ومن هنا يستنتج (ما هو معروف سابقا) ان:

$$\text{النمط المظهري} = \text{النمط الجيني} + \text{مؤثرات البيئة}$$

وهذا يعني ان DNA وتوالياته لا يعني الكل المطلق وانما تساهم بجزء قد يصل الى 50%. والمواضيع الوراثية غير المتعلقة بالتواليات تنطوي تحت الوراثة اللاجينية او الخارجية وما تضمه يسمى الجينوم الخارجي Epigenome ومقابلته الحقل الذي يهتم بدراسته Epigenomics وقد كان لتحديد توالي الجينومات الأثر الكبير في تطور الدراسات في هذا

الحقل. وتتضح معالم مجال الدراسة هذا من حقيقة انه مثلا كل خلايا الإنسان تحمل بصمة وراثية واحدة ولكن يلاحظ ان الخلايا العصبية تختلف عن خلايا الجلد او اي نسيج او عضو آخر.

ويظهر تأثير الوراثة اللاجينية في المراحل المختلفة من الحياة وكذلك الاختلاف في الجنس والأنماط المظهرية التي تظهر بتأثير ظروف المناخ والغذاء وغيرها. وتشير الدراسات الى ان تواليات DNA في البشر متشابه الى حد 99.9% وان الاختلافات 0.1% تعزى الى تغيرات النيوكلو تيد المفرد SNPs المارة الذكر فضلا عن مساهمة الوراثة اللاجينية في ذلك. ومن هنا يلاحظ ان الوراثة اللاجينية تدمج كل اللغات الكيماوية التي يستعملها الجينوم والبيئة للتواصل وبالتالي فهم التطور ودراسة الحالات الطبيعية من الحالات المرضية.

وتشمل الدراسة اللاجينية عدة حقول منها تحويلات الكروماتين وداينميكية الجسيمات النووية Nucleosome dynamics وجزيئات RNA غير المشفرة التنظيمية، وعملية إسكات الجينات كما في تثبيط الكروموسوم الجنسي X-chromosome inactivation وكذلك انزلاق الهستونات من الجسيم النووي بواسطة بعض العوامل البروتينية مثل SWI/SNF الموجودة في الخمائر وبعض الحشرات والنباتات والإنسان التي تؤدي الى انزلاق الجسيم النووي وإبعاد الهستونات المركزية والإحلال مكان المزدوجات H₂A-H₂B لإعادة تشكيل الكروماتين والسماح بانتساخه ثم العودة الى الحالة الطبيعية.

وطمغ الجينوم Genome imprinting والذاكرة الانتساخية Transcriptional memory، وكل هذه المعلومات تجمع وتوضع في خرائط لاجينية تساعد في التمييز بين الحالات الطبيعية (الصحية) والحالات غير الطبيعية (الأمراض). ولأهمية هذا المجال بدأ وضع مشروع الجينوم البشري اللاجيني Human epigenome project (HEP) عام 1999 حتى قبل اكتمال مشروع الجينوم البشري Human genome project (HGP). وأوجدت للموضوع قواعد بيانات والإمكانات الكبيرة الخاصة بالإنسان والحيوانات المستعملة كموديلات دراسية للمساعدة في البحث والتقصي للوصول الى أفضل الحالات ، وظهر العلاج اللاجيني Epigenetic therapy على غرار العلاج الجيني والتي تتناول السرطانات بشكل خاص باعتبار بعض منها أمراض لاجينية، كما ان تأثيرات الوراثة اللاجينية او الخارجية يمتد ليطول الأمراض النفسية مثل التوحد وانفصام الشخصية (التي مر مناقشتها سابقا) وكذلك بعض أمراض المناعة الذاتية مثل SEL .

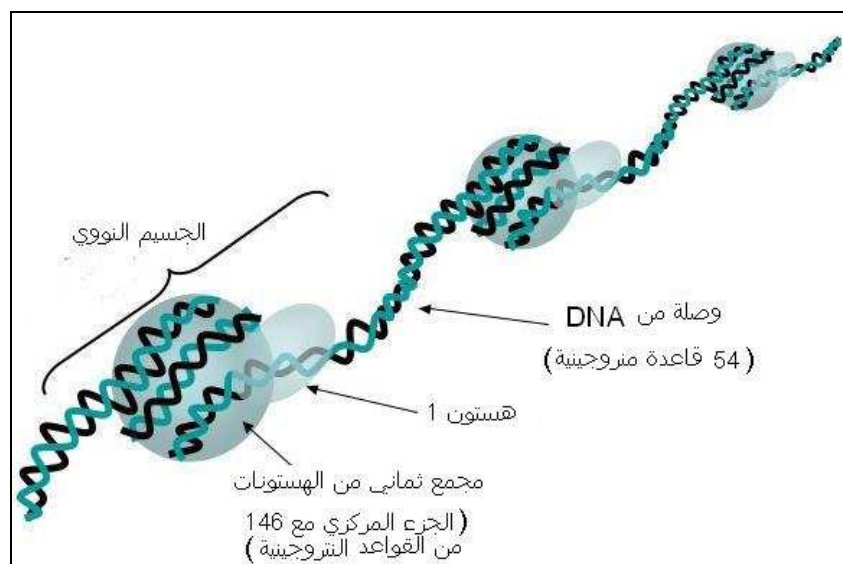
تركيب الكروموسوم في الخلايا حقيقية النواة

مر مناقشة بعض هذه الجوانب فيما سبق مثل تواليات DNA وتوزيعها ولكن المهم في هذا المجال ليس DNA وانما متعلقاته مثل الهستونات والبروتينات من غير الهستونات . فالكروماتين يتكون من أشرطة DNA ملتفة على مركز من الهستونات ، والارتباط الوثيق بين المادة الوراثية DNA والهستونات يجعل الأخيرة اللاعب الرئيس في الوراثة اللاجينية.

الهستونات Histones

بروتينات صغيرة تتراوح أوزانها الجزيئية بين 11-22 كيلو دالتون تساعد في التفاف أشرطة DNA التي تكون بشكل حلزونات مزدوجة ، المركز مكون من الهستونات H₂A ، H₂B ،

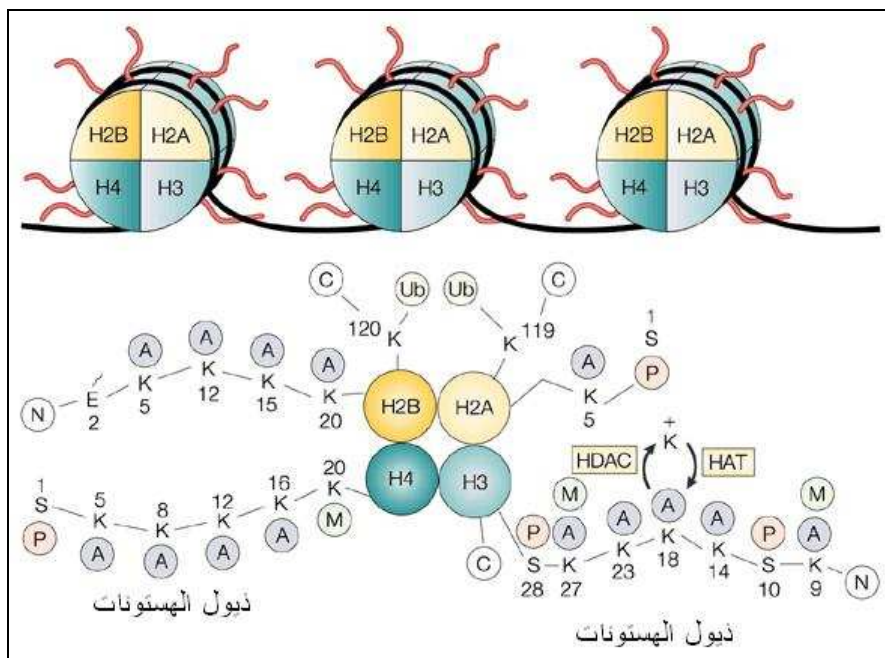
H_4 ، H_3 بشكل تجمع ثماني الوحدات اي اثنين من كل هستون مكونة الجسيم النووي اما H_1 فيربط الجسيمات النووية مع بعضها كما توضح في الشكل 92.3:



شكل 92.3 : تركيب الكروماتين في الخلايا حقيقية النواة

ويتم تثبيت الجسيمات النووية المكونة من مركز الهستونات وحوالي 147 قاعدة من DNA من الأشرطة المزدوجة الملففة حولها بتداخلات كهربائية مستقرة بين مجاميع الفوسفات السالبة الشحنة للـ DNA والمجاميع الموجبة لمجموعة ϵ -amino group لثمالة اللايسين وكذلك مجموعة Guanidino group لثمالة الأرجنين. وتتكون الهستونات من جزء كربوكسيلي يكون ضمن كرية الجسيم النووي وذيل الطرف الاميني الذي يبرز من سطح الجسيم النووي كما موضح في الشكل (93.3).

وفي هذه الذبول هناك ثمالات متعددة تكون عرضة للتحويل. وفضلا عن الهستونات هناك البروتينات اللاهستونية التي ترتبط بالمواد الوراثية ومن أفضل الأمثلة عليها Polycomb proteins و Trithiorax proteins التي تلعب دوراً مهماً في الوراثة اللاجينية.



شكل 93.3 : تركيب الجسيم النووي وذيل الهستونات

الآليات المستعملة في الوراثة اللاجينية

هناك عدة آليات تتم بواسطتها إظهار تأثيرات الوراثة اللاجينية ومنها:

- تداخلات RNA (RNAi) وقد مر ذكرها سابقا في موضوع تواليات RNA غير المشفرة ncRNA.
- المثيلة وتحدث هذه لكل من تواليات DNA والبروتينات المرتبطة به.
- الأسئلة وتحدث بشكل أساسي للهستونات
- إضافة البايوتين Biotinylation
- إضافة بروتين Ubiquitin (Ubiquitylation)
- Sumoylation إضافة بروتين صغير .

ولذلك فالمجال يهتم بالتحويلات التي تحصل للكروماتين التي تشمل التغيرات التركيبية مثل طوي بروتينات الكروماتين، وارتباطه بالأرضية النووية وطريقة تعبئة والتفاف DNA حول الهستونات التي تؤدي الى تكوين الكروماتين الحقيقي الذي يسمح لتواليات DNA بالانتساخ ، او تكوين الكروماتين المتباين الذي يكون في اغلب الأحيان ملازما لحالة منع انتساخ تواليات DNA. فضلا عن تحويل ذيل الهستونات التي تشمل المثيلة والاسئلة والفسفرة وغيرها.

مثيلة DNA: وتعد أكثر الآليات المدروسة في مجال الوراثة اللاجينية. وتضاف مجموعة المثل في اغلب الأحيان الى ذرة الكربون الخامسة من السايترز (وهذه القاعدة النروجينية تشكل بقع ساخنة للتحويل) في مزدوجات CpG او ما يسمى بجزر CpG وهي تواليات مكررة تتوزع على الجينوم مشكلة إحدى مظاهر تكرار التواليات في خلايا حقيقية النواة ، ففي الجينوم

البشري هناك حوالي 290,000 جزيرة وتقع أغلبها عند الطرف '5 في العديد من الجينات البشرية، وهذا يشير الى ان حوالي 60% من الجينات البشرية تحوي على هذه الجزر والأغلبية منها لا تحوي على المثل في اغلب أطوار الحياة. وتكثر في المناطق السابقة للمهدات اذ تشكل 40% من مهدات اللبائن.

وتختلف عملية المثيلة من حيث النمط اعتمادا على عدة ظروف مثل العمر، ففي جينوم الإنسان (الخلايا الجسمية) تصل نسبة السايتوزين الحاوي على مجموعة المثل الى حوالي 1% من القواعد الكلية للـ DNA وهذا يعني ان هناك حوالي 70-80% من الجزر الثنائية CpG حاوية على المثل على الذرة الخامسة من السايتوزين بغض النظر عن حالة التعبير عن الجينات . وهناك حالات من المثيلة منتشرة في DNA تسمى المثيلة الشاملة Global methylation وتحدث في بعض مراحل العمر وتكون مختلفة بين الأحياء ومختلفة في حالات الصحة والمرض مما يشير الى ان عملية المثيلة تشترك في وظائف أخرى غير التحكم في عمليات الانتساخ فهي تمثل إحدى الوسائل الدفاعية ضد الفيروسات الغازية او عمليات القفز للعناصر القافزة غير المرغوب فيها وكذلك تساعد في التمايز وتخصص الأنسجة وغيرها من المظاهر. وكما ذكر أنفاً فان المثيلة تختلف بين الأحياء ويكون ذلك بال نوعية او المستوى، فمثلا بعض الحشرات تكون مناطق المثيلة هي CpT بدلا من CpG الشائعة في اللبائن، وفي البكتريا تتم مثيلة الأدينين الواقعة ضمن تواليات محددة اما في الفطر *Neurospora* فتتم مثيلة الأدينين في Tpa. وقد سجلت أعلى حالات المثيلة في جينومات المملكة الحيوانية في الحيوان البحري Sea squirt وعلى النقيض فان الدودة (النموذج الدراسي) *Caenorhabditis elegans* لا توجد فيها سايتوزينات تحوي المثل ولا تمتلك الإنزيمات الخاصة بالمثيلة. فضلا عن ذلك وكما ذكر أعلاه ان المثيلة وأنماطها تختلف وفق المرحلة العمرية، فمثلا الأجنة المبكرة للفيران تكون درجة مثيلة جينوماتها تصل الى 70-80% وتهبط الى 30% في المراحل اللاحقة. وتختلف درجة المثيلة حتى بين أنواع الجنس الواحد.

آليات الإسكات بمثيلة DNA: تكون الجينات الممثلة بدرجة عالية Hypermethylation عادة ساكنة وهناك طريقتين لأداء هذه المهمة:

- تداخل مجموعة المثل في CpG للـ DNA مع البروتينات التي يفترض ان ترتبط الى جزر CpG وبالتالي تؤثر في عملية الانتساخ او مشاركة عناصر أخرى مثل المشجعات لأداء مهمتها.

مثيلة DNA تؤدي الى منع ارتباط عوامل الانتساخ الى مواقع التميز Recognition sites في المهدات لان معظم عوامل الانتساخ تتنافر مع مجموعة المثل.

وعلى العموم فان الإحباط في الانتساخ نتيجة المثيلة هي ليست حالة مطلقة فهناك بعض المناطق الحاوية على المثل تكون نشطة وذلك نتيجة لاستعمال بروتينات الحجب Masking proteins التي تغطي المواقع الحاوية على المثل اي ان هناك عملية إزالة مثل غير مباشرة Passive demethylation وبالتالي تفك هذه الجينات من الإسكات الذي يضيفه وجود مجاميع المثل.

فوائد عملية المثيلة: ان حالة الإسكات التي تضيفها حالة المثيلة تساعد الخلايا في الدفاع ضد DNA الغازي الذي يدخل الخلية وكذلك إحباط عمليات القفز . والملاحظ ان Alu sequence التي تشكل مجموعة التواليات المكررة الشائعة جدا في الجينوم البشري يكون القليل منها حوالي 1% فعال والأغلبية صامتة نتيجة المثيلة ولكن في حالات الإجهاد يمكن ان تنشط فعاليتها حتى بوجود المثل وبذلك يمكن ان تؤدي الى حث عدد من الأمراض مثل التلاسيميا وارتفاع كولسترول الدم وبعض أنواع الشلل وبعض أورام الجهاز العصبي. وتتدخل جزيئات RNA في نمط المثيلة مثل الأشرطة المزدوجة dsRNA او غيرها من الجزيئات غير المشفرة تؤدي الى زيادة المثيلة وإسكات الجينات.

ويعد نمط المثيلة من الواسمات المهمة والمميزة لبعض حالات السرطان ويمكن في بعض الحالات تصنيف السرطانات وفقا لدرجة مثيلة CpG. تغيير نمط المثيلة يمكن ان يستعمل في بعض الحالات ضمن العلاج اللاجيني فبعض الأدوية المثبطة للمثيلة تستعمل لعكس سكوت الجين المسبب عن المثيلة.

ويفيد دراسة نمط المثيلة في بعض الحالات فهناك ما يسمى Methylation variable positions (MVPs) وهي تشبه حالة تباير النيوكلوไทيد المفرد SNPs وتستعمل MVPs في اللبائن لتحديد مواقع CpG وهي تعكس فعالية الجين ونوعية النسيج والمرحلة العمرية فضلا عن استعمالها في تحديد بعض الحالات المرضية في الإنسان على المستوى الجزيئي. فالملاحظ ان فقدان او انخفاض المثيلة Hypomethylation واضطرابها يؤدي الى تشوهات الكروموسومات الذي يتبعه تشوهات مرضية في النباتات مثلا، وكذلك تؤدي هذه الحالة الى زيادة الحالات المرضية في الإنسان نتيجة لزيادة حالات القفز للعناصر المتحركة.

آليات مثيلة DNA

هناك أكثر من طريقة لإيجاد حالة المثيلة وكذلك وجود طرق للحفاظ عليها ، فعملية إيجاد المثيلة De novo methylation تتم بواسطة إنزيمات المثيلة، اما إزالة المثل Demethylation تتم بواسطة إنزيمات إزالة المثل Demethylases. الحالة الأولى تكون فعالة في المراحل الجنينية المبكرة ويمكن ان تحفز بشكل كبير عند إصابة الخلايا الجنينية بالفيروسات، وذلك لانها تكون في الحالات الطبيعية بطيئة.

اما عمليات الحفاظ على حالة المثيلة فهي تشمل عمليات اكتساب المثل وأخرى لمنع فقدان المثل. والحفاظ على المثيلة يعبر من جيل الى آخر، فعند تضاعف DNA شبه المحافظ فتكون الأشرطة الجديدة لا تحوي على المثل اي تحصل حالة شبه مثيلة او نصف مثيلة Hemimethylation فيقوم إنزيم المثيلة DnmtI بمثيلة CpG في الأشرطة الجديدة لتصبح مشابهة للأشرطة الأبوية للحفاظ على المعلومات والنمط اللاجيني وهذا ما يسمى بالشفرة اللاجينية او شفرة الوراثة اللاجينية Epigenetic code لذلك يطلق على الإنزيم DnmtI إنزيم الإدامة ويكثر في الخلايا الجسمية وتكون كفاءته في مثيلة الأشرطة شبه الممثلة أكثر من 10-40 مرة من الحالات الأخرى أي ان تكون له ألفة عالية لمثل هذه الأشرطة التي تكون حالة عابرة ولكنها تكون كارثية بالنسبة للخلايا ، وعملية الإدامة قد لا تصل الى مرحلة الكمال ويمكن ان تفشل ويصل فقدان حالة المثيلة للـ CpG الى مستوى 5% \ CpG \ انقسام خلوي

ويؤدي الى انخفاض نمط المثيلة بنسبة 4% لكل جيل من الخلايا وهذه تشارك فيها عوامل أخرى غير DnmtI عند غياب الأخير او انخفاض تركيزه. ومن الإنزيمات الأخرى العاملة في المثيلة Dnmt3a و Dnmt3b وهذه يكون التعبير عنها قوي جدا في المراحل الجنينية اذ تتم إضافة المثل لكل CpG بدون تميز في معظم الخلايا وفي أثناء هذه المرحلة تحصل برمجة لعملية المثيلة .

ان مناطق الجينوم لا تكون متاحة لإضافة المثل بشكل مطلق فمثلا Dnmt3b يكون ضروريا لإضافة المثل الى جزر CpG في حالة تثبيط كروموسوم الجنس X (كما سيأتي ذكره لاحقا) لذا فهو يمكن ان يضيف مجاميع المثل الى مناطق أصلا صامتة من الكروماتين مثل الكروماتين المتباين وتوجد بعض البروتينات التي تؤدي الى اضطراب الكروماتين حتى تمكن إنزيمات المثيلة للوصول الى DNA وهذه تكون مهمة في حالة الكروماتين المتباين والمناطق الصعبة الوصول.

وأكثر المناطق عرضة للمثيلة هي التواليات المكررة مثلما يحدث في الفطريات والنباتات وتساعد في إسكات الجينات التي تطولها وقد شكلت حالة إسكات الجينات المنقولة اليها Ttransgenes إحباطا" في عمليات تحسين النبات لذلك تؤخذ الاحتياطات لإبعاد الجينات المنقولة ضمن عمليات التحسين الوراثي عن المناطق الصامتة. اما حالة إزالة المثل فهي اما تكون بشكل داينميكي وبدون مشاركة الإنزيمات حيث تكسر الأصرة بين C-C لمجموعة المثل والبريميدين. اما الحالات الأخرى فتشارك فيها الإنزيمات المزيلة للمثل وتحصل عملية إزالة المثل في مراحل مختلفة سواء المراحل الجنينية او الخلايا العادية وترافق عمليات انخفاض المثيلة عدم ثبوت الجينوم وزيادة معدل الطفرات وظهور الكروموسومات بأشكال غريبة. وفي العموم تتبادل عملية المثيلة مع بروتينات PcGrx الأدوار في الحياة الخلوية ويمثلان ذاكرة الخلية التي تعتمد عليها.

وفي البكتريا تمثل مثيلة DNA أحد عمليات التنظيم التي تحدث بعد تضاعف DNA للسيطرة على التداخل بين DNA والبروتينات، وفيها تتم مثيلة الأدينين بدلا من السايتوزين كإشارات للوراثة اللاجينية . وتختلف المجاميع البكتيرية في المكررات التي تستهدف في المثيلة فمثلا α -Proteobacteria يكون التوالي GANTC (N تمثل أي نيوكلوئيد) هو المستهدف وتضاف بإنزيمات مثيلة خاصة وتنظم دورة حياة الخلية والأحداث التي تمر بها الخلايا ، فهي تنظم توقيت تضاعف DNA وعزل وتوزيع الكروموسومات الناتجة على الخلايا البنوية وعمليات الانتساخ وتكون في اغلب الأحياء ذات تأثير محبط .

والمثيلة تجعل الخلايا سريعة الاستجابة للظروف البيئية التي تتصف بها وحيدات الخلايا. وتساعد في تعبئة العائيات البكتيرية بعد تضاعفها، وتساعد في التقليل من عمليات قفز العناصر المتحركة وبالتالي تؤثر في التعبير الجيني وإصلاح DNA ونقل البلازميدات ، والعملية الأخيرة اي الاقتران تكون حساسة لاحتواء بعض المناطق من DNA على المثل. وفضلا عن ذلك فان بعض الممهدات تكون فعالة جدا عندما تكون شبه حاوية على المثل اي بعيد انتهاء التضاعف وقبل اكتمال إضافة المثل اليها.

وفي المجموعة البكتيرية δ -Proteobacteria تحدث المثيلة للأدنين ضمن التوالي GATC وهذه تقع في أماكن خاصة وعند ذلك تمنع ارتباط البروتينات الى DNA ويمكن ان تعبر الحالة من الخلايا الأم الى الخلايا الناتجة كما يحصل في تباير الطور Phase variation ، وفي اغلب السلالات من *E.coli* يحصل تغير في اللاصقات Adhesins وبالتالي تتغير الحالة بين ممرضة وغير ممرضة.

وإنزيمات المثيلة في البكتريا تسمى إنزيمات المثيلة اليتيمة Orphan methylase لانه تنقصها الإنزيمات القاطعة Restriction enzymes المرافقة. وتلعب إنزيمات المثيلة دورا مهما في أنظمة القطع والتحويل R/M systems فالإنزيمات القاطعة وهي من النوع الداخلية تميز تواليات معينة وتقطعها ما لم تكن حاوية على المثيل ، والأنظمة تستهدف الفيروسات الداخلة الى الخلايا، وفي بعض الأحيان تخادع الفيروسات وتتجنب أنظمة R/M . فالمعروف ان هناك حوالي 20,000 من تواليات GATC في جينوم *E. coli* قد تكون كلها حاوية على المثيل ويعتمد هذا على عدد من الظروف مثل حالة النمو والظروف البيئية المحيطة بالخلايا ووجود هذا العدد يعني انها تتكرر في جينوم *E. coli* بمعدل 232\1 ويشبه ما موجود في العاثي T₇ الذي يوجد فيه 256\1 وهذا يمثل وجود حوالي 141 موقع في جينوم العاثي الصغير ولكن الواقع العملي يشير الى وجود 6 مواقع فقط من GATC وكذلك الحال مع عاثيات أخرى اذ تكون القيم العملية اقل بكثير من القيم المحسوبة لجينوماتها وهذه الندرة في GATC في جينومات العاثيات تقلل من فرص هضمها بـ MutH endonuclease. وبعض العاثيات تحوي على مواقع GATC مشابه لتكرار هذه التواليات في الخلية المضيفة (يعني الواقع العملي) ولكنها تحوي على آليات تجعلها تقلت من أنظمة R/M.

تحويل الهستونات:

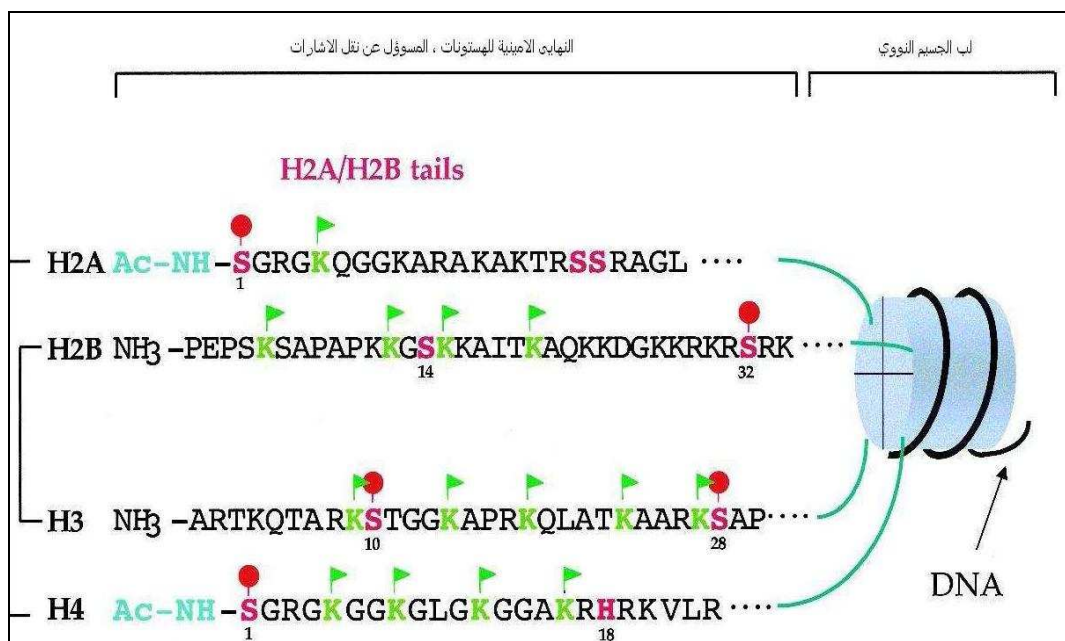
تمثل أحد أوجه التنظيم اللاوراثي اي غير المتعلق بتواليات DNA، فالهستونات يمكن ان تنشط الجينات او تسكتها وهذا يوضح لماذا تتمايز وتتخصص الخلايا.

أشرطة DNA تلتف حول الهستونات بشكل خاص حول مراكز الهستونات كما موضح في الشكل (93.3 ، 94.3) وعند تغير نمط الالتفاف فان ذلك يؤثر في التعبير الجيني. ولكن هناك تغيرات أخرى على المستوى الجزيئي يمكن ان تؤثر في فعالية الكروماتين وبالتالي التعبير عن الجينات وتوجد عدة مستويات للتحويل تعمل بطرق مختلفة وفقاً للموقع المحور وهذا يشير الى إمكانية قراءة التحويلات من قبل أنظمة الانتساخ ، ومن هذه المستويات:

- التحويلات بعد الترجمة ولعلها تكون الأهم ، اذ تحور الحوامض الامينية الداخلة في تركيب الهستونات التي تكون بشكل سلاسل طويلة، فاذا كانت الحوامض الامينية مشحونة فان يمكن ان يؤدي الى تغيير كرية الهستون وبالتالي لا تستطيع أشرطة DNA الالتفاف حولها بشكل صحيح أثناء عمليات التضاعف وهذا يمكن ان ينقل الى النسخ الجديدة حيث تعمل الهستونات المحورة كقوالب لتكوين هستونات جديدة التي تكون بشكلها الجديد المتغير عن الأصل، وهذا ما يحصل في الخلايا المتميزة Differentiated cells وتبقى بهذه الحالة ولا تعود الى حالة الخلايا الجذعية.

وأهم مناطق الهستونات التي تحور هي ذيول الهستونات وتشمل التحويرات المثيلة أو الأسئلة أو الفسفرة أو Ubiquitylation أو Sumoylation. وأهم التحويرات هي الأسئلة التي ترتبط بنشاط الجين وزيادة التعبير عنه وهي أكثر الحالات المدروسة ، فمثلا تضاف مجموعة الاستيل الى ثملات اللايسين K₉ و K₁₄ لذيل الهستون H3 بواسطة إنزيم أسئلة الهستونات HAT (كما موضحة في الشكل 94.3) الذي يؤدي الى زيادة كفاءة الانتساخ ، والسبب هو فيزيولوجي بطبيعته، فاللايسين له شحنة موجبة على ذرة النتروجين لذا يرتبط بمجموعة الفوسفات السالبة في DNA وبالتالي يمنع تنافهم ولكن ارتباط الاستيل الى اللايسين يحول الشحنة الموجبة الى حالة التعادل وبالتالي تكون هناك فرصة لوصول عوامل الانتساخ والإنزيمات الى DNA . ومن جهة ثانية فان الجسيم النووي يمكن ان يرتبط الى الجسيمات المجاورة مؤدية الى حدوث التكثيف ومنع وصول عوامل الانتساخ الى DNA ولكن إضافة مجموعة الاستيل الى اللايسين يمكن ان يمنع هذا الانضغاط. وعليه فان أسئلة اللايسين يمكن ان تساعد وتجلب او ترشد عددا من عوامل الانتساخ والإنزيمات المحورة ومكونات مكائن الانتساخ الى DNA للتعبير عنه. ويمكن ان تتأزر عملية الأسئلة مع آليات أخرى مؤدية الى تغيير تصرف الجسيم النووي وهذا بمجموعه يسمى شفرة الهستون .

- مثيلة الهستونات وهنا مرة أخرى تكون ذيول الهستونات هي المرشحة للتحوير كما موضح في الشكل 94.3 :



شكل 94.3 : الجسيم النووي ومواقع تحوير ذيول الهستونات
(موضحة بالأعلام)

وتتم عملية المثيلة بواسطة إنزيمات المثيلة وتزال بإنزيمات مقابلة فضلا عن مشاركة عدد من البروتينات، فمثيلة ثملات اللايسين رقم K_3 و K_4 و K_{79} للهستون H3 تكون مرافقة لنشاط الانتساخ في حين مثيلة ثملات اللايسين رقم K_9 و K_{27} للهستون H3 و K_{20} للهستون H4 تكون مرافقة لإخماد الانتساخ وهذه التغيرات يمكن ان تعبر للأجيال القادمة او تقتصر على حياة الخلية الواحدة، ولذا يلاحظ ان معطيات المثيلة تعتمد على الثملات المعرضة للتحويل. وعلاوة على ذلك فان هناك علاقة وثيقة بين مثيلة DNA والهستونات فمثلا في الفطر *Neurospora* عند حدوث طفرة في ثمالة اللايسين التاسعة للهستون الثالث H3/K₉ فانه يؤدي الى اختفاء مثيلة DNA وهذا يحدث في الخمائر واللبائن.

- إضافة البايوتين الى الهستونات وهنا ايضا يضاف البايوتين الى ثملات اللايسين في ذبول الهستونات بمساعدة الإنزيم Biotinidase وتكون الإضافات الى H_2A/K_9 ، H_2A/K_{13} و H_3/K_4 ، H_3/K_9 ، H_3/K_{18} و H_4/K_{12} ، H_4/K_{18} والإضافات يمكن ان تتغير نتيجة لبعض الأضرار الحاصلة في DNA مثل كسور الأشرطة المزدوجة (DSBs) التي تقلل من إضافة البايوتين في حين ان تكوين مزدوجات الثايمين تزيد من قابلية إضافة البايوتين.
- فسفرة الهستونات: تمثل آلية أخرى من آليات تحويل الهستونات ويمكن ان تحصل في مواقع مختلفة ، وتكون بشكل عام مشجعة للـ DNA للتفاعل والتداخل مع عوامل الانتساخ او عوامل التنظيم الأخرى ويكون لمستوى الفسفرة أثراً في التأثير الذي تضيفه.
- إضافة بروتين Ubiquitin (Ubiquitylation) وهي من التحويلات التي تحصل للهستونات اذ يرتبط Ubiquitin وهو بروتين صغير يصل وزنه الجزيئي الى 8.5 كيلودالتون ويرتبط الى ثملات اللايسين ويؤدي الى تحويل فعاليتها او يقود البروتين الى الجسيمات البروتينية لغرض تفكيكه. والعملية تحصل بثلاث خطوات إنزيمية رئيسية ، وكل الهستونات الأربعة المكونة للجسيمات النووية تكون عرضة لهذه العملية ولكن بمعطيات مختلفة، فعند التفاعل او الارتباط الى الهستون H_2A فانه يكون مرافقا لإسكات الجين ، والارتباط الى H_2B يكون في الغالب مرافقا لتنشيط الجين وفي أحيان قليلة يؤدي الى الإسكات .
- وهناك عمليات وتحويلات أخرى تحصل للهستونات يمكن ان تؤدي الى إزالتها، وعمليات تشارك فيها وصفات الهستونات وغيرها التي قد تكون درست بشكل أفضل في البروتينات اللاهستونية.

- Sumoylation وهي إضافة بروتين صغير يشبه البروتين المدمر Ubiquitin.

دور البروتينات اللاهستونية في الوراثة اللاجينية

هناك بروتينات أخرى لاهستونية تساهم في تنظيم الجينات في خلايا حقيقية النواة الراقية ولعل أهمها التي درست بشكل جيد هي معقدات البروتينات Polycomb group (PcG) و Trithorax group (TrxG). تستطيع هذه البروتينات ان تحافظ على نسبة الانتساخ والاحتفاظ كذلك بذاكرة الانتساخ أثناء عمليات التطور والتمايز. وتتغير سيطرة البروتينات أثناء حياة الخلية حقيقية النواة استجابة للظروف الخارجية وذلك بآليات تتم بعد الانتساخ (PTMs) فمثلا PcG تؤثر في عملية المثيلة او الأسئلة التي تؤثر في فعاليتها الحيوية، وتؤثر هذه المجموعة من البروتينات في مثيلة وتحويلات أخرى تجري على الهستونات بعد ارتباطها

بمواقع خاصة بها. اما TrxG فهي ترتبط الى المواقع نفسها التي ترتبط بها PcG ولكنها في الغالب تؤدي الى تنشيط الانتساخ عبر تحويل الهستونات وإعادة تشكيل الكروماتين. وتعاني البروتينات اللاهستونية ايضا من إضافة البروتين المدمر Ubiquitylation الذي يقودها الى جسيمات البروتين او ما يسمى بمقبرة الخلية لتفكيكها، ولكن إضافة البروتين المدمر يمكن ان يقوم بوظائف أخرى غير قيادة البروتينات المطموعة به الى جسيمات البروتينات الحالة.

وتقوم بعض أصناف PcG بإضافة البروتين المدمر وأكثر الهستونات استهدافا هو H₂A مما يؤدي الى إخماد المنطقة التي تحصل فيها العملية وتعد العملية من أهم الأحداث على المستوى الجزيئي لتعطيل الكروموسوم الجنس X ، وبعض البروتينات اللاهستونية PcG تعد من الإنزيمات اللحم Ligases. وتكون عملية إضافة البروتين المدمر بعض الأحيان ذاتية، كما انها يمكن ان تحصل في أكثر من موقع مثل الارتباط الى ثملات اللايسين K₆ ، K₂₇ ، K₄₈ ، والبعض منها يشارك في عملية الإسكات وليس كلها.

وتحدث حالة Sumoylation التي تقوم بها بروتينات PcG للهستونات وتعرض هي نفسها لهذه العملية وبآليات مختلفة تؤدي في النهاية الى كبح الانتساخ، وكذلك تؤثر في بعض الإنزيمات مثل Dnmt3a في الطرف الكربوكسيلي الذي يجعله يتداخل مع عملية إزالة أسئلة الهستونات اي مع وظيفة Histone deacetylase، والخاصة ان Sumoylation تشبه حالة إضافة البروتين المدمر Ubiquitylation تحصل بعد عملية الانتساخ وتشارك في تنظيم التعبير الجيني ولكن الاختلاف ان Sumoylation يكون في حالة تداخل داينميكي مع البروتينات غير العائدة الى PcG.

وتحدث الفسفرة لـ PcG وتعد من الأحداث المهمة بعد الانتساخ وتعمل هذه البروتينات على فك مجاميع الفوسفات المرتبطة اليها في أواخر مرحلة التخليق من دورة الخلية، في حين توجد مجموعة منها تقوم بفسفرة البروتينات بشكل دائم في مرحلة G₂/M. اما بروتينات Thiorax فتقوم بفسفرة بعض البروتينات بشكل يعتمد على دورة الخلية.

ان عملية فسفرة البروتينات تؤثر في تحديد مواقع البروتينات مثل توزيعها في السايكوبلازم او النواة، فالبروتينات الأكثر حركة تكون في السايكوبلازم اما البطيئة فتبقى في النواة. وقد أشارت دراسة مكنون البروتينات المفسفرة Phosphoproteomics الخاصة بتصنيف مواقع الفسفرة بـ PcG الى ان هناك 118 موقعا في جينوم الإنسان و 32 في جينوم الفيران و 14 منها مشتركة بين الإنسان والحيوان. والملاحظ ان توالي الحوامض الامينية حول موقع الفسفرة هي التي توجه إنزيمات الفسفرة Kinases الى الموقع لفسفرته، وهذه المواقع (الجانبية) تتصف بوجود البرولين فضلا عن بعض الحوامض الامينية من صنف الحامضية ولكنها اقل أهمية، والكانيزات التي تميز وترتبط الى البرولين هي تابعة للـ (Mitogen activated protein) MAP kinases وهو أحد المسارات المهمة العاملة داخل الخلايا. وفي العموم فان عملية الفسفرة تساعد في إعادة برمجة فعاليات الجين كما في Epigenetic remodeling events. والمعقدات البروتينية PcGs التي ترتبط الى الكروماتين وتؤثر فيه ، ويوجد منها حوالي 60-100 في اللبائن ولها مواقعها الخاصة التي تعاني التحويل بالآليات المذكورة أعلاه، وكل

نوع من التحويل يرتبط بعمليات تعتمد على الكروماتين التي تؤدي دورا مهما فيه ومنها داينميكية الانتساخ، والإسكات الطويل الأمد وتضاعف DNA والاستجابة للتدمير الحاصل للـ DNA وغيرها من الأحداث لغرض استمرار بقاء الخلايا. ومن جهة ثانية هناك بعض الأمراض المرتبطة بهذه المجموعة من البروتينات، فالتعبير الزائد عنها معروف في بعض حالات السرطان والهرم والسمنة وداء السكري وغيرها ولذلك فهي يمكن ان تشارك في بعض العلاجات.

وتتعرض بروتينات PCG الى التفكك بواسطة الجسيمات النووية الحالة او بالتعرض الى إنزيمات Caspases وكذلك يحصل لها تحلل أثناء عملية الاستماتة والتي ربما تكون مسئولة عن تجزئة وتكثيف DNA التي تعد أحد مظاهر الاستماتة.

البريونات Prions

تمثل أحد عوامل الوراثة الخارجية اللاجينية وهي أحد أشكال البروتينات المسببة للإصابة بالأمراض. وحقيقتها انها بروتينات مطوية بشكل غير صحيح تؤدي الى تخریب الوظائف الخلوية وفي الفطريات والخمائر تستورث هذه المكونات دون التأثير في الجينوم . وتأثيرها يكون من خلال حجزها البروتينات في تجمعات خاصة والبعض منها يزيد من عملية الترجمة بالمساعدة بعبور شفرات الوقف وبالتالي اما تؤدي الى التخلص من الطفرات التافهة او تزيد من طول البروتين عن المقنن له (وقد مر ذكر بعض الجوانب الخاصة بها في موقع سابق) .

بعض المظاهر الهامة للتنظيم

هناك بعض المظاهر العامة للتنظيم التي تكون واضحة على النمط المظهري للأحياء . ومنها الطمع Imprinting وتعطيل كروموسوم الجنس X وغيرها وسيتم تناول المذكورة منها.

الطمع Imprinting

ظاهرة وراثية يتم فيها التعبير عن الجين بشكل معتمد على الأصل فيما اذا كان من الأم او الأب وهي عملية مستورثة مستقلة عن الوراثة المندلية، والجين المطموغ يعبر عنه من النسخة (الليل) الآتي من الأم او الأب، والمناطق المطموغة تكون في العادة ذات مثيلة عالية واقل نشاطا من حيث الانتساخ، وهذه الطمغات يمكن ان تمحي او تجدد في الحياة الجنينية المبكرة، وتوجد في الحشرات واللبائن والنباتات المزهرة. وعملية الطمع كما موضح آنفاً هي عمليات تتم خارج نطاق توالي القواعد في DNA اي انها ضمن الوراثة اللاجينية وتشمل عمليات المثيلة للـ DNA وتحويل الهستونات وغيرها للوصول الى التعبير عن نسخة واحدة من الجينات دون تغير تواليه . وهذه الواسمات اللاجينية تستحدث في الخلايا الجرثومية التناسلية وتبقى في الخلايا الجسمية للكائن وعملية التعبير عن الجينات المطموغة يكون ضروريا في عمليات التطور وعند حصول اضطرابات فيها تؤدي الى عدة أمراض منها Angelman syndrome ، Prader-Willi syndrome و Silver-Russell syndrome وغيرها.

في الأحياء ثنائية مجموعة الكروموسومات تحوي الخلايا الجسمية على نسختين من الجينوم ويكون كل جين ممثل بنسختين او أليلين آتية من أحد الوالدين وفي عموم الخلايا الجسمية يكون التعبير آتياً من النسختين بشكل متزامن. وفي اللبائن هناك حوالي 1% من الجينات تكون مطموغة اي ان التعبير يتم من أحد النسختين فقط وهذا يعتمد على أصل الجين هل هو آتي من الأم او الأب فمثلا الجين المشفر لمستقبل الأنسولين Insulin-like growth factor 2

(IGF2/Igf2) يعبر من النسخ الآتية من الأب فقط. وقد اكتشفت الحالة في الحشرة *Pseudococcus nipae* التي تنشأ من بيضة مخصبة، وفي البيوض التي تعطى إناث تبقى كروموسوماتها في حالة كروماتين حقيقي اما في الذكور فان مجموعة من الكروموسومات تكون في حالة كروماتين متباين بدأ من الانقسام السادس للبيضة المخصبة وتبقى كذلك في معظم الأنسجة وتكون هذه نتيجة عملية الطمع والتي بدورها تحدد جنس الكائن. وعملية الطمع تكون مميزة لبعض اللبائن ففي الفيران وجد ان الحاجة ماسة لمشاركة كل الجنوم الأبوي والأموي ، واما في الإنسان فان هناك نسبة كبيرة من الجينات المطموغة تعمل أثناء نمو الجنين والمشيمة وتطور الكائن.

وهناك عدة طرق لتحديد عملية الطمع منها طرق تقليدية ولكن التطورت في الوقت الحاضر والطرق الجديدة التي استعملت لرسم خرائط طمع جينومات مختلفة أشارت الى ان 80% من الجينات المطموغة تتجمع على شكل عناقيد وتسمى الدومينات المطموغة والتي تشير الى وجود سيطرة منظمة على توزيعها.

آلية الطمع

عملية الطمع هي عملية متحركة (داينميكية) اي يمكن محو الطمع او استعادته أثناء الجيل الواحد اي انها عمليات تتم بعيداً عن تواليات DNA، لذلك يلاحظ ان الطمغات يمكن ان تمحي وتعود اعتماداً على جنس الكائن (ذكر او أنثى) ، فالملاحظ انه في النطفة تعود طمغة الأب للظهور وفي البويضة تعود طمغة الأم للظهور، وعمليات محو الطمغة وإعادتها ضرورية جداً فهي متعلقة بالجنس كما في الانسان. وفي النباتات واللبائن تعمل آليتان في هذا المجال وهي مثيلة DNA وتحويل الهستونات.

تنظيم الطمع

ان تجمع الجينات المطموغة بعناقيد (كما ذكر أعلاه) يوفر إمكانية المشاركة في عمليات التنظيم ويتم بواسطة مناطق سيطرة الطمع Imprinting control regions (ICRs) وهذه المناطق توجد في DNA وتكون غنية بـ G و C ويكون فيها السايتوزين حاوياً على المثل من نسخة واحدة من مزدوج DNA. وعلى العموم فان عملية المثيلة هذه ليست مرتبطة بعملية الإسكات بشكل دائم وإنما تعتمد على الحالة التي يتم فيها الطمع.

وظيفة الجينات المطموغة

ان التعبير عن أنواع مخصصة من الجينات بواسطة عملية الطمع تشترك فيها اللبائن التي توجد فيها المشيمة وكذلك الجربيات Marsupials مثل الكنغر وكذلك النباتات المزهرة. ويمكن ان يحصل طمع كلي لبعض الكروموسومات التي لوحظت في بعض الحشرات وكذلك طمع الكروموسوم الجنس في بعض أنسجة الفيران والإنسان وغيرها من ثنائيات الجنس. وأكثر الجينات المطموغة لها دور في نمو المشيمة وبعضها يعمل في التطور بعد الحياة الجنينية التي لها علاقة بعملية الرضاعة والايض.

منشأ الطمع

ان عملية الطمع لا بد منها لإضفاء اللياقة المفيدة للكائن وجيناته التي يحويها، فالجينات الأبوية تميل لتشجيع النمو والأموية تميل لجعل النمو محدداً، والملاحظ ان التأثير الأموي هو الغالب في البيضة المخصبة. وهناك تفسير آخر يخص عملية الطمع منها ان الانتخاب الطبيعي للطمغ

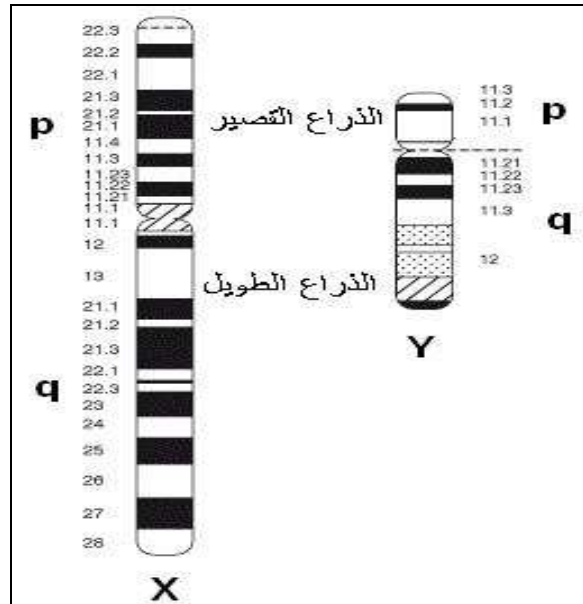
الجينومي يحتاج الى تغاير وراثي في المجتمع مما يساعد أنظمة دفاعات المضيف المسؤولة عن إسكات الجينات الغريبة مثل الإصابة بالفيروسات وكذلك إسكات الجينات التي هي صامتة أصلاً ولكن يحصل تعبير عنها بالخطأ وهي التي تدفع بحدوث عمليات الطمع، ولذلك يلاحظ ان بعض الفيروسات والجينات الارتدادية القافزة Retrotransposed genes اذا دخلت الى الجينوم قرب مناطق صامتة فهي تكتسب هذه الصفة وتعطل فعاليتها.

مشاكل الطمع

تؤثر عمليات الطمع في حالة كونها مؤدية الى حالة الإسكات في عمليات الكلونة خاصة مع DNA غير حاوي على المثيل عند دخولها الى مناطق صامتة وقد كانت هذه عقبة أمام تطوير النباتات لوقت طويل، وكذلك تشارك بآليات مختلفة في حث بعض سرطانات الثدي والمبيض.

تعطيل كروموسوم الجنس X

كروموسوم X واحد من الكروموسومات المحددة للجنس في العديد من الأنواع الحيوانية وبضمنها اللبائن ويكون النمط الأنثوي XX والذكري XY ، ويحوي كروموسوم الجنس X على حوالي 153 مليون من النيوكليوتيدات وتمثل في الإناث ما يقرب من 5% من الجينوم وفي الذكور 2.5%. ويحوي الكروموسوم X حوالي 2000 جين من المجموع الكلي لجينات الإنسان البالغة 30,000-25,000 (على آخر تقديرات الجهات المختصة) اما الكروموسوم Y فيحوي على 200 جين، ويمثل الشكل التالي (94.3) مخطط للكروموسوم X و Y:



شكل 95.3 : كروموسومات الجنس (X , Y)

وحصول الطفرات في الكروموسوم الأنثوي X chromosome يطلق عليها الطفرات المرتبطة بالجنس ويعطل أحد كروموسومي X في الإناث في الحياة الجنينية المبكرة وينحسر في تركيب يسمى Barr body في الخلايا الجسمية، لإعطاء فرصة متكافئة لكلا الجنسين ومنع

زيادة جرع الجينات في الإناث. الكروموسوم الأنثوي يكون أكبر من الكروموسوم الذكري (Y chromosome) كما موضح في الشكل أعلاه وفيه كروماتين حقيقي أكثر من الذكري، وتوجد فيهم مناطق متشابهة ويفتقر الكروموسوم الذكري الى بعض المناطق الثابتة الموجودة في الكروموسوم الأنثوي. وبما ان الذكور تحوي نسخة من الكروموسوم الأنثوي فهم عرضة للأمراض المتعلقة به. وتظهر العديد من الاضطرابات نتيجة الاختلاف في عدد الكروموسومات الأنثوية بشكل خاص فمثلا زيادة الكروموسوم الأنثوي XXY في الذكور تؤدي الى زيادة المواد الوراثية من الكروموسوم الأنثوي التي تتداخل مع تطور الذكور وتقلل من هرمون Testosterone وتمنع Testicles من العمل بشكل طبيعي، وهناك حالات أخرى مثل XXXXY ، XXXY ، XYYY هذه في العادة ترتبط بانخفاض القابلية على التعلم والقراءة، ويكون معدل الذكاء IQ في حالة XXY في الحدود الطبيعية ولكن ما بعدها من الزيادات تقلل من مستوى الذكاء . كما ان الإناث ذات النمط XXX يكون مستوى الذكاء عندهم 90 مقارنة بـ XX (100)، ولكن في مجموعة XXX تكون الإناث أطول قامة وخصبات اي كثرات الولادة وهي صفة لا تورث لأبنائهن. ويمكن ان تكون الخلايا الجسمية موزائكية او مختلطة البعض XY طبيعي والآخر XXY ، كما انه توجد حالات كثيرة ترافق بعضها صفات مظهرية واضحة.

ولهذه الأسباب يحصل تعطيل للكروموسوم الأنثوي وتسمى ايضا Lyonization باسم مكتشفتها Mary Lyon. ويكون التنشيط لأحد الكروموسومات X في الإناث وذلك بتغليفه بكروماتين متباين معطل للانساخ، واختيار أي من الكروموسومين للتعطيل قد يكون عشوائيا في الفيران والإنسان ذات المشيمة وعند تعطيله يستمر طوال حياة الخلية ، اما في الجربيات فان التعطيل يكون للكروموسوم الآتي من الأب. وتحصل عملية التعطيل في الحياة المبكرة اي في مرحلة تكون 2-4 خلايا جنينية في المشيمة والأنسجة الأخرى المساندة للجنين ويبقى الكروموسوم الأموي فقط هو الفعال، وتكون عملية التعطيل غير قابلة للرجوع في حياة الخلية ويرجع في بيوض الإناث.

والإناث الطبيعية تحوي على كروموسومين أحدهما نشط يرمز له Xa والآخر معطل Xi وفي حالة احتواء الخلايا على أكثر من اثنين من الكروموسومات الأنثوية يبقى واحد فعال والأخرى تكون معطلة. وهناك عامل يرتبط الى الكروموسوم ويمنع تنشيطه اما الباقي فيتم غلقها لمحدودية هذه العوامل. ويرتبط العامل المذكور الى منطقة محددة على الكروموسوم Xa تسمى X-inactivation center (XIC) تسيطر على إسكات الكروموسوم والعامل المفترض يرتبط الى هذه المنطقة . ان XIC تكون ضروريا وكافيا لتعطيل الكروموسوم وذلك واضحا عند حصول التناقل بين الكروموسومات فان انتقال القطعة XIC الى كروموسوم جسمي يؤدي الى تعطيله، فضلا عن ان الكروموسومات الأنثوية التي ليس فيها XIC لا تعطل.

وتحوي منطقة XIC على جينين للـ RNA غير قابلة للترجمة هما Xist و Tsix العاملة في إسكات الكروموسوم الأنثوي، والمنطقة تحوي على مواقع ارتباط للبروتينات المنظمة المعروفة الهوية وأخرى لا تزال غير معروفة. والجين X-inactive specific transcript (Xist) يعطى جزيئة RNA غير مشفرة مسئولة عن إسكات الكروموسوم الذي تنتسخ منه حيث يقوم

Xist RNA بتغطية الكروموسوم المعطل وليس Xa ولذلك فالكروموسوم الذي ينقصه Xist gene لا يمكن ان يعطل ونقل الجين Xist الى الكروموسومات الجسمية وانتساخه يؤدي الى تعطيلها. والملاحظ انه قبيل عملية التعطيل فان كلا الكروموسومين الأنثوية تكون فعالة بشكل ضعيف لإعطاء Xist RNA وخلال عملية التعطيل فان أحدهما Xa يوقف انتساخ Xist ، في حين ان Xi (المحدد مسبقا) يزداد فيه إنتاج Xist RNA ويغطي الكروموسوم بدءاً من XIC ويسكت الأغلبية العظمى من الجينات الموجودة فيه.

اما Tsix فهو الآخر يشفر لجزيئة RNA كبيرة لا تشفر للبروتينات وتنتسخ من الشريط النقيض للـ Xist ، ويكون Tsix المنظم السلبي للـ Xist ، لذلك فالكروموسوم الذي ينقصه Tsix يكون التعبير فيه عند مستويات عالية من انتساخ Xist ويعطل بشكل اكبر من الكروموسوم الطبيعي، ومثلما هو الحال قبيل عملية التعطيل فان كلا الكروموسومين الأنثوية تنتج مستويات واطئة من Tsix RNA من الجين Tsix gene وعند التعطيل فان الكروموسومات التي مصيرها Xi توقف التعبير عن Tsix RNA ، وتزيد من التعبير عن Xist في حين يستمر Xa في التعبير زيادة عن Tsix لعدة أيام .

الإسكات

الأغلبية العظمى من الجينات في الكروموسوم المعطل لا يعبر عنها مقارنة بالكروموسوم غير المعطل وذلك نتيجة لإسكات لجينات Xi بجزيئات RNA والكروماتين المتباين. والملاحظ ان Xi فيه مستوى عالٍ من مثيلة DNA مقارنة بالكروموسوم الفعال Xa وكذلك يلاحظ مستوى واطئ من أسئلة الهستونات وكذلك مستويات واطئة من مثيلة الهستون H3 عند ثمالة اللايسين الرابعة في حين يكون مستوى مثيلة ثمالة اللايسين التاسعة عالياً للهستون H3 وهذه كلها أحداث مرتبطة بإسكات الجين، فضلاً على احتواء Xi على نوع من الهستونات هي Macro H2H الذي يوجد على الجسيمات النووية له.

ولكن بعض جينات Xi يمكن ان تتجاوز التعطيل، كما ان بعض الجينات يعبر عنها بشكل متكافئ مع جينات Xa. وتجاوز عملية التعطيل تحصل في جينات الفيران اما في الإنسان فهناك حوالي 25% من الجينات على Xi المعطل تتجاوز التعطيل ويعبر عنها وتكون هذه بشكل عناقيد.

ومن جهة ثانية هناك بعض الجينات التي تتجاوز عملية التعطيل تحوي على بعض المناطق الموجودة على الكروموسوم الذكري هذه تسمى المناطق الجسمية الكاذبة Pseudoautosomal region ، وبما ان كلا الجنسين يستلم نسخة من الجينات الحاوية على هذه المنطقة لذلك لا حاجة الى عمليات التعويض والتكافؤ في جرع الجين، وهذه المناطق في Xi لا تحوي على التحويلات الموجودة في باقي Xi وبذا فهي تمتلك وترتبط بالقليل من Xist RNA المسكتة. ان وجود جينات فعالة على الكروموسوم المعطل وبمساعدة بعض الاضطرابات في عدد الكروموسومات مثل حالة XO تؤدي الى بعض الأمراض.

الجوانب العملية لدراسة الأنظمة الحيوية

كما ذكر في مستهل الكتاب ان المعلوماتية الحيوية هي الوجه الحاسوبي لعلم الحياة الجزيئي ، وهذا يعني ان التعامل يكون بمستوى الجزيئات وفي مقدمتها DNA ثم البروتينات و RNA. وقد تم في العديد من الأحيان الربط بين طرق دراسة هذه المكونات او الجزيئات الحيوية والبرامج والوسائل المستعملة في المعلوماتية الحيوية، لذلك كان لابد من المرور على بعض هذه الطرق التي يمكن ان تشارك في توجيه وضع برامج الحاسوب.

الطرق المتعلقة بـ DNA

تكاد تكون هي المجال الأكثر تطوراً وتشعباً. ويمكن تكثير قطع DNA المطلوبة اما داخل الكائنات الحية *In vivo* كما في عمليات الكلونة او خارج الكائنات الحية.

كلونة DNA (DNA cloning)

تمثل وسيلة ملائمة للحصول على كميات كبيرة من توالي محدد من DNA مثل جين معين لدراسته ، ويكون ذلك بعملية التضخيم التي تنتج قطع عديدة متجانسة. وقد تكون القطع غير متجانسة وايضا يتم إنتاج عددا منها لتستعمل في تصنيع المجسات Probes لتكون وسيلة في التعرف على القطع المكمل لها فيما بعد أي بعد تحويلها الى أشرطة مفردة بعملية المسخ. وتشمل عملية الكلونة بعض الخطوات المحددة الرئيسة:

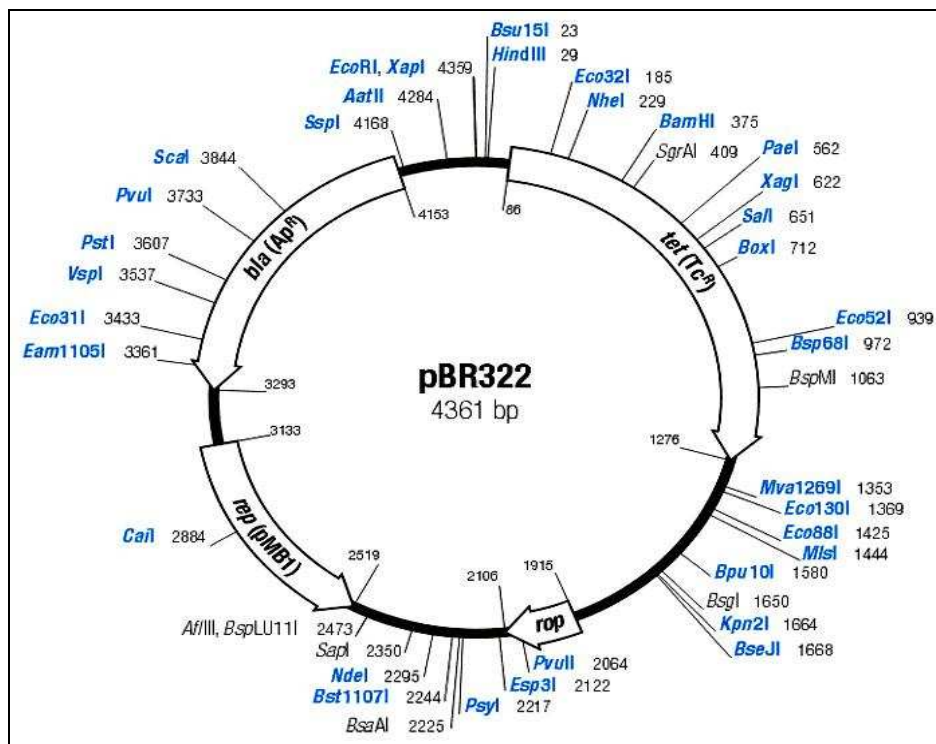
- تحديد قطع DNA المطلوبة ويتم ذلك بتقطيع DNA المراد أخذ القطع منه باستعمال إنزيمات القطع الداخلي التي تقطع DNA في مواقع محددة من التواليات المستهدفة وبهذه الإنزيمات يمكن تقطيع جينوم الخلايا الى قطع صغيرة من DNA ونظرا لتنوع إنزيمات القطع تنتج قطع مختلفة الأحجام يمكن فرزها على أساس الحجم ويفضل استعمال الإنزيمات التي تنتج نهايات لاصقة لتسهيل عمليات لحماها مع قطع أخرى من DNA تحمل أصل للتضاعف لتكثيرها ، ويمكن فرز القطع على هلام الاكروز بالترحيل الكهربائي اذ تكون القطع الصغيرة اسرع في الهجرة من القطع الكبيرة ويتم تحديدها بطريقة التهجين أي وصمة سوزن Southern blot التي سيأتي ذكرها لاحقا، وترتبط القطع المطلوبة بواسمات خاصة مثل جينات مقاومة المضادات الحيوية او اي صفة مظهرية مميزة ، والمتضاعفات التي تربط اليها القطع قيد الدراسة تدخل الى خلايا مضيفة مثل البكتريا والخمائر، والأولى أكثر استعمالا مثل *E. coli* وذلك لان هذه البكتريا معروفة الخلفية الوراثية، فضلا عن انها سريعة النمو مما يساعد في تقليص الوقت. والجزيئات المتضاعفة المهجنة سوف تتضاعف داخل الخلية المضيفة بشكل مستقل عن كروموسومها عند تنميتها بظروف ملائمة لإنتاج مجاميع خلوية متجانسة ، واذ كانت القطع المدخلة متنوعة عندها يمكن الحصول على مكتبة من النسائل Clone library ويطلق على المتضاعف الحاوي على DNA المستهدف مادة الناقل Vector material.

نواقل الكلونة Cloning vectors

كثيرة هي نواقل الكلونة الموجودة في الوقت الحاضر والمسوقة على نطاق تجاري وهذه تكون بمواصفات كثيرة يمكن ان تلاؤم بعض عمليات الكلونة ومن أهمها:

- البلازميدات وتستعمل لكلونة القطع الصغيرة من DNA بين 10,5 كيلوقاعدة وتكون حاوية على بعض الواسمات الوراثية لتسهيل تمييز الخلايا التي تلتقطها، كما انها تحوي

على مواقع انفلاق بـانزيمات القطع ليسهل إدخال المواد المراد كلونتها. وأكثر البلازميدات استعمالاً هو pBR322 وخصائصه موضحة في الشكل 96.3 :



شكل 96.3 : خارطة ناقل الكلونة pBR322

ويحتوي على مواقع انفلاق بالإنزيمات Pst7 ، EcoR₁ ، Sal 1 فضلا عن احتوائه على جينات المقاومة للمبسلين والتتراسايكلين وجينات المقاومة هي التي تأوي مواقع القطع بالإنزيمات المذكورة وتستعمل في انتخاب الخلايا المطلوبة (النسائل Clones) كما موضح في الشكل أعلاه .

ويمكن استعمال واسمات أخرى غير المقاومة للمضادات الحيوية مثل استعمال LacZ كجين إعلان Reporter gene ، وفي الوقت الحاضر أصبحت النواقل البلازميدية اقل استعمالا وذلك لتطوير عدد من النواقل التي تحمل قطع اكبر من DNA مثل النواقل المشتقة من الخمائر YACs وغيرها. كما ان التطورات طالت DNA المراد نقله لذلك استعمل cDNA وهي قطع مفردة الأشرطة يتم تخليقها باستعمال mRNA بعملية النسخ العكسي وتكون حاوية على الاكسونات اي المناطق المشفرة، وقطع cDNA اي المكملة تستعمل كمجسات فعند حصول اي تغيير في اكسونات الجينات يمكن الكشف عنها. لذلك فان تحضير cDNA وكلونته تكون مهمة، فمن تتبع تو الياته يمكن الحصول على استنتاجات أساسية حول الجين ونواتجه.

تضخيم DNA خارج الأنظمة الحية

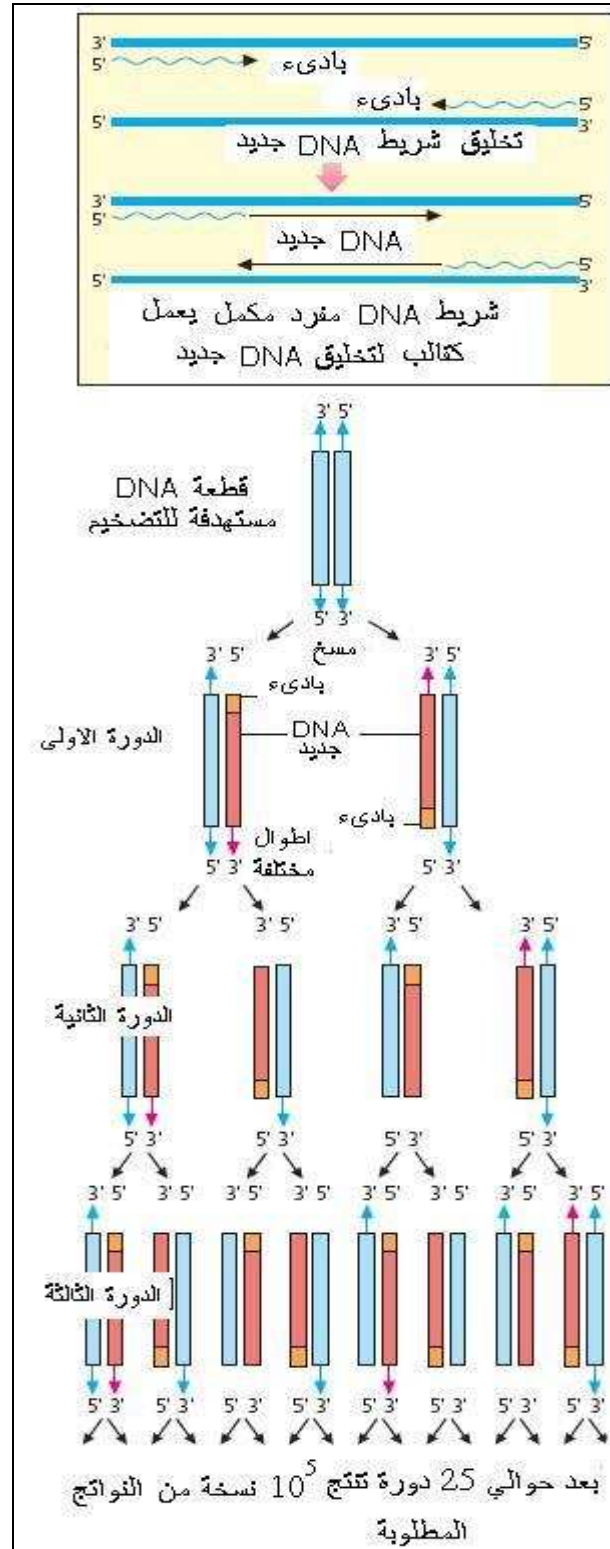
تتم باستعمال طريقة PCR التي فتحت عصراً جديداً من الوراثة الجزيئية وانتشرت الطريقة بشكل كبير وتطورت الأجهزة الأوتوماتيكية لأجرائها فضلاً عن تطوير العديد مما يلزم لإجراء الطريقة مثل تطوير برامج لتحديد البوادي التي تتم كوثره DNA عليها واستعملت في المجالات الأكاديمية والتطبيقية.

وطريقة PCR تتم خارج الأنظمة الحية وتكون حساسة لتكثير قطع DNA وان كانت لها بعض المساوئ والتي استغلت أيضاً لتطوير الطريقة. وتستخدم بكثرة للبحث عن التغيرات والطفرات في DNA. ومن محاسنها انه يتم تكثير DNA من قطع صغيرة جداً من المواد الأصلية. وعملية التضخيم تحتاج الى بعض المعلومات الأولية حول توالي DNA المستهدف مثل المناطق المحيطة بها ونوعية القواعد المشاركة، وبناءاً على هذه المعلومات يستعمل 2 من البوادي بطول حوالي 15-25 قاعدة وتكون مكتملة للتواليات خارج القطعة المطلوبة عند النهايات 3' و 5' وترتبط بشكل متخصص بها، والطريقة Polymerase chain reaction يشير الى حدوث سلسلة من التفاعلات لان الأشرطة الأولى ستمسخ وتستعمل قوالب لبناء أشرطة جديدة وهكذا ويمكن ان تستمر دورة التفاعلات هذه حوالي 25-35 دورة متعاقبة. ونظراً لتضاعف DNA في كل دورة ففي النهاية يكون هناك نسخ كثيرة من التوالي المستهدف وهذه يمكن ان تظهر على شكل حزم بأحجام مختلفة في الهلام بعد إجراء الترحيل الكهربائي. وكل دورة تشمل ثلاث تفاعلات محددة بشكل دقيق من ناحية التوقيت وتتم في الوقت الحاضر باستعمال جهاز الدورات الحرارية Thermal cyclers والدورات هي:

- مسخ أشرطة DNA المزدوجة لإنتاج أشرطة مفردة وتحدد الظروف بمحتوى الأشرطة من GC وبالنسبة للـ DNA البشري تكون الحرارة بين 93-95°م.
- الخطوة الثانية ربط البادي الى الأشرطة المفردة بدرجة حرارة يمكن ان تتراوح بين 50-70°م اعتماداً على درجة انصهار أشرطة DNA.
- الخطوة الثالثة تخليق أشرطة جديدة مكتملة للأشرطة المفردة الناتجة من الخطوة الثانية باستعمال إنزيم كوثره DNA ثابت للحرارة والذي يشتق من أحياء محبة للحرارة مثل *Thermophilus aquaticus* (Taq polymerase) الذي يعمل بدرجة حرارة 70-75°م.

وعند نهاية كل دورة تعمل الأشرطة المتكونة بعد مسخها كقوالب للدورات اللاحقة في التخليق.

وقد طورت الطريقة بشكل كبير لتلبي العديد من الاحتياجات. وقد أمكن من معرفة توالي حوامض أمينية في بروتين معين الى تصنيع DNA الخاص به ثم تضخيمه بالـ PCR، وكذلك يمكن باستعمال mRNA تخليق cDNA لغرض تضخيمها وعليه فعند توفر RNA بكميات صغيرة يمكن الحصول على ما يقابلها من الاكسونات وهذا يكون مفيداً عندما تكون الاكسونات متباعدة في الجين الأصلي. ويمكن إجمال الخطوات في الشكل 97.3 :



شكل 97.3 : مجمل خطوات PCR

الكشف عن قطع DNA

تختلف طرق الكشف عن DNA اعتمادا على النموذج المراد فحصه، فمثلا اذا كانت القطع مرحلة على الهلام فعندها يمكن الكشف عنها بوصمة سودرن التي سميت على اسم مكتشفها E. Southern عام 1975 ، وفيها يتم طبع الحزم من الهلام على أغشية من النايلون أو Nitro cellulose وعلى الأغشية الأخيرة يتم مسخ DNA بالقلويات للحصول على أشرطة مفردة وتثبت الأخيرة على الأغشية بمعاملة حرارية معتدلة تصل الى 80 °م أو تثبت باستعمال الأشعة فوق البنفسجية لعمل اتصالات عرضية بين الغشاء وقطع DNA الموجودة عليه. ثم يتم حضن الأغشية مع مجسات معلمة بمواد مشعة مثل ^{32}P المستعمل بكثرة والمجسات سوف ترتبط بقطع DNA المكتملة لها من ناحية القواعد النتروجينية اي تحصل عملية تهجين وبذا يمكن الكشف عن DNA بوضع أفلام أشعة X على الغشاء اذ تظهر DNA المعلم بشكل بقع سوداء بعد إظهار الفلم اي الحصول على Autoradiogram.

ويمكن معرفة حجم القطعة وموقعها من استعمال المعايير Ladder من قطع DNA المعروفة الحجم على هلام الترحيل الكهربائي.

تحديد توالي DNA sequencing

معرفة توالي النيوكلووتيدات في الجين يعطي معلومات مهمة حول التركيب والوظيفة والتطور والعلاقات بين الجينات ودرجة التشابه مع جينات أخرى في الكائن نفسه او في كائنات مختلفة. وقد يكون التوالي الناتج هو نسخة أولية Draft sequence ويحتاج هذا الى تغطية غير دقيقة لان هذه يمكن ان تعالج في المراحل النهائية ، وفي هذا التحديد غير المشذب Sequence redundancy تكون هناك زيادة تصل على الأقل ستة أضعاف ويغطي حوالي 90% من الجينوم وهذا النوع من التحديد يمكن ان يستعمل عندما يكون بدرجة عالية من النوعية للتعرف على الجينات وغيرها من مظاهر الجينوم دون الحاجة الى العمل التجريبي. وهناك أكثر من طريقة لتحديد توالي الجينوم منها الطرق الكيماوية مثل طريقة Maxam & Gilbert التي وضعت أسسها عام 1977 وطريقة إنزيمية وهي طريقة Sanger التي استعملت عام 1981.

التحديد بالطرق الكيماوية

تعتمد الطريقة على فلق جزيئات DNA بواسطة بعض المواد الكيماوية الخاصة. وتستعمل 4 مواد كيماوية في أربع خلائط من التفاعلات كل مادة في خليط، وكل تفاعل ينتج قطع DNA مختلفة الأحجام ويحدد الحجم بموقع النيوكلووتيد الذي تم فلقه في DNA . والمواد الأربعة المتخصصة مثل المادة Dimethyl sulfate (DMS) تهاجم قاعدة الكوانين وتضيف اليها مجموعة مثيل وحامض الفورميك يحور كل من الأدينين والكوانين، اما الهيدرازين Hydrazine فيحور كل من الثايمين والسايوتوزين وعند إضافة تراكيز عالية من الملح (5 مول) الى خليط التفاعل يصبح التفاعل خاص بالسايوتوزين ومع هذه المواد يحوي خليط التفاعل على Piperidine التي تقوم بفلق العمود الفقري للـ DNA في الموقع القادم للقاعدة المحورة. و DNA المراد تحديد تواليه يكون بشكل أشرطة مفردة ويعامل الشريط بإنزيم حل الفوسفات القاعدي Alkaline phosphatase لإزالة الفوسفات عند النهاية 5' ثم يلي ذلك تعليم الشريط

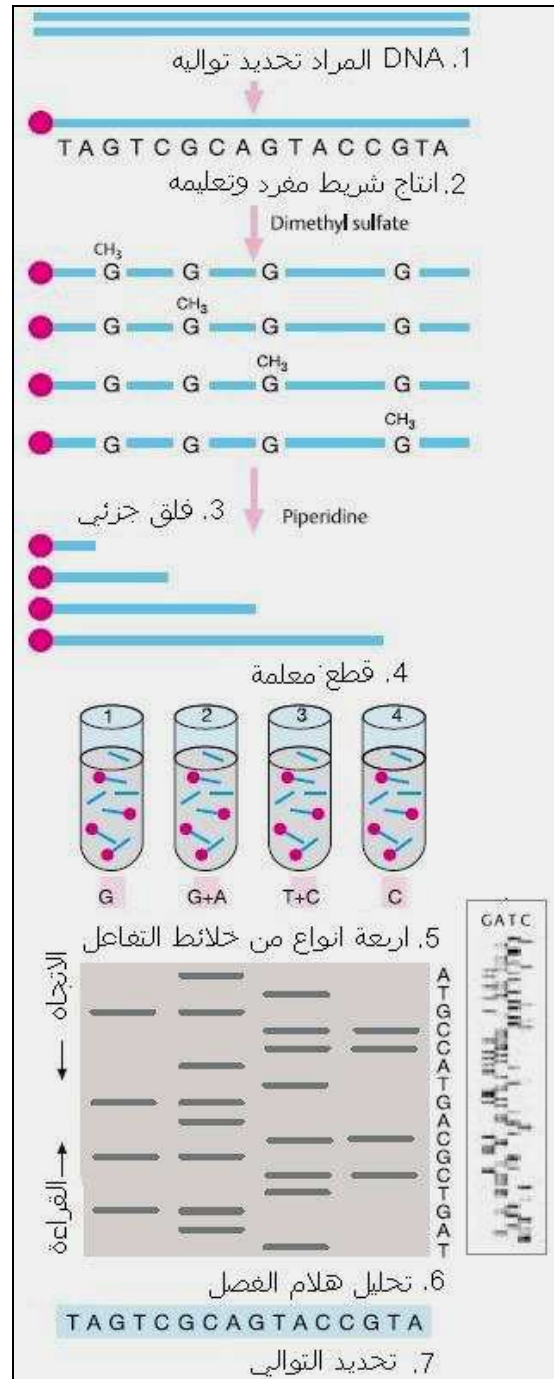
بالفسفور المشع وذلك بمفاعله بـ ATP حاوي على ^{32}P وبمساعدة إنزيم فسفرة النيوكليوتيدات المتعددة Polynucleotide kinase يربط الفسفور المشع الى النهاية '5. ثم يتم تقسيم نموذج DNA المهيأ للتفاعل الى أربعة أقسام لتتم مفاعلة القواعد النتروجينية بالمواد الخاصة بها كما مذكور أعلاه فضلا عن إضافة Piperidine لفصم السلسلة عند القواعد المحورة. ثم يتم ترحيل كل نموذج في الهلام لتحديد مواقع القواعد النتروجينية في DNA كما موضح في الشكل (98.3).

والطريقة تحتاج الى كميات من DNA النقي كما ان عملية إجرائها بشكل آلي غير متوفرة ولكنها لا تزال تستعمل لبعض الأغراض مثل حالة تقفي الأثر Footprinting.

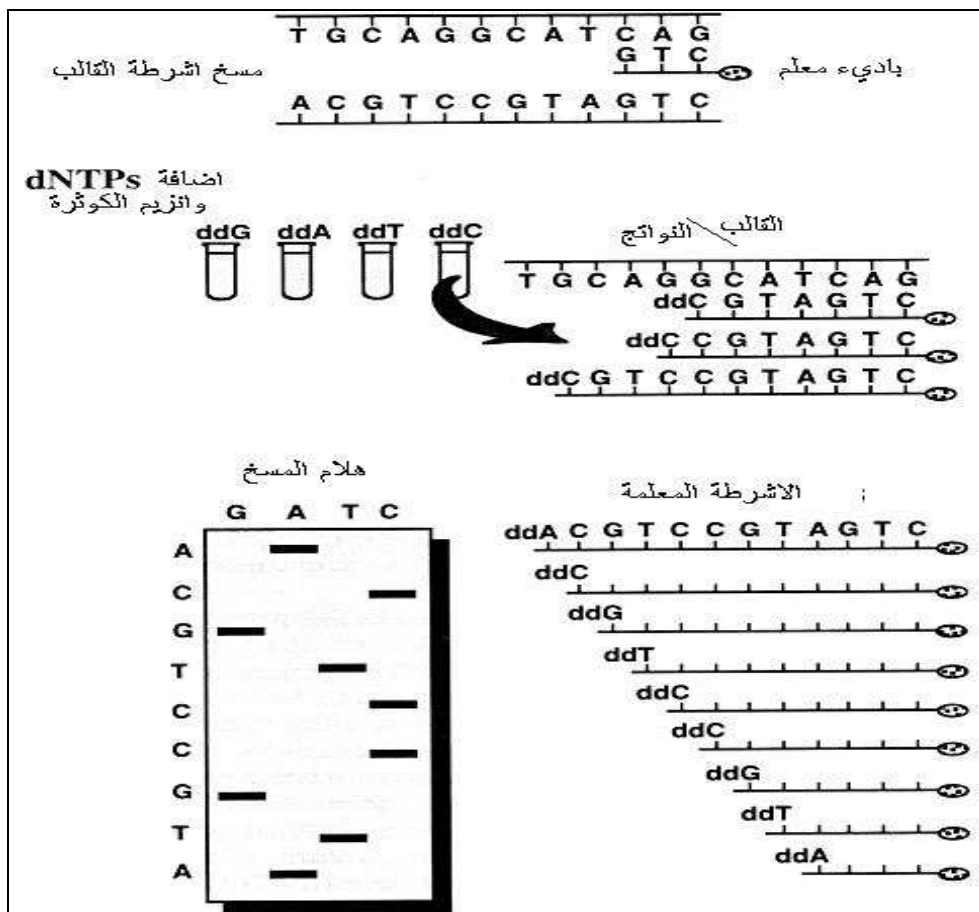
الطريقة الإنزيمية

الطريقة أكثر استعمالا من طريقة الفلق الكيماوي وتعتمد أساسا على تخليق DNA الذي يتم إيقافه عند وجود نيوكليوتيدات غير طبيعية وهي dideoxy nucleotides (ddNTP) وهي نيوكليوتيدات مضاهية للنيوكليوتيدات الطبيعية الا انه ينقصها OH- على الموقع الكربون 3' وعندما تدمج في السلسلة سوف لا يكون هناك OH- 3' في الموقع لإضافة نيوكليوتيد جديد لذا يتوقف التخليق عند هذه النقطة. و DNA المراد تحديد تواليه يكون بشكل أشرطة مفردة ويتم بدء التخليق باستعمال البوادي وإضافة نوع واحد من النيوكليوتيدات وهي اما ddTTP او ddGTP او ddATP او ddCTP وتكون معلمة بـ ^{32}P في مجموعة الفوسفات. ويمكن استعمال صبغات متفلورة.

وتعتمد الطريقة على اندماج أحد ddNTP في أشرطة DNA النامية باستعمال إنزيم كثررة DNA (DNAP I) وهي تعد من طرق كثررة DNA خارج الأنظمة الحية، واستعمال القواعد الشاذة يؤدي الى إيقاف أو إنهاء التفاعل بعد اندماجها ولذلك تنتج قطع مختلفة الطول التي يتم تحديدها بعد الترحيل الكهربائي اما باستعمال الأنابيب الشعرية والاكروز او هلام Acrylamide ثم قراءة توالي DNA. وعليه فان خليط التفاعل يتكون من قطعة DNA المراد تحديد تواليها، و بوادي خاصة ثم يقسم الى أربعة أقسام، تضاف الى كل قسم النيوكليوتيد الشاذ الذي يضاهي النيوكليوتيدات الأربعة الطبيعية. وتتم العملية وفق الشكل (99.3).



شكل 98.3 : تحديد توالي DNA بالطريقة الكيماوية



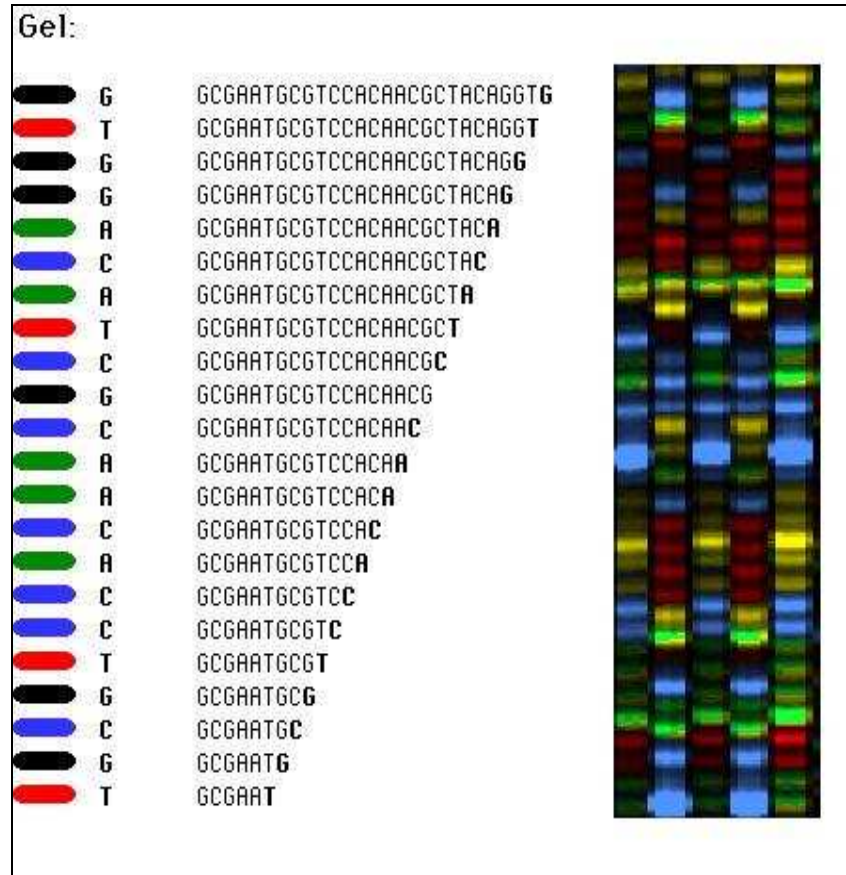
شكل 99.3 : الطريقة الإنزيمية لتحديد توالي DNA

اما في حالة استعمال الصبغات فتستعمل صبغة خاصة بكل قاعدة وعند الفصل والفحص بأشعة الليزر ويمكن ان تكون مخرجات النموذج بالشكل 100.3 :

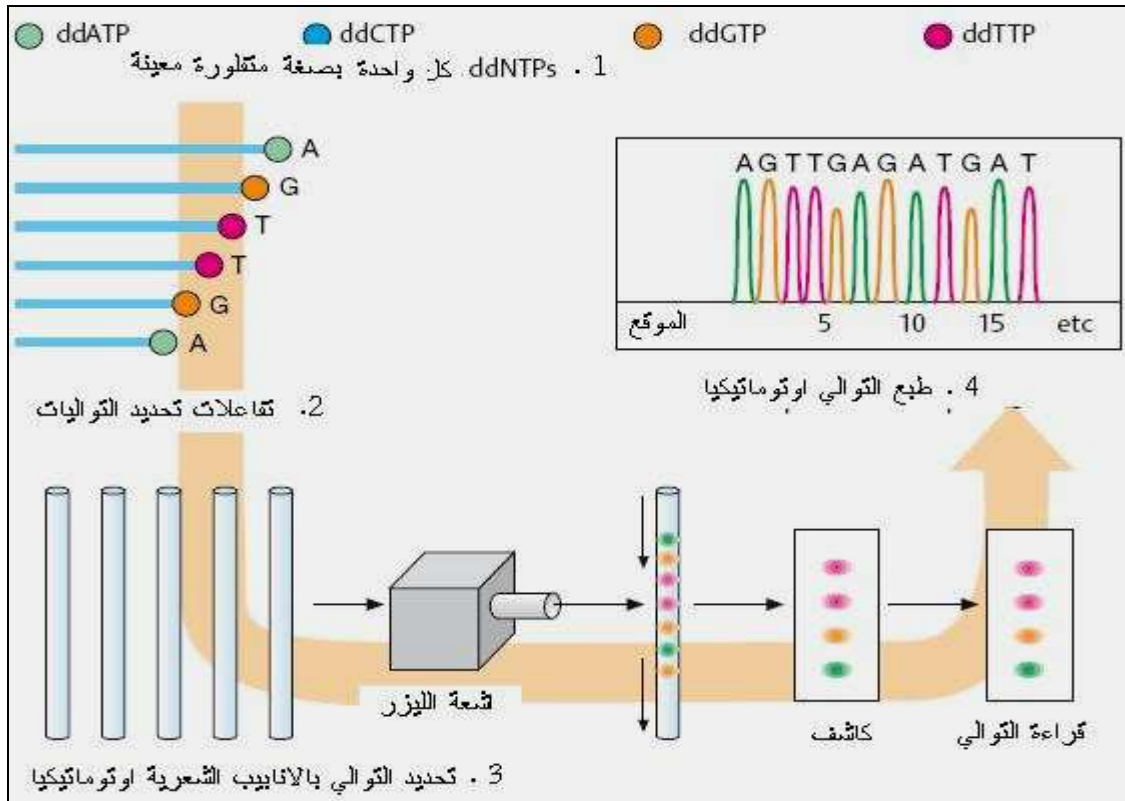
عملية تحديد توالي DNA آلياً Automated DNA sequencing

تحتاج عمليات تحديد تواليات DNA الكبيرة الى طرق آلية، وهذه الطرق تعتمد على تعليم DNA بصبغات متقلورة خاصة بكل قاعدة ونظام مناسب للكشف عنها. والتعليم بالصبغات المتقلورة تتم اما بطرق مباشرة او غير مباشرة وعند ربط هذه الصبغات الى القواعد تعطي الأخيرة ألوان مميزة عند تعريضها لضوء بأطوال موجية محددة. ومن أهم الصبغات المستعملة هي Fluorescein ويعطي لون اخضر باهت عند تعريضه لطول موجي 494 نانومتر، و Rhodamine ويعطي لون احمر عند طول موجي 555 نانومتر، وصبغة Aminomethyl coumarin acetic acid تعطي لون ازرق عند طول موجي 399 نانومتر، ويمكن مزج الصبغات لإعطاء لون آخر او رابع، وبذلك يمكن تعليم القواعد الأربعة بشكل مميز والمستعملة ddA تكون بلون اخضر، ddC تكون بلون ازرق، ddG تكون بلون اصفر و ddT تكون احمر، أي ان كل التواليات التي تنتهي عند الأدينين تكون خضر والتي تنتهي عند C زرق

وهكذا، وفي العادة تستعمل طريقة الترحيل الكهربائي في أنابيب شعرية ويتم الكشف عن الحزم باستعمال الليزر الذي يساعد في إظهار اللون المتفلور ويظهر على شكل قمم متبادلة بأحد الألوان المستعملة وملخص العملية يمكن ملاحظته في الشكل 101.3 :



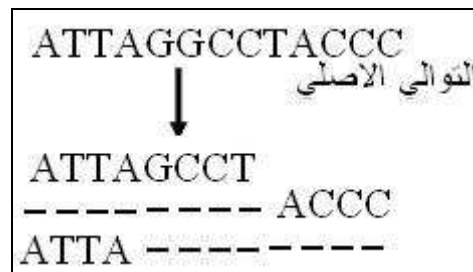
شكل 100.3 : استعمال الصبغات المتفلورة في تحديد التواليات



شكل 101.3 : تحديد التواليات بالطرق الآلية

تحديد التوالي بطريقة التشظية Shotgun

يمكن تحديد التوالي بالطرق المذكورة عندما تكون القطع صغيرة ولكن في الجينومات الكبيرة والكروموسومات تتم باستعمال طريقة التشظية حيث تجزأ جزيئات DNA الكبيرة ثم يتم تحديد التوالي لكل قطعة لإنتاج المقروءات Reads التي تمتد تواليها باتجاهين لإعطاء Mate pairs هي المقروءة ونهاية المقروءة End read وتجمع في صف طويل من المتجاورات Contigs وتستعمل برامج حاسوب ملائمة لتجمعها كما موضح في المثال:



وبالجمع يمكن الحصول على التوالي الأصلي. وقد كانت الطريقة هي المفضلة لعقد من الزمن 1995-2005 ولكن تم تطويرها إلى طرق أخرى مثل Next generation read وتغطي

25-500 قاعدة وهي سريعة وتعطي تغطية جيدة للتوالي وفي وقت قصير ولكن الدقة ليست بالعالية.

وتعاني طريقة التنشيطية من بعض عدم الدقة عند وجود المكررات المتشابهة ولكن يعتقد ان تطوير برامج الحاسوب يمكن ان تتجاوز هذه المشكلة. وقد طبقت طريقة التنشيطية عام 1995 لتحديد توالي البكتريا *Haemophilus influenzae* وأصبحت الطريقة القياسية لتحديد توالي بدائيات النواة وتستعمل بنجاح مع تواليات حقيقيات النواة.

انضغاط DNA (DNA compression)

تكون العملية الأخيرة في تحديد تواليات DNA هي الترحيل الكهربائي باستعمال ظروف المسخ في الهلام Denaturing polyacrylamide لفصل جزيئات DNA اعتمادا على الحجم ويفترض في هذه الطريقة توضيح حتى النيوكليوتيد الواحد من بين عدة مئات عند استعمال الأجهزة البسيطة للكشف. ولكن الطريقة تواجه بعض الصعوبات، فالهجرة الكهربائية وتحت ظروف المسخ تعتمد على الاختلاف في الحجم وهذا يتطلب إلغاء التركيب الثانوي للـ DNA والذي توفره ظروف المسخ المذكورة آنفاً مثل استعمال اليوريا بتركيز عالية وإجراء الترحيل بدرجات حرارية مرتفعة نوعاً ما، وهذا يكون كافياً لمعظم DNA والنواتج تعطي حزم منفصلة ولكن يلاحظ بعض الأحيان انضغاط في الحزم حيث تظهر في الموقع نفسه مرتبطة مع بعضها وتظهر بأكثر من ممر Lane في الموقع ذاته، والحزم التي قبلها تكون بعيدة أكثر من الحد الطبيعي عند مناطق الانضغاط وهذا دليل على ان التراكيب الثانوية لا تزال موجودة وتسمى الكواذب المضغوطة Compression artifacts ، لذلك تعالج الطريقة باستعمال تراكيز أعلى من اليوريا تصل الى 7 مولر ودرجة حرارة 50 °م التي تؤدي الى مسخ A/T للتراكيب الثانوية ولكن لا تصلح للتراكيب التي فيها 3 أو أكثر من أزواج G/C. والأخيرة تعالج باستعمال مضاهيات dGTP أثناء عملية التخليق (التي تحصل أثناء تحديد التوالي) ولكن هذه تكون ملائمة عند استعمال طريقة ddNTP وهي الطريقة الأسهل والأكثر كفاءة من استعمال المسخ باستعمال Formamide مع هلام تحديد التوالي.

ومن المضاهيات الملائمة استعمال dITP و 7-deaza-dGTP وعندما تحل هذه مكان dGTP في تفاعلات تحديد التواليات فان نيوكليوتيدات الكوانين في DNA سوف تستبدل بالمضاهيات المستعملة تؤدي الى تخريب التركيب الثانوي لمعظم تواليات DNA وتوصي الجهات المختصة باستعمال المادتين او الطريقتين لان كل واحدة ستفكك عقدة قطع محددة من DNA. وعملية الانضغاط هذه تستغل في بعض برامج الحاسوب للكشف عن بعض الحقائق البايولوجية كما سيتم ذكره في فصول قادمة.

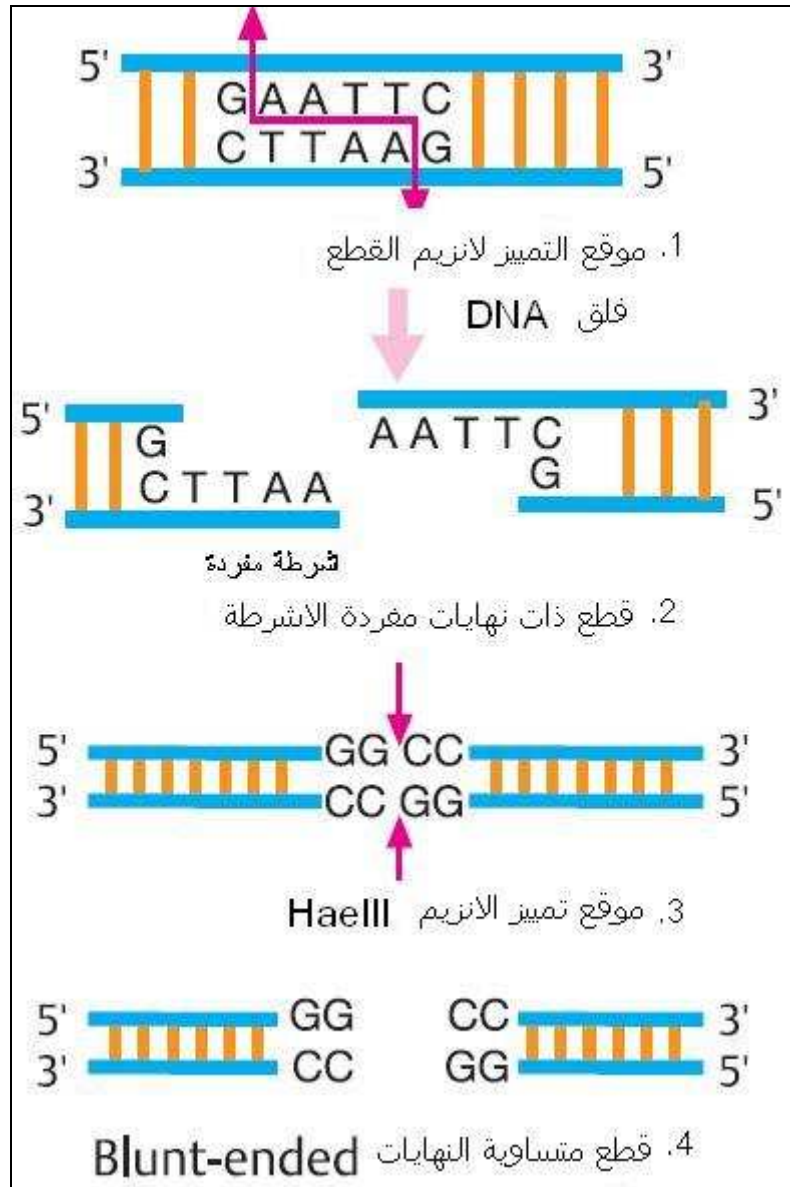
الخرائط الوراثية

تستعمل جزيئات DNA لوضع أنواع من الخرائط او الأطالس الخاصة وهي تمثيل لتوزيع المواقع Loci او الجينات على جينوم الأحياء وتستعمل لتحديد مواقع الجينات او مدى ترابط توارث بعض الجينات ولغيرها من الأغراض ومنها:

خرائط التقيد Restriction maps: وهي الخرائط التي تمكن من توصيف DNA بتبيان مواقع القطع فيها بالإنزيمات القاطعة ونمط توزيعها. وتتم باستعمال الإنزيمات البكتيرية التي

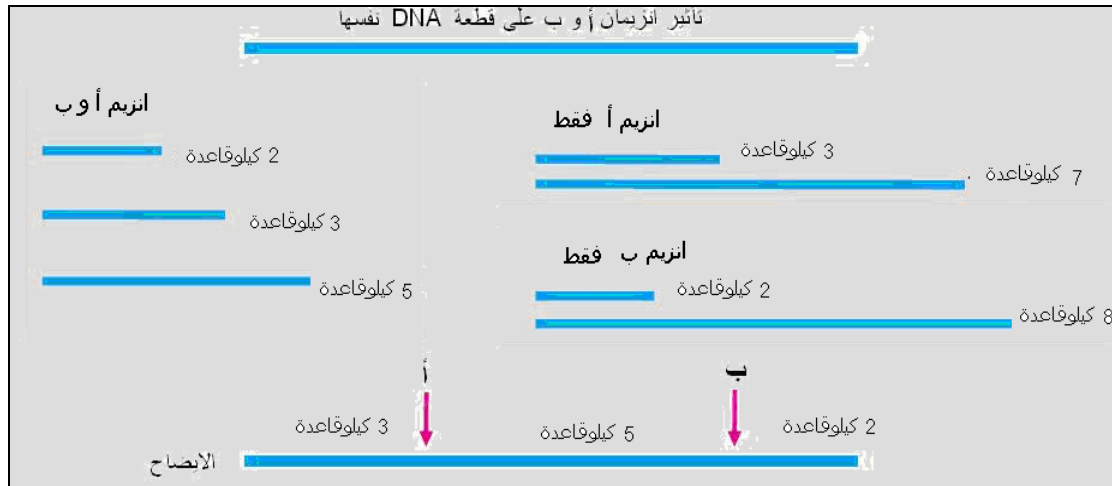
تنتج العديد منها كوسائل دفاعية ضد DNA الغريب الذي يدخلها، وكل إنزيم له توالي محدد من القواعد النتروجينية مكون من 4-8 قواعد واغلب الأحيان يكون 6 قواعد يمكن ان يميزه الانزيم ويعمل على قطعه ويسمى هذا بموقع القطع. وحجم القطع الناتجة يعتمد على توزيع هذه المواقع وهناك مئات من الإنزيمات المنتجة ومنها المسوق تجاريا.

تسمى الإنزيمات وفق ضوابط محددة فمثلا الإنزيم EcoR1 مشتق من *E. coli* Restriction enzyme 1 والإنزيم *Hind II* مشتق من بكتريا *Haemophilus influenzae* Rd ويقطع في أماكن محددة مثل GTpy↓puAC حيث يمكن ان يقطع في التوالي المذكور عندما تكون البريميدينات هي T او C وكذلك البيورينات هي A او G ومثل هذا الموقع يتكرر بكثرة مما يؤدي الى إنتاج قطع صغيرة، اما الإنزيمات التي تكون تواليات مواقعها نادرة فنتنتج قطع كبيرة. وقد يكون القطع متناظر للشريطين فينتج نهايات ملساء او غير متناظر ينتج نهايات لزجة كما في الشكل 102.3 :



شكل 102.3 : مواقع القلع المختلفة بإنزيمات القلع

وعلى سبيل المثال لو استعمل إنزيمين لقطع قطعة بطول 10 كيلوقاعدة الأول يقطعها الى قطعتين 3 و 7 كيلوقاعدة والآخر 2 و 8 كيلوقاعدة فعند استعمال الاثنين تكون القطع الناتجة 2، 3، 5 كيلوقاعدة فهذا يعني ان مواقع الفلق لهما تبعد 5 كيلوقاعدة كما موضح في الشكل 103.3



شكل 103.3 : قطع الإنزيمات المختلفة لقطع DNA

والجدول التالي (10.3) يوضح بعض الإنزيمات المستعملة في خرائط التقييد.

جدول 10.3 : بعض الإنزيمات القاطعة المستعملة في خرائط التقييد

توالي القطع	الإنزيم	البكتريا
5' GAA TTC AAG CTT	<i>EcoRI</i>	<i>Escherichia coli</i> KY 13
5' GTPy PuAC CAPu PyTG	<i>HindII</i>	<i>Haemophilus influenzae</i> Rd
5' AAG CTT GAA TTC	<i>HindIII</i>	<i>Haemophilus influenzae</i> Rd
5' GTT AAC TTT CAA TTG	<i>HpaI</i>	<i>Haemophilus parainfluenzae</i>
5' CC GG CC GG	<i>HpaII</i>	<i>Haemophilus aphrophilus</i>
5' GG CC GG CC	<i>HaeIII</i>	<i>Haemophilus aegyptius</i>

ومثل هذه الخرائط تكون مهمة من النواحي الطبية ومجال البحث في حالة دراسة التطور. كما انها تكون مهمة للكشف عن التغيرات التي تحصل في DNA . ومن أهم التغيرات هي تغيرات النيوكليوتيد المفرد SNPs وكذلك تستعمل في دراسة مظاهر أخرى . والمعروف انه لكل حوالي 100 قاعدة يمكن ان يوجد اختلاف في DNA، لذلك فان توالي التميز للإنزيمات يمكن ان يوجد على أحد الكروموسومات (في الأحياء مزدوجة المجموعة الكروموسومية) ولذلك

فان حجم قطع التقييد تختلف عند ذاك الموقع وهو ما يسمى Restriction fragment length polymorphism (RFLP) ويستفاد منها في معرفة كون الشخص متجانس الزيجة Homozygous او متباين الزيجة ، وفيما اذا كانت هناك طفرات في قطع معينة من الكروموسوم مثل قطعة من 5 كيلوقاعدة المعروف انها تأوي موقع للقطع بأحد الإنزيمات فغياب هذا الموقع يعني حدوث طفرة في هذه القطعة. وتغايرات DNA يكشف عنها بطرق أخرى غير استعمال إنزيمات القطع مثل استعمال تهجين وصمة سوذرن وغيرها من الطرق.

الخرائط الفيزيائية Physical maps

خرائط تعتمد على التحليل المباشر للـ DNA وتحدد المواقع بين Loci بعدد القواعد، ويوجد منها نوعان : الخرائط القصيرة التي تشمل 30 كيلوقاعدة والتي تقرب من حجم الجين، والخرائط الطويلة التي تصل الى طول ميكاقاعدة وتستعمل فيها الإنزيمات ذات مواقع القطع النادرة ثم يتم فصل القطع باستعمال ترحيل كهربائي خاص Pulsed field gel electrophoresis (PFGE) الذي يمكن ان يفصل قطع يصل حجمها 6 ميكاقاعدة او أكثر وبعد ذلك يكشف عن المواقع باستعمال وصمة سوذرن وبمجسات خاصة وتستعمل الطريقة للقطع الكبيرة التي يتم نقلها باستعمال كروموسوم الخميرة الصناعي Yeast artificial chromosome (YAC). ويستعمل هذا النوع من الخرائط لتحديد مواقع آلاف الجينات على الكروموسوم.

خرائط الارتباط Linkage maps

وتسمى ايضا خرائط التأشب وهي النوع الأول من الخرائط التي طورت بعد دراسات مندل. وتعتمد على تحليل التوارث التقليدي للصفات الوراثية المظهرية للأفراد الناتجة وفي الحقيقة هي تقع ضمن الخرائط الكروموسومية والخرائط الفيزيائية. وفي هذا النوع تقاس المسافة بين المواقع Loci بـ Centimorgan الذي يمثل معدل عبور 1%.

الخرائط الكروموسومية

وتسمى ايضا خرائط الوراثة الخلوية Cytogenetic maps وتعتمد على التتميط النووي وتشمل تحديد الحزم على الكروموسوم.

ومن بين الخرائط أعلاه يلاحظ ان الخرائط الفيزيائية هي التي تعطي المسافة الصحيحة حول طول قطعة من DNA التي تفصل بين مواقع الكروموسوم ، والأخريات المفروض ان تعطي المسافات الصحيحة ولكن ليست مثل الخرائط الفيزيائية ، وفي الحقيقة تعد الخرائط الكروموسومية صورة أخرى للخرائط الفيزيائية لانها تعتمد على مؤشرات أخرى، ولذلك فيعتقد ان وحدة القياس Centimorgan ربما تختفي لتحل محلها ميكاقاعدة.

مكتبات DNA

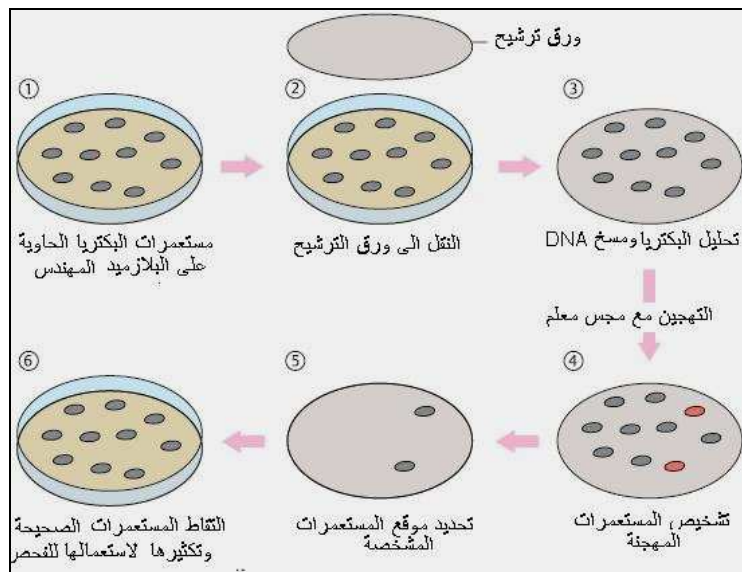
مجمع لقطع DNA التي تكون بمجموعها الجينوم الكلي للكائن وتعد نقطة البداية لعملية كلونة الجينات التي لا يعرف موقعها على الكروموسوم. ولإنتاج المكتبات يهضم DNA بالإنزيمات القاطعة وترتبط القطع الناتجة الى نواقل وتكثر في مضاييف بكتيرية بمواصفات خاصة، ويتم إنتاج عدد من الكلونات كي يكون هناك تمثيل لكل

قطعة على الأقل مرة واحدة في الخلية المضيفة. وفي إعداد المكتبات تستعمل البلازميدات والعائيات أو النواقل المهجنة منها مثل الكوزميدات Cosmids كنواقل تربط إليها قطع DNA وفي حالة القطع الكبيرة يمكن استعمال خلايا الخمائر ونواقلها. وهناك نوعان من المكتبات: **المكتبات الجينومية**: نسائل أو كلونات Clones (خلايا مضيفة) تحوي قطع من الجينوم من كل كروموسوم وتحوي على المناطق المشفرة وغير المشفرة. ولإيجاد جين معين يجب إجراء مسح على المكتبة الكاملة.

مكتبات cDNA: نسائل أو كلونات تحوي على المناطق DNA المشفرة ولذلك تفضل بعض الأحيان على مكتبة الجينومات. ولإنشائها تحتاج إلى توفر mRNA لغرض بناء cDNA. وهذه المكتبات لا تعطي فكرة عن تركيب الجين. وجزيئات mRNA يتم الحصول عليها من الخلايا التي يعبر فيها عن الجين المعني. وتتم عملية النسخ العكسي لإنتاج cDNA ثم يفك من mRNA الذي عمل قالباً لتخليقه، ثم تستعمل أشرطة cDNA لتخليق الأشرطة المكملة لها وتربط إلى نواقل ملائمة يتم تكثيرها في البكتيريا. ويمكن استعمال cDNA لتحديد تواليات البروتينات التي تنتجها باعتبارها تمثل اكسونات الجينات القابلة للتشفير.

مسح مكتبات DNA

يكون المسح بالبحث عن البكتريات التي التقطت النواقل الحاوية على القطع المطلوبة بتميمتها على إطباق بتري لإعطاء المستعمرات، وتؤخذ لها طبعة Replica imprint على أغشية، ثم يتم نسخ DNA الموجود على الأغشية بالمحاليل القلوية بعد التخلص من الخلايا، ثم يتم البحث عن قطعة DNA المطلوبة بعملية التهجين باستعمال مواد مشعة أو غيرها من المجسات المعلمة، وبعد التهجين تظهر الإشارات على الأغشية في المواقع الحاوية على قطعة الجين أي القطعة المكملة للمجس ومنها يمكن تحديد المستعمرة على الطبق الحاوية على الجين المطلوب كما موضحة في الشكل 104.4 :



شكل 104.3 : مسح مكتبات DNA

الطرق المتعلقة بـ RNA

تعتمد الطرق الخاصة بقياس RNA أو الكشف عنه على تركيب الجزيئة اذ وكما مر سابقا هناك أنواع من جزيئات RNA ذات مواصفات خاصة ، والبعض منها يمكن ان يستعمل لأغراض خاصة وليس للكشف. وبصورة عامة يكون الكشف عن RNA باستعمال وصمة نورذن .

ومعظم الطرق تعتمد على تحضير cDNA من mRNA وهذه الطريقة لها أغراضها الخاصة ولعل في مقدمتها التخلص من الانترونات والتركيز على الاكسونات لتحديد بعض المؤشرات مثل نوع النسيج أو لغيرها من الأغراض. وفي هذه الطريقة يؤخذ النسيج المطلوب دراسته ويتم عزل mRNA منه ويتم نسخ DNA منه باستعمال إنزيم النسخ العكسي بعد التزويد ببوادئ خاصة. وبما ان جزيئات mRNA تحوي على ذيل الأدينين (Poly (A عند النهاية 3' فيمكن استعمال بادئ من Poly (T) عندها يستطيع إنزيم النسخ العكسي تخليق cDNA بدءاً من 5' بالاتجاه الى 3' ثم تتم إزالة RNA باستعمال إنزيم Ribonuclease، ويستعمل cDNA (شريط مفرد) كقالب لبناء الشريط المقابل باستعمال إنزيم الكوثر DNAP ، ويمكن بتحليل القطع الناتجة أو حتى جزيئات mRNA معرفة أي من الجينات أو أي من اكسونات الجينات هي النشطة في النسيج المستعمل.

تقنيات المصفوفات الدقيقة Microarray technology

تسمى أحيانا رقائق الجين Gene chips اعتمادا على المادة المراد دراستها. تستعمل هذه التقنية في الأوقات الحالية لدراسة مكنون النسخ في الخلايا أو بصورة أدق لدراسة التعبير الجيني في الخلايا المتمثل بـ mRNA والتي لا تترجم كلها، لذلك فان هذه الطريقة لا يمكن اعتمادها في تفسير نتائج ما بعدها أي نوعية البروتينات التي يمكن ان تنتج . الذي يميز الطريقة هي قابليتها للكشف عن التعبير لآلاف الجينات في الوقت الواحد والتي كانت غير ممكنة في السابق. وتختلف الطرق المستعملة سواء كانت للكشف عن مكنون النسخ أو عن الجينات أو غيرها من الأغراض إلا ان الأساس واحد وهو استعمال مواد سائدة مثل الأغشية أو الشرائح الزجاجية الخاصة التي قد تكون مغطاة بمواد معينة، ويتم تقييد المجسات بواسطة روابط تساهمية الى المادة السائدة ، ويكون ربط المجسات في بقع منتظمة، والمجسات هنا تكون بشكل أشرطة مفردة للمفردات الحيوية المراد الكشف عنها ولذلك في حالة وجود هذه المفردات في النماذج التي تضاف لهذه الشرائح وتكون مكملتها لها سوف ترتبط بها. وطريقة الكشف تعتمد على ان الجزيئات المراد الكشف عنها (النموذج) تعلم بصبغات متفلورة والتهجين على الشرائح سوف يعطي إشارات خاصة بحدوث التهجين، وشدة الإشارة أو التفلور لكل بقعة تعكس مدى التعبير للمواد في النموذج، ويمكن إجراء المقارنات من حساب التغيير في التعبير الجيني والاستعانة بالحاسوب.

وللطريقة الكثير من التحويرات تجعلها مناسبة للعديد من الاحتياجات في مجال الدراسات الحيوية.

الطرق الخاصة بالبروتينات

الطرق الخاصة بتحليل البروتينات والكشف عنها كثيرة منها ما هو قديم ومنها ما تطور من الطرق التقليدية (وقد ذكر بعضها سلفاً). فمثلاً الطريقة التقليدية وصمة ويسترن Western blot والمسماة Immunoblotting تستعمل للكشف عن البروتينات باستعمال الأجسام المضادة . وهناك طرق الترحيل الكهربائي المختلفة سواء الترحيل باتجاه واحد لتحديد الحزم البروتينية أو الترحيل باتجاهين باعتماد مؤشرات مختلفة 2D-gel electrophoresis التي يمكن أن تستعمل في رسم خرائط الببتيدات. فضلاً عن استعمال طرق الكروماتوغرافي بأنواعها المختلفة وأسسها المختلفة.

والطرق الحديثة تعتمد على استعمال رقائق البروتين بشكل مشابه لرقائق DNA ولكن باختلافات جوهرية، فالمساند الصلبة المستعملة يجب أن تغطي ببروتينات خاصة أو رباطات التي تستطيع قنص البروتينات المكمل لها. أو استعمال الطرق المناعية بتوفير الأجسام المضادة لكل البروتينات التي تؤلف المكون البروتيني ولكن المعوقات في هذه الطريقة أن الأجسام المضادة الطبيعية يحصل لها مسخ سريع ويمكن أن يكون هناك مجالا لاستعمال الأجسام النانوية Nanobodies وهي شكل مصغر للأجسام المضادة وتحتوي على الجزء القابل للارتباط بالمستضدات فضلاً عن ميلها للتداخل مع مستضدات غير متخصصة، كما أنه من الصعوبة إنتاج أجسام مضادة لكل بروتين في الكائن الحي وكذلك كون هذه العملية مكلفة.

وللتغلب على مشكلة الأجسام المضادة ومشاكلها تم تطوير تقنيات جديدة منها استعمال منصات البروتين Protein scaffolds لاقتناص البروتينات المستهدفة. والمنصات هذه مشابهة للأجسام المضادة ولكنها أصغر وأكثر ثباتاً وتخصصاً في ارتباطها للبروتينات المستهدفة ، ويمكن أن تصنع دون الحاجة إلى الخلايا الحية، وترتبط إلى صبغات متفلورة، ويعتمد الكشف في هذه الحالة على نقل طاقة التهيج لجزيئات الصبغات المتفلورة وتعتمد كفاءة نقل الطاقة على الاختلاف بين الصبغات وتستعمل أجهزة خاصة للتعامل معها.

ومن الطرق الأخرى هي طمع البروتين وهي الأساس لطريقة الطمع الجزيئي Molecular printing method وفيها تستعمل مواد كيميائية بسيطة لتغطية البروتين المستهدف وبعد حصول التغطية الكاملة أو الكثرة وإزالة البروتين المستهدف فإن القالب الناتج يشبه البروتين المستهدف وهذه القوالب تستعمل لاقتناص الجزيئات المشابهة لها بتخصص عالي. وهناك العديد من التقنيات الأخرى في دورها للتطوير ويمكن أن تعمل بشكل مشابه لرقائق أو المصفوفات الدقيقة للـ DNA ويمكن تحليل صورها وعنقدها، فضلاً عن إمكانية إيجاد البرامج الخاصة بها لتحديد البروتينات التي يتم تنظيمها بشكل متزامن.

وتستعمل طرق مختبرية مختلفة للتعرف على البروتينات وصفاتها ومن أهمها:

الترحيل الكهربائي الهلامي Gel electrophoresis

توجد عدة طرق لإجراء الترحيل الكهربائي والطرق التي تبحث في تفاصيل دقيقة للبروتينات تشمل الترحيل الكهربائي ثنائي الاتجاه وهي طريقة ذات كفاءة عالية في فصل البروتينات بالاعتماد على الشحنات أو الكتلة. فمثلاً يتم ترحيل البروتينات في هلام متدرج قيم الرقم الهيدروجيني باتجاه واحد تحت ظروف غير ماسخة لفصل البروتينات اعتماداً على نقاط تعادلها الكهربائية Isoelectric points (pI) ثم ترحيل النواتج المفصولة بشكل متعامد تحت ظروف

المسخ لفصل البروتينات اعتماداً على الوزن الجزيئي، ومن ثم استعمال أملاح الفضة للتصبيغ لأنها أكثر طرق التصبيغ حساسية لتوضيح موقع البروتينات على الهلام التي تظهر بشكل خريطة ثنائية الاتجاه، والبقع الناتجة تمثل أحد البروتينات التي تم التعبير عنه، ومثل هذه الصور تسمح لتحليلها. ولكن في العموم ليس كل البروتينات يمكن أن تفصل عن بعضها وتصبغ بهذه الطريقة بشكل مقبول خاصة بروتينات الأغشية الكارهة للماء والتي لا تذوب بسهولة وتميل إلى التكتل في المحاليل المائية لهلام الترحيل مثل بروتينات الأغشية، وللتغلب على هذه المشكلة يمكن أن تجزأ بروتينات الغشاء باستعمال طرق خاصة ودواريء ترحيل خاصة.

وعملية التجزئة للبروتينات لغرض ترحيلها تستعمل مع بروتينات أخرى وليست البروتينات الغشائية فقط مثل فصل الأجزاء النووية وبروتينات الهيكل الخلوي وغيرها. في العموم فإن تحليل الصور الناتجة تمثل خطوة متقدمة بعد الترحيل الكهربائي التي تساعد في توضيح نمط التعبير البروتيني وتوجد العديد من البرامج الخاصة بهذه الأغراض التي سيتم تناولها في الفصول القادمة.

تحديد توالي الحوامض الأمينية

من الطرق الأساسية المستعملة في هذه الحالة هي طريقة Edman degradation أو Tandem mass spectrometry. وتحتاج عملية تحديد توالي الحوامض الأمينية إلى الانتباه وذلك أن بعض التحويرات التي تجري على البروتينات بعد الترجمة مثل الفسفرة وتكوين الجسور الكبريتيدية الثنائية أو إضافة السكريات يمكن أن تعد جزءاً من التركيب الأولي ولكنها لا يمكن أن تقرأ من توالي الجين. فضلاً عن الأخذ بنظر الاعتبار الجوانب الأخرى مثل انتظام سلسلة الحوامض الأمينية في تراكيب ثانوية محددة مثل حلزونات ألفا أو صفائح بيتا.

تحديد وظائف البروتينات

تتم هذه في الوقت الحاضر باستعمال البرامج الخاصة وذلك لأن الطرق التجريبية قد تعد متخلفة نوعاً ما مقارنة بالطرق السريعة Throughput وسيتم تناولها في الفصول القادمة.

تداخلات البروتينات

يتم الكشف عن التداخلات بالطرق التقليدية مثل استعمال Yeast-two-hybridization وفيها يتم تخليق بروتينات ذات مواصفات خاصة وفيها دومينات قادرة على الارتباط بالبروتينات الأخرى مثل عوامل الانتساخ و DNA مما يؤدي إلى التعبير عن جين الإعلان. وهذه الطريقة وإن لاقت شيئاً من الاستحسان وتم إنشاء مستلزماتها إلا أنها تعاني من بعض العجز وذلك لأنها طريقة غير مباشرة وتعطي نتائج موجبة غير حقيقية لتداخلات كاذبة التي قد تصل إلى 50% من النتائج.

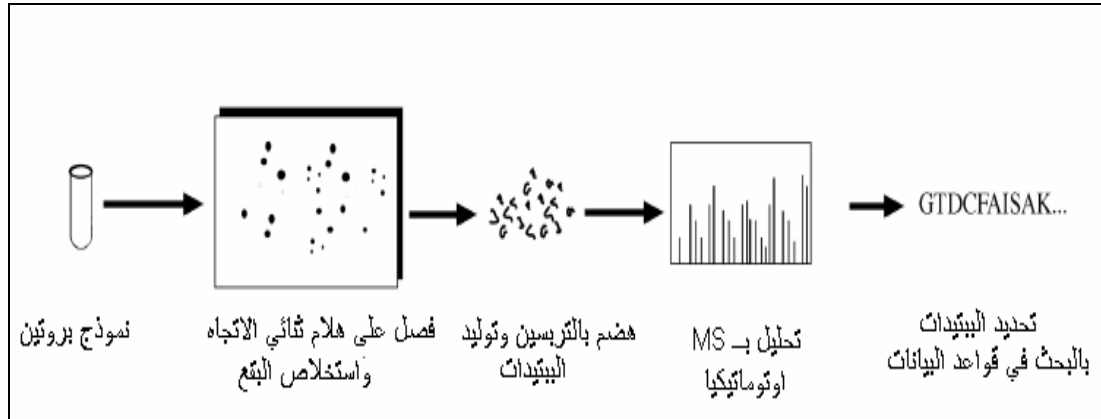
ولذلك فإن بدائل هذه الطريقة هي الأفضل ومنها استعمال طرق تنقية البروتين وطرق كروماتوغرافي الألفة وبالتالي تعريض البروتينات النقية إلى التحليل بطرق أخرى مثل المطياف الكتلي لتحديد المكونات المتداخلة. وكذلك استعمال طرق المصفوفات البروتينية الدقيقة التي تمثل الطرق السريعة وتساعد هذه في التقليل من النتائج الموجبة الكاذبة والسالبة الكاذبة. وفي العموم يكون استعمال أكثر من طريقة الحل الملائم لتقليل النتائج الكاذبة.

تحديد التركيب البروتيني

يمكن تحديد التركيب العام للبروتينات بعد عزلها على هلام ثنائي الأبعاد اي باستخدام أكثر من صفة لعمليات فصل البروتينات ثم الانتقال الى توصيفها أكثر باستعمال المطياف الكتلي Mass spectroscopy ، اذ بعد عمليات الفصل على الهلام يتم استقطاع البروتينات وضمها بالبروتينات مثل التربسين او غيرها، فعملية الهضم تؤدي الى إنتاج نمط فريد من قطع الببتيدات باوزان جزيئة مختلفة يطلق عليها بصمة الببتيد Peptide fingerprint والأخيرة هي التي يمكن ان تحلل بالمطياف الكتلي لتحديد عدد من مواصفاتها وتوجد في الوقت الحاضر طرق متطورة.

تقنيات تحليل التعبير البروتيني

تُعنى هذه الطرق بتوصيف البروتين بعد تخليقه اي بعد مرحلة التعبير الجيني وتهدف الى تحديد المؤشرات التقليدية مثل كمية البروتينات، وكذلك فصل البروتينات بالاعتماد على طرق استخدام الهلام ثم تحليل الصور للحزم الناتجة ثم يلي ذلك تحديد بصمة الببتيد وتحديد تواليات الحوامض الامينية فيه ثم في التالي التعرف على البروتين بالبحث عن مشابهاة في قواعد البيانات ويمكن تلخيص العملية بالشكل 105.3 :



شكل 105.3 : تقنيات تحليل التعبير البروتيني

فرز وتوزيع البروتينات

يعد وجود البروتينات في أماكن خاصة في الخلية أساسا لوظيفتها ولذا تقوم البروتينات بوظائفها بعد تخليقها ونقلها. ودراسة آليات فرز وتوجيه البروتينات الى مواقعها يقع ضمن الحقل الدراسي " توجيه البروتينات " والذي أصبح من المواضيع الأساسية في دراسة بايولوجية الخلية الحديثة. ويعد تحديد الموقع من الضروريات للإعلان عن الوظيفة اذ يقلل من احتمالات التنبؤ.

وفي الخلايا حقيقية النواة فان البروتينات يجب ان تنقل الى أجزاء محاطة بالأغشية نظرا لوجود التراكيب المميزة في داخل الخلية مثل المايكوندريا والبلاستيدات (في حالة الخلايا النباتية) ، والنواة والأجسام البيروكسيمية. ولحمل البروتين الى المكان المخصص له يحتاج الى ببتيدات إشارة خاصة موجودة على البروتينات المخلقة حديثا والتي تحمل شفرة المنطقة او رقم المنطقة

Zip code التي توجه البروتين الى الجزء المخصص له، وبعد نقل البروتين الى الجزء او العضية الخاصة به يحدث فصل لببتيد الإشارة بواسطة بروتينات خاصة لإنتاج البروتين الناضج الفعال. وهذه يمكن ان تمثل أحد التحويلات التي تحدث بعد عملية الترجمة والعملية تحدث في الخلايا حقيقية النواة وبدائية النواة على حد سواء. وفي بدائية النواة يتم توجيه البروتينات الى الغشاء الداخلي او الخارجي او منطقة الفسحة المحيطة.

وتواليات الإشارة هذه تكون في العادة غير متشابهة وتختلف في الطول بشكل كبير جدا وتختلف في الموقع ولكن هناك بعض المظاهر العامة فيها ومنها ان كل تواليات الإشارة تحوي على منطقة مركزية كارهة للماء تسبقها 1-2 من الثمالات المشحونة بشحنة موجبة (في الأغلبية) ، ولكن مثلا في تواليات الإشارة الخاصة بالبروتينات المتوجهة الى المايكوكوندريا تتراوح أطوالها بين 20-80 ثمالة وتقع على الطرف الاميني من البروتين وتكون غنية بالثمالات المشحونة بشحنة موجبة مثل الارجنين وكذلك الثمالات الحاوية على الهيدروكسيل لكل من السيرين والثريونين وخالية من الثمالات السالبة وتميل لتكوين حلزونات ألفا ثنائية الصفات Amphiphilic α -helices وينفلق التوالي عندما يصبح البروتين داخل المايكوكوندريا. اما تواليات الإشارة في البلاستيدات الخضراء وتسمى الببتيدات العابرة Transit peptide فتقع على الطرف الاميني من البروتين وطولها بين 25-100 ثمالة، تحوي القليل من الشحنات السالبة ولكن تحوي عددا "كبيرا" من الثمالات الحاوية على الهيدروكسيل مثل السيرين وتكون هذه القطع مجزأة وكل جزء يقوم بوظيفة خاصة.

اما توجيه البروتينات الى النواة فتكون بببتيدات إشاراتها بطول يتراوح بين 7-41 ثمالة وتوجد في المناطق الداخلية للبروتين وليس عند الأطراف وتحوي على قطيفة متوافقة Consensus motif هي K(K/R)X(K/P) ، وتواليات الإشارة هذه لا تنفلق بعد دخولها الى النواة وانما تبقى في البروتين. وكل هذه المواصفات تؤخذ بنظر الاعتبار للتنبؤ في البرامج التي سيتم ذكرها في فصول قادمة.

الفصل الرابع

الانترنت والإحصاء وتطبيقاته في الحاسوب

المعلوماتية الحيوية والانترنت
جبر المصفوفات
مفاهيم إحصائية
معنوية الاختبارات الإحصائية
سلاسل ماركوف
الشبكات العصبية الاصطناعية
التحليل العنقودي

الانترنت والإحصاء وتطبيقاته في الحاسوب

يعد الحاسوب الأداة الأكثر أهمية للباحثين والمهنيين في مجال المعلوماتية الحيوية، وقد كانت هذه المهمة منوطة لغاية سنوات قليلة باستخدام جهاز حاسوب كبير (له عدة محطات طرفية) Mainframe computers لإجراء الحسابات والبحث في قواعد البيانات في مجال المعلوماتية. ونظرا للتطور في قابلية الحواسيب الشخصية الحديثة Personal computers (PC) وإمكانية إجراء الكثير من المهام بالحواسيب الشخصية، فإن الحاجة إلى استخدام أجهزة الحاسوب الكبيرة انحصرت فقط عند التعامل مع كميات كبيرة من المعلومات، على الرغم من أن أنظمة التشغيل Mac OS و Windows تستخدم بصورة رئيسة في المنزل والمكتب ... الخ، فإن نظام التشغيل UNIX يستخدم بصورة عامة في مجال المعلوماتية الحيوية. وبانطلاق الحاسوب الشخصي الأول في أوائل السبعينيات، فإن العديد من أنظمة التشغيل تم تطويرها. في العام 1974 قدم نظام التشغيل CP/M للبحوث ، الذي كان واحدا من أوائل أنظمة التشغيل للعديد من المعالجات البسيطة . في بداية الثمانينيات جهزت شركة IBM منتجها الجديد من الحواسيب الشخصية IBM PC بنظام تشغيل مايكروسوفت Microsoft operating system. (MS DOS) ، ويقتصر نظام التشغيل MS DOS على سطر الأوامر (Command line) ، الذي يتم التعامل من خلاله بين المستخدم للحاسوب والجهاز. وبخط موازي انبثقت فكرة تنفيذ بيئة المستخدم التصويرية Graphical user interface (GUI) في العام 1983، وقد سوقت شركة Apple Lisa أول حاسبة شخصية تعمل على نظام تشغيل يستخدم بيئة المستخدم التصويري للتفاعل مع المستخدم . ونتيجة لاستخدام بيئة المستخدم التصويري أدى إلى نجاح حاسوب Apple Macintosh وأدركت الشركة فائدة بيئة المستخدم التصويرية وبدأت بتطوير نظام تشغيلي مماثل . ظهر نظام التشغيل Windows (نوافذ) في الأسواق في منتصف الثمانينيات، ولاقى الإصدار Windows 3.0 نجاحا كبيرا في بداية السبعينيات. كان نظام التشغيل Windows 3.0 مقتصرا على مستخدم واحد فقط ، لكن قدم الإصدار Windows 3.11 وظائف الشبكة. وبانطلاق نظام التشغيل Windows 95 شرعت شركة مايكروسوفت بتطوير نظام تشغيل Windows جديد، الذي يحتوي أفكار أنظمة التشغيل UNIX في إصدارات أنظمة التشغيل XP و Windows server 2003 و Windows vista وفي المقابل طورت شركة Apple نظامها التشغيلي Mac OS وتوفر النسخة الحالية Mac OS X نظام تشغيلي UNIX يعتمد على بيئة المستخدم التصويرية متقنة.

المعلوماتية الحيوية والانترنت Bioinformatics and the internet

يمثل علم المعلوماتية الحيوية مجموعة متوافقة من العلوم والتي تنمو باستمرار وذلك باستخدام طرق حسابية (عددية) للإجابة على الأسئلة البايولوجية. وتتطلب الإجابة على هذه الأسئلة الاستفادة من مجموعة البيانات الكبيرة والمعقدة للوصول إلى الاستنتاجات البايولوجية. يلعب حقل المعلوماتية الحيوية مع الاكتشافات الكبيرة للمعلومات الخاصة بالتواليات والتراكيب الخاصة بها المتاحة للباحثين دورا متزايدا في دراسة أساس المشاكل الطبية الحيوية. الهدف أو (التحدي) هو كيفية تقديم المساعدة في اكتشاف الجين وغيرها من المعلومات ، والتجارب من

الأنواع الأخرى التي يمكن ان توجي الى احتمالية وجود علاقات غير معروفة مسبقا نسبة الى تركيب ووظيفة كل من الجين والبروتين. أصبح هذا الهدف نوعا ما مثبط للهمة في ضوء كمية المعلومات الكبيرة الخاصة بمشاريع الجينوم البشري والكائنات الأخرى. وقبل المباشرة بمناقشة الطرق الحسابية لحل المشاكل البايولوجية كان من الضروري تحديد الأساسيات العامة التي تمكن المستخدم من الوصول وتنفيذ الخوارزميات. سيتم في البداية عرض مفهوم الانترنت والمصطلحات الخاصة به والربط الى الانترنت ومجهز الخدمة والمجال الجغرافي ونماذج الاتصالات وتقنيات النقل ومدى التردد والتركيب البنيوي للشبكة وأخيرا إدارة الشبكة.

الشبكة العنكبوتية للانترنت (World Wide Web (WWW

ظهرت شبكة الانترنت (شبكة تبادل ونقل المعلومات) من قبل وزارة الدفاع الأمريكية عام 1969 عن طريق تمويل مشروع من اجل ربط مؤسسات الجيش الأمريكي مع الجامعات والمؤسسات البحثية التي تعمل على أبحاث ممولة من القوات المسلحة، وسميت هذه الشبكة باسم أربا (ARPA) Advance Research Project Agency، صمم المشروع باستخدام طريقة إعادة التوجيه الديناميكي Rerouting Dynamic بحيث انه في حالة فشل (عطل) جزء من الشبكة لا يؤدي الى التوقف التام. ثم أصبحت هذه الشبكة فيما بعد مشهورة جدا خاصة مع استخدام مجال البريد الالكتروني (e-mail). وبجانب كل من مجهزي الخدمة WWW والبريد الالكتروني ظهر مجهز خدمة آخر هو بروتوكول نقل الملفات (FTP) File Transfer Protocol الذي يلعب دورا كبيرا في مجال المعلوماتية الحيوية، والنجاح الذي حققه مجهز الخدمة WWW يرجع الى بساطة وسهولة استخدامه والتوافق بين مختلف بروتوكولات الاتصال.

كان تطوير مجهز الخدمة WWW في الأصل ، لغرض التبادل العلمي للبيانات، ومن ثم اصبح مصدرا مهما في مجال المعلوماتية الحيوية. فقد جهزت قواعد البيانات البايولوجية من خلال مجهز الخدمة WWW العلماء والباحثين بمختلف البيانات في مجال المعلوماتية الحيوية.

الربط الى الانترنت Connecting to the internet

قبل البدء باستخدام جميع الإمكانيات والموارد التي توفرها شبكة الانترنت، لابد من إجراء الربط -فيزيائيا- بين جهاز الحاسوب المستخدم وبين قناة المعلومات الرئيسية The information superhighway وتنقسم عملية الربط الى جزأين منفصلين: الجزء الأول هو عملية الربط الحقيقي (فيزيائي) الذي يشمل الكيبل Wire بدءا من الكيبلات الخاصة بجهاز الحاسوب المستخدم وصولا الى العمود الفقري للشبكة. والجزء الثاني هو مجهز الخدمة، الذي يتعامل مع موضوع الطريق ومحتويات الاتصال في الوقت ذاته .

مجهز الخدمة Providing

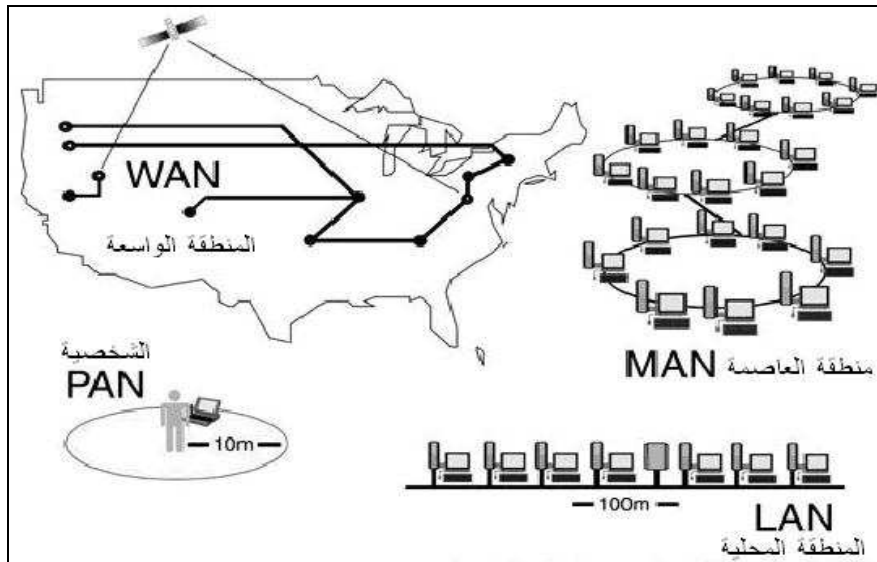
يحتاج مستخدم شبكة الانترنت الى الربط مع مجهز خدمة ويكون على صنفين رئيسيين، الخدمة المباشرة Online services مثل America Online (AOL) الذي يوفر عدداً من الخدمات الرقمية التفاعلية، وبضمنها استرجاع المعلومات والبريد الالكتروني (e-mail)

ولوحة إعلانات وغرف محادثة تمكن المستخدمين الذين هم على الخط مباشرة من التحدث مع بعضهم البعض حول مختلف المواضيع في الوقت نفسه .

ويمكن ان تكون خدمات الانترنت الاتجاه المعاكس (ISPs) Internet service providers فبدلاً من التركيز على حجم وكمية التجهيز فان ISPs تجهز الأدوات الضرورية للمستخدمين لإرسال واستلام الرسائل الالكترونية وتحميل وتنزيل الملفات والتجول في الشبكة العنكبوتية للانترنت WWW لإيجاد المعلومات المطلوبة في المواقع البعيدة . والفائدة الرئيسة هي سرعة الاتصال وفي الغالب فان مجهز خدمات صغير يمكن ان يوفر سرعة اتصال أسرع مما يستطيع نظيره في مجهز الخدمات المباشر .

المجال الجغرافي The geographical scope

يعد المجال الجغرافي للشبكة ذو أهمية كبيرة وله علاقة تحدد بمدى التردد Bandwidth وأمن الشبكة Security ، ومدة الاستجابة Response time ، ونوعية الحسابات المحتملة. وعلى الرغم من ان الحدود الجغرافية تبدو عشوائية إلا انه يمكن ان يعبر عن الشبكات بالعديد من الأنواع منها: شبكات المنطقة الشخصية (Personal area networks) PANs وشبكات المنطقة المحلية (Local area networks) LANs وشبكات المنطقة العاصمة (Metropolitan area networks) MANs وشبكات المنطقة الواسعة (Wide area networks) WANs كما هو موضح بالشكل 1.4:



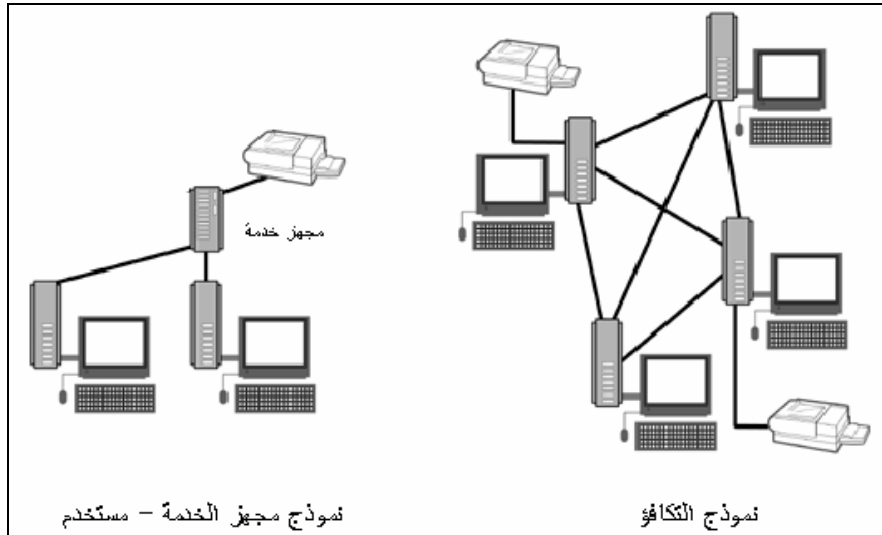
شكل 1.4 : المجالات الجغرافية للتغطية بالانترنت

تؤمن شبكات المنطقة الشخصية للمستخدم منطقة نصف قطرها حوالي 10 متر والتي تؤسس باستخدام تكنولوجيا اللاسلكي والاتصال عن بعد. في حين تمتد شبكات المنطقة المحلية الى حوالي 100 متر من مركز مجهز الخدمة ، او في طابق واحد من احد الأبنية ويبدأ مجال عمل

شبكات منطقة العاصمة من حيث انتهى مجال عمل شبكة المنطقة المحلية (LANs) ، حيث تغطي هذه الشبكة أبنية بأكملها وتمتد الى مسافات تبلغ مئات الكيلومترات. اما شبكات المنطقة الواسعة (WANs) فتتمد عبر مناطق واسعة جدا .

نماذج الاتصالات Communications models

هنالك نوعان من نماذج الاتصالات، الأول هو نموذج جهاز الخدمة - مستخدم Client-server model، حيث يوفر جهاز الخدمة البيانات الى مستخدم او أكثر. اما النموذج الآخر، نموذج التكافؤ Peer to peer model وفيه تعمل كل حاسبة كمجهاز خدمة ومستخدم الى غيرها من الحاسبات على الشبكة، وفي مثل هذه الحالة فان حاسبة ما قد تعمل في لحظة معينة كمجهاز خدمة وفي اللحظة التالية كمستخدم كما موضح في الشكل 2.4 :

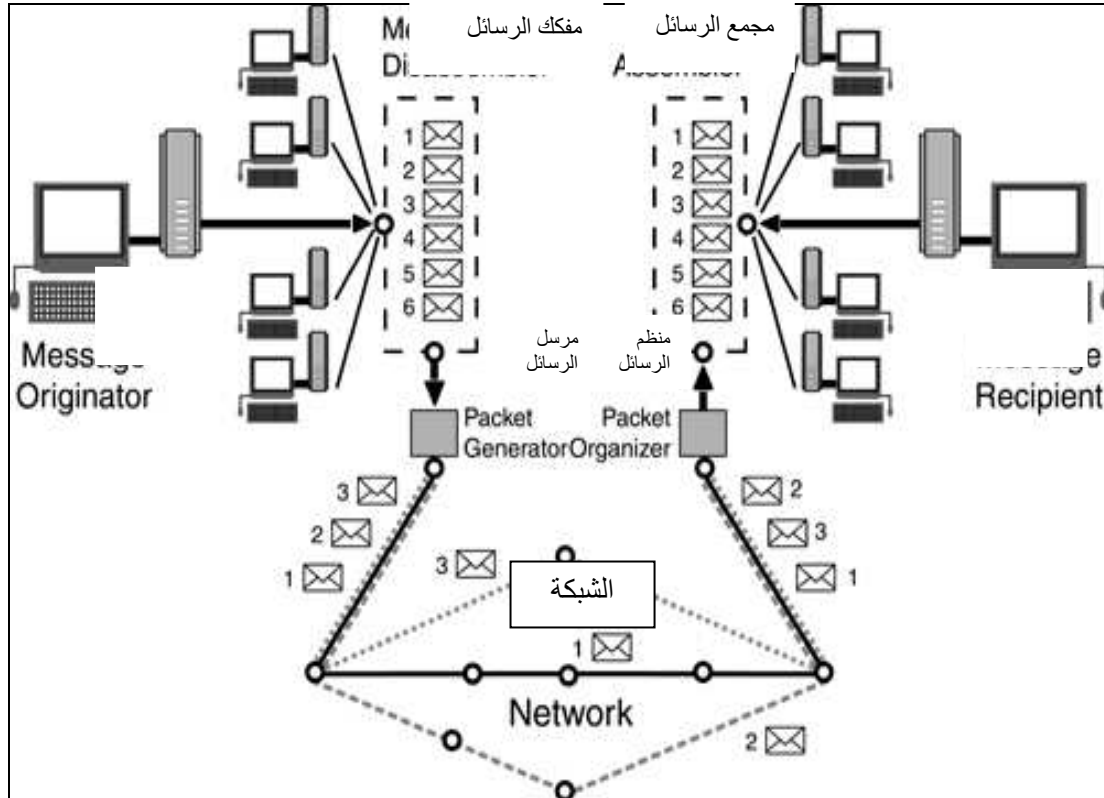


شكل 2.4 : نماذج الاتصالات

وتعد الأخيرة أبسط أنواع شبكات الحاسوب في مجموعة عمل صغيرة. في هذا النموذج تعمل كل محطة على أساس انها جهاز خدمة ومستخدم للمحطات الأخرى . يكمن الجانب السلبي لهذا النوع من الشبكات هو الاستخدام غير المنتظم للموارد . كذلك هناك محددات آخر لهذا التصميم وهو ان إدارة المعلومات في هذا النوع من الشبكات تواجه صعوبات كبيرة ، على سبيل المثال فان المستخدم الذي يعمل في مجموعة عمل معينة عليه ان يعمل أرشفة خاصة لبياناته لضمان سرية معلوماته ، كذلك عليه ان يحدث باستمرار الأنظمة المضادة للفيروسات على جهازه. في المقابل يوفر نموذج جهاز الخدمة - المستخدم جهاز للخدمة مركزي برامج وملفات البيانات التي يمكن للمستخدم ان يدخلها من محطات عمل المستخدم على الشبكة. وتكمن فائدة هذا النوع ان نظام تشغيل الشبكة الموجود على جهاز الخدمة يوفر أمان معين للشبكة والمستخدمين فيها. حيث يسمح للمستخدم فقط بالدخول على الملفات والبرامج الموجودة.

تقنيات النقل Transmission technology

ان فكرة تطوير بدائل نموذج الشبكة التي تعمل على المفتاح التقليدي واحدة من الأساليب المبتكرة لتكنولوجيا الاتصالات، وكان استخدام شبكة الاتصال هذه حوالي القرن التاسع عشر. والنقطة السلبية الرئيسة لنموذج شبكة الاتصال التي تعمل على المفتاح التقليدي هو عدم إمكانيةها توفير المزيد من قنوات الاتصال زيادة عن عدد المفاتيح المتوفرة. اما النموذج الآخر فهو نموذج الحزمة ، وفيه تجزأ الرسائل الى قطع صغيرة بواسطة مفكك الرسائل Message disassembler (الشكل 3.4). ثم ترحل هذه الحزم بواسطة طرق مختلفة من خلال الشبكة الى الجهة المقصودة، وتكون محدودة بكثافة المرور عند كل عقدة على طول الطريق. ولكون الحزم ترسل بصورة ديناميكية الى العديد من العقد المختلفة كوظيفة الطالب الأنى لعنوان الاتصال الفعلية، فان بعض الحزم يمكن ان تتقدم على البعض الآخر وخارج تسلسلها الصحيح . عند الجهة المقصودة تجمع حزم البيانات المرسله بواسطة مجمع الرسائل Message assembler الى ترتيبها الأصلي. وبهذا فان المشترك يمكن ان يستلم الرسائل في أي وقت بدون ان يكون هناك توقف ومنع الوصول له من قبل قنوات الاتصالات عند إجراء اتصالات أخرى من قبل مشتركين آخرين وهذا مغاير لنموذج شبكة الاتصال بواسطة المفتاح والتي تحجز وتقيد القناة حتى ينتهي عمل المشترك السابق او الأول ويحررها، والتي تبقى مقيدة حتى في حالة عدم وجود بيانات مرسله.



شكل 3.4 : عملية نقل الرسائل في الشبكة

البروتوكولات The protocols

ان واحدة من الفوائد الأساسية لشبكة الحاسوب هو مشاركة العديد من أنظمة التشغيل المختلفة التي تعمل على مختلف الحواسيب والبيانات والموارد الأخرى من خلال الشبكة. إضافة الى ذلك ازدياد الأجهزة التي يمكن إجراء الاتصالات فيما بينها من خلال الشبكة مما يؤدي الى ازدياد أهمية هذه الشبكة.

المنظمات المعيارية بروتوكولات الشبكة الموجودة في المجاميع المتداخلة للأنظمة المفتوحة OSI Open system interconnection (OSI) منها على سبيل المثال المعهد الكهربائي الإلكتروني الهندي The Institute of electrical and electronics engineers (IEEE) والمعهد الوطني الأمريكي The American national standards institute (ANSI) وغيرها هي التي عملت على التعريف بالبروتوكولات والاتفاق عليها.

بدأت المجاميع المتداخلة للأنظمة المفتوحة (OSI) في أواخر السبعينيات من القرن الماضي، حيث عرفت معمارية الاتصالات ذات المستويات العالية، متضمنة نموذج (OSI) (جدول 1.4). يعرف هذا النموذج ابتداء من الناقل الفيزيائي الى دلالات والألفاظ الخاصة بالرسائل على الشبكة.

جدول 1.4: طبقات عمليات الاتصال بالأنظمة المفتوحة OSI

الطبقة	الاسم	الفعالية
7	التطبيق	علم الدلالات والألفاظ
6	التمثيل	تدقيق العمل
5	جلسة	تنسيق الحوار
4	النقل	نقل البيانات المعتمدة
3	الشبكة	تحديد الطريق والتتابع
2	ربط البيانات	النقل الخاص-للتكنولوجيا
1	الفيزيائي	الربط الفيزيائي

مدى التردد Bandwidth

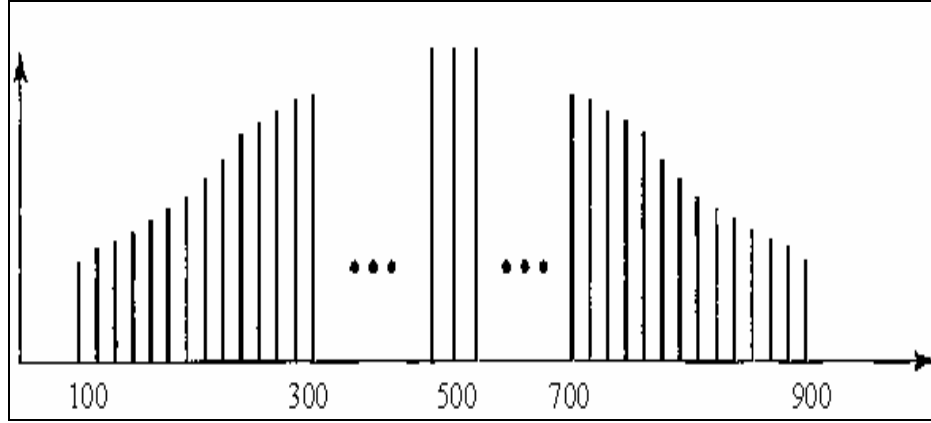
عبارة عن الفرق بين أعلى تكرار وأوطأ تكرار للإشارة المركبة. ويمكن أن تقاس أيضا بقدرة الشبكة على نقل المعلومات. على سبيل المثال، إشارة دورية تحللت إلى خمس موجات بترددات مقدارها 100، 300، 500، 700، 900 (كيلو هرتز) KHz (شكل 4.4). يمكن حساب مدى التردد Bandwidth كالآتي:

$$B = f_h - f_L = 900 - 100 = 800 \text{ KHz}$$

حيث أن:

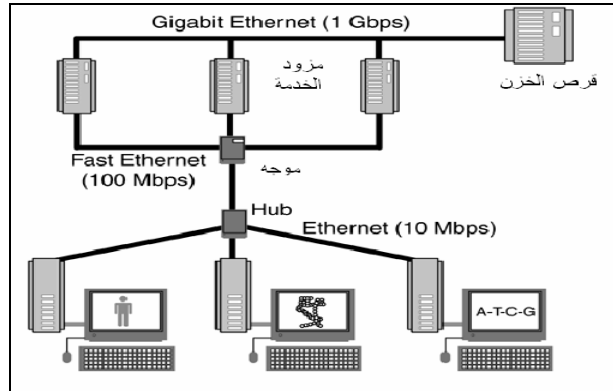
f_h : مقدار تكرار العالي.

f_L : مقدار تكرار الواطي



شكل 4.4 : مدى التردد (KHz)

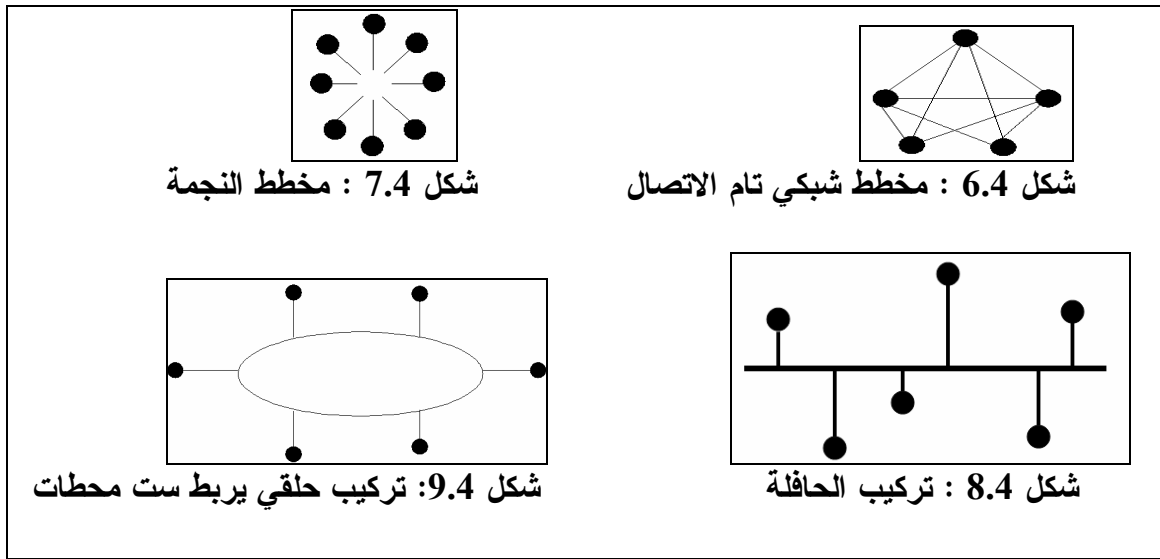
ويشير المصطلح أعلاه (مدى التردد) أيضا الى عدد من البتات (bits) لكل ثانية ترمز الى ما يمكن ان تنقله قناة او شبكة من المعلومات. يوضح الشكل (5.4) تطبيق عملي لاسترجاع المعلومات المخزونة على قرص الخزن Storage area network disk array من خلال شبكة مجهزة بنظام شبكة الصفوف المتضمن شبكة النقل عبر الأثير Gigabit Ethernet . حيث يوفر هذا البروتوكول 1Gbps لنقل المعلومات من المصدر الرئيسي للبيانات الى جهاز الخدمة. في حين توفر شبكة النقل عبر الأثير السريعة (Fast ethernet) ما مقداره 100Mbps لنقل المعلومات بين جهزي الخدمة والتطبيقات المختلفة الأخرى وأخيرا توفر شبكة النقل عبر الأثير المعيارية (Standard ethernet) ما مقداره 10 Mbps لنقل المعلومات بين المحطات الطرفية.



شكل 5.4 : سعة نقل المعلومات عبر شبكة الانترنت

التركيب البنوي للشبكة Topology

يشير التخطيط الفيزيائي للشبكة إلى التركيب البنوي لها (Topology) استنادا إلى الوظيفة العملية التي يمكن أن يؤديها هذا التركيب والتي تكون مفروضة عليه من قبل البيئة المحيطة به والبروتوكولات الواجب توفرها وكلفة التنصيب . هنالك أربعة أنواع أساسية محتملة لتراكيب الشبكة: التركيب الشبكي Mesh (شكل 6.4) والتركيب النجمي Star (شكل 7.4) وتركيب الحافلة Bus (شكل 8.4) والتركيب الحلقي Ring (شكل 6.4) كما موضح في الآتي :



التركيب الشبكي Mesh topology

يرتبط كل جهاز في التركيب الشبكي (شكل 6.4) ارتباط الجهاز المتبادل مع كل الأجهزة الموجودة في هذه الشبكة. ويوفر الارتباط حركة مرور بين الجهازين المرتبطين فقط. يوفر هذا النوع من الشبكات عدة فوائد ، أولها ضمان أن الرابط (الذي يربط بين جهازين حاسوبيين) يحمل البيانات الخاصة به وهذا يزيل مشاكل المرور التي تحدث عندما يكون هناك اشتراك أكثر من جهاز بالخط نفسه . ثانيا تعد هذه الشبكة قوية وفعالة، ففي حالة كون أحد الروابط غير مستخدم فإن ذلك لا يؤثر في مقدرة النظام الكلية في العمل. وهناك فائدة ثالثة وهي الخصوصية والسرية التي يمكن أن تنتج عن هذا النوع من الشبكات ، فعندما تنتقل الرسائل داخل الشبكة فإنها تذهب إلى الجهة المطلوبة حيث تمنع الحدود الفيزيائية المستخدمين الآخرين من إمكانية الحصول على هذه الرسائل. أخيرا هناك إمكانية أن يكتشف مدير الشبكة مكان العطل الذي يمكن أن يحدث وتحديد أسبابه والطول الخاصة به . أما العائق الرئيس لهذا النوع (التركيب) هو كمية الأسلاك وأعداد نقاط التغذية للمدخلات والمخرجات المطلوبة.

التركيب النجمي Star topology

في التركيب النجمي (شكل 7.4) يرتبط كل جهاز بجهاز السيطرة المركزي ، الذي يطلق عليه بالمحور Hub وان هذه الأجهزة لا ترتبط بعضها مع البعض الآخر. على النقيض من التركيب

الشبكي، فإن التركيب النجمي لا يسمح بالمرور المباشر (للمعلومات والبيانات) بين الأجهزة المرتبطة داخل الشبكة. حيث يعمل جهاز السيطرة كمبدل، فإذا كانت هناك رغبة في إرسال البيانات والمعلومات من أحد الأجهزة إلى جهاز آخر داخل الشبكة، فإن عملية الانتقال تبدأ من الجهاز الأول إلى جهاز السيطرة ثم تمر إلى الجهاز المعني. إن كلفة التركيب النجمي أقل من التركيب الشبكي حيث إن كل جهاز في هذا التركيب يحتاج إلى نقطة تغذية واحدة للمدخل / والمخرج لربطه مع أي عدد من الأجهزة الأخرى الموجودة في هذه الشبكة. الفائدة الأخرى لهذا التركيب هو أنه في حالة عدم تأمين الاتصال بأحد هذه الأجهزة فإنه لا يؤثر في ارتباط بقية الأجهزة. يكمن العائق الرئيس لهذا التركيب هو في اعتماد التركيب بأكمله على نقطة واحدة، وهو المحور Hub ، وفي حالة عطل المحور فإن النظام بأكمله يتوقف.

تركيب الحافلة Bus topology

وصفت الأمثلة التركيبية السابقة على أن اتصالها بشكل نقطة-إلى-نقطة. في حين نجد أن اتصال تركيب الحافلة يكون على شكل نقاط متعددة، يتكون هذا التركيب من سلك طويل يمثل العمود الفقري لربط جميع الأجهزة الموجودة في الشبكة (شكل 8.4). ترتبط العقد مع سلك تركيب الحافلة بواسطة السلك النازل من المحطة (الجهاز) إلى نقطة التفرع التي تربط المحطة مع السلك الرئيس. في هذا التركيب تضعف عملية نقل المعلومات داخل الشبكة كلما بعدت المسافة لهذا السبب هناك محددات معينة لعدد التفرعات المسموح بها داخل هذا التركيب ومقدار المسافة بين هذه التفرعات. ومن مزاياه سهولة تنصيبه، وقلة استخدامه للأسلاك مقارنة مع التركيب النجمي والشبكي. أما عوائق هذا التركيب فهي صعوبة إعادة الربط، كذلك هناك صعوبة في إضافة أجهزة أخرى لهذا التركيب.

التركيب الحلقي Ring topology

في هذا التركيب يتصل كل جهاز على شكل نقطة-إلى-نقطة مع الجهازين الذين يقعان على طرفيه. ثم تمر الإشارة على طول الحلقة باتجاه واحد. من جهاز إلى جهاز آخر حتى تصل إلى الجهاز المطلوب، ويتصل كل جهاز في هذا التركيب مع معيد (Repeater)، فعندما يستلم الجهاز إشارة موجهة إلى جهاز آخر فإنه يعاد إرسالها مرة أخرى (Routers). هذا التركيب يمتاز بسهولة نسبية في التنصيب وإعادة التشكيل، حيث يربط كل جهاز إلى جهازين متجاورين (شكل 9.4). وفي حالة إضافة أو حذف جهاز يتطلب ذلك تغيير اثنين من الروابط فقط، وفي حالة وجود عطل في أي جهاز سوف يطلق تحذير بذلك ويمكن معالجته. والمعوقات إن عطل (توقف) أي جزء يؤدي إلى عطل الشبكة بكاملها.

أدارة الشبكة

بعد تأسيس شبكة الانترنت، لابد من إدارة هذه الشبكة لأدراك القوة الكاملة لها. تتضمن عملية إدارة الشبكة والتزود بالاحتياجات اللازمة لتغطية الأزمات الممكن حصولها، التحميل المتوازن، إدارة مدى التردد والحفاظ على أمن الشبكة. على سبيل المثال تتضمن خطة تغطية (معالجة) الأزمات (والكوارث) التي يمكن أن تحدث للشبكة هو توفير كهرباء للشبكة وتوفير بدائل للأعطال التي تصيب الوسائط الموجودة في الشبكة والتحسب لحالة اشتعال حريق في الشبكة وكذلك اضطراب (عدم استقرار) الكهرباء داخل الشبكة للحصول على التحميل المتوازن، تشير هذه العبارة ضمن سياق إدارة الشبكة، إلى طريقة توزيع حجم البيانات - المطلوب نقلها - عبر

عدة طرق (مسالك) بحيث يمكن نقل اكبر كمية من البيانات والمعلومات خلال النظام بأكمله ، على سبيل المثال، اذا كان هناك طريقين متكافئين في شبكة الانترنت، يحمل احدهما 10 بالمائة من حركة المرور في الشبكة والآخرى 90 بالمائة ، فان الخط الأول غير مستخدم بصورة كاملة من حيث ان الخط الثاني يكون عليه زخم (مروري) كبير .

يتضمن مبدأ التحميل المتوازن على تشكيل موجهات وأجهزة الكترونية أخرى داخل الشبكة بحيث ان حركة المرور داخل الشبكة تكون منفصلة قدر الإمكان عبر أجزاء الشبكة المختلفة والأجهزة الموجودة فيها. لتأمين نقل اكبر ما يمكن نقله من المعلومات خلال الشبكة وبأسرع وقت ممكن . تتضمن إدارة مدى التردد الحصول على تحميل متوازن كمثال زيادة كمية تدفق المعلومات باتجاه (طريق) معين لتحقيق التوازن عند الضرورة لغرض توفير نقل المعلومات خلال الشبكة. وتعتمد إدارة مدى التردد على الكلفة وإمكانية توفر مدى عالي من التردد .

وقد يكون التحدي الأكبر لإدارة الشبكة هو الحفاظ على امن معلوماتها وهذا يستلزم جهد مراقبة الشبكة (الأجهزة المرتبطة بالشبكة) يوميا للتحقق من الفيروسات الجديدة ومنع المتطفلين وغير المسموح لهم بالدخول الى الشبكة.

جبر المصفوفات Matrices algebra

أهمية جبر المصفوفات

تتبع أهمية جبر المصفوفات بارتباطه بالعديد من الأساليب والطرق في الرياضيات والإحصاء واستخدامه في العديد من طرق متعدد المتغيرات. لهذا السبب فمن الضروري الإلمام ببعض الجوانب من العلم في هذا المجال .

جبر المصفوفات والمتجهات

لو افترض إن مصفوفة $A_{m \times n}$ ، وهي عبارة عن مجموعة أعداد مكونة من m هي الصفوف و n من الأعمدة ، وتكتب بالشكل الآتي:

$$A_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

في حالة $m = n$ ، تكون لدينا مصفوفة مربعة . إذا كانت المصفوفة مكونة من عمود واحد، على سبيل المثال

$$\underline{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$$

عندئذ فإن $\underline{c}$ يدعى متجه عمودي. أما إذا كانت المصفوفة مكونة من صف واحد ، على سبيل المثال

$$\underline{r}' = (r_1, r_2, \dots, r_n)$$

عندئذ فإن $\underline{r}$ تسمى متجه صفي .

يمكن إيجاد مبدلة المصفوفة (Transpose of a matrix) بتغيير صف محل عمود. وهكذا فإن مبدلة المصفوفة $A_{m \times n}$ هي $A_{n \times m}$ وتكتب كالآتي:

$$\hat{A} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & & a_{mn} \end{pmatrix}$$

وللتوضيح ، نفترض لدينا المصفوفة $B_{4 \times 3}$ و كالآتي:

$$B_{4 \times 3} = \begin{pmatrix} 5 & 6 & 3 \\ 8 & 7 & 4 \\ 6 & 9 & 1 \\ 2 & 5 & 8 \end{pmatrix}$$

فان مبدلة المصفوفة $B_{3 \times 4}$ تكتب بالشكل الآتي:

$$\hat{B}_{3 \times 4} = \begin{pmatrix} 5 & 8 & 6 & 2 \\ 6 & 7 & 9 & 5 \\ 3 & 4 & 1 & 8 \end{pmatrix}$$

كذلك فان $\underline{c}$ يصبح متجه صفي ويكتب كالآتي:

$$\underline{c} = (c_1, c_2, \dots, c_m)$$

وان $\underline{r}$ يصبح متجه عمودي ويكتب كالآتي:

$$\underline{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

هنالك عدداً من المصفوفات الخاصة والتي سيتم التطرق إلى الجزء المهم منها في هذا الفصل. تتكون المصفوفة الصفرية (Zero matrix) من عدد من العناصر الصفرية، وكالاتي:

$$Z = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{pmatrix}$$

المصفوفة القطرية هي مصفوفة مربعة تمتلك عناصر صفرية باستثناء عناصر القطر الرئيسي، والتي تكون بالشكل الآتي:

$$D = \begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{pmatrix}$$

المصفوفة المتماثلة هي مصفوفة مربعة التي لا تتبدل قيمها عن قيم المصفوفة المبدلة، بعبارة أخرى فإن المصفوفة A لها هذه الخاصية عندما تتحقق فيها $A=A'$. أخيراً، فإن المصفوفة المحايدة (Identity matrix) هي مصفوفة قطرية، وأن جميع قيمها القطرية مساوية للواحد.

$$I = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

تتساوى مصفوفتان فقط إذا تساوت جميع عناصرهما. على سبيل المثال:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

تتساوى هاتين المصفوفتين فقط إذا $a_{11} = b_{11}, \dots, a_{33} = b_{33}$ بعبارة أخرى ، تساوي القيم المتناظرة.

يحدد اثر المصفوفة (Trace of matrix) من جمع العناصر القطرية للمصفوفة، حيث ان:

$$\text{tr}(A) = a_{11} + a_{22} + \dots + a_{nn}$$

للمصفوفة المربعة $n \times n$.

عمليات على المصفوفات:

يمكن إجراء العمليات الحسابية الاعتيادية على المصفوفات. ان عمليتي الجمع والطرح يمكن تطبيقها بسهولة على المصفوفات وذلك بإجراء العمليتين على كل عنصر من عناصر المصفوفتين المعنيتين. على سبيل المثال المصفوفتين $A_{3 \times 2}$ و $D_{3 \times 2}$ وكلاهما بالحجم نفسه فان:

$$A_{3 \times 2} + D_{3 \times 2} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} + \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{pmatrix} = \begin{pmatrix} a_{11} + d_{11} & a_{12} + d_{12} \\ a_{21} + d_{21} & a_{22} + d_{22} \\ a_{31} + d_{31} & a_{32} + d_{32} \end{pmatrix}$$

وفي عملية الطرح تكون :

$$A_{3 \times 2} - D_{3 \times 2} = \begin{pmatrix} a_{11} - d_{11} & a_{12} - d_{12} \\ a_{21} - d_{21} & a_{22} - d_{22} \\ a_{31} - d_{31} & a_{32} - d_{32} \end{pmatrix}$$

هاتين العمليتين يمكن أن تطبقان على مصفوفتين من الحجم نفسه فقط. في جبر المصفوفات، وعند ضرب مصفوفة ذات حجم معين برقم عددي معين فان ناتج الضرب يكون مساويا لحاصل ضرب الرقم العددي في جميع عناصر المصفوفة وعلى سبيل المثال:

نفرض ان المصفوفة $A_{3 \times 2}$ فان ناتج حاصل ضربها بالعدد K هو :

$$KA = \begin{pmatrix} Ka_{11} & Ka_{12} \\ Ka_{21} & Ka_{22} \\ Ka_{31} & Ka_{32} \end{pmatrix}$$

أما في حالة ضرب مصفوفتين ، فان الإجراء يختلف اذ ان هنالك شرط واحد لإمكانية إيجاد حاصل ضرب المصفوفتين وهو أن يكون عدد الأعمدة في المصفوفة الأولى مساوي إلى عدد الصفوف في المصفوفة الثانية وعلى سبيل المثال:

المصفوفتين $A_{m \times n}$ و $B_{n \times p}$ فان :

$$A \times B = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1p} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2p} \\ \vdots & & & & \\ b_{n1} & b_{n2} & b_{n3} & \dots & b_{np} \end{pmatrix}$$

فمثلا المصفوفتان $A_{3 \times 2}$ و $B_{2 \times 2}$

$$A_{3 \times 2} = \begin{pmatrix} 4 & 9 \\ -2 & 6 \\ 8 & 5 \end{pmatrix} \quad B_{2 \times 2} = \begin{pmatrix} 5 & 9 \\ -3 & 6 \end{pmatrix}$$

يكون حاصل ضربهما $A_{3 \times 2} \times B_{2 \times 2}$ كالآتي:

$$\begin{aligned} A_{3 \times 2} \times B_{2 \times 2} &= \begin{pmatrix} \xrightarrow{4 \quad 9} \\ \xrightarrow{-2 \quad 6} \\ \xrightarrow{8 \quad 5} \end{pmatrix} \times \begin{pmatrix} \downarrow 5 \quad \downarrow 9 \\ \downarrow -3 \quad \downarrow 6 \end{pmatrix} \\ &= \begin{pmatrix} (4 \times 5 + 9 \times (-3)) & (4 \times 9 + 9 \times 6) \\ ((-2) \times 5 + 6 \times (-3)) & ((-2) \times 9 + 6 \times 6) \\ (8 \times 5 + 5 \times (-3)) & (8 \times 9 + 5 \times 6) \end{pmatrix} \\ &= \begin{pmatrix} -7 & 90 \\ -28 & 18 \\ 25 & 102 \end{pmatrix} \end{aligned}$$

معكوس المصفوفة Matrix inversion

يُنظر مفهوم معكوس المصفوفة في العمليات الرياضية الاعتيادية عملية القسمة ، إذا كان العدد K فإن: $K \times K^{-1} = 1$. وبطريقة مماثلة، إذا كانت A مصفوفة مربعة فإن معكوسها هو A^{-1} حيث إن $A \times A^{-1} = I$ والناتج هو المصفوفة المحايدة. يعرف معكوس المصفوفة فقط للمصفوفات المربعة ، بالمقابل ليس كل مصفوفة مربعة لها معكوس. عندما A^{-1} تكون موجودة فإن:

$$A^{-1} \times A = A \times A^{-1} = I$$

مثال : لحساب معكوس المصفوفة A حيث أن :

$$A = \begin{pmatrix} 4 & 1 \\ 3 & 5 \end{pmatrix}$$

يتبع الآتي:

$$\Delta = a_{11} a_{22} - a_{12} a_{21} = 17$$

$$\begin{pmatrix} 4 & 1 \\ 3 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5/17 & -1/17 \\ -3/17 & 4/17 \end{pmatrix}$$

ولغرض التثبت من الحل نتبع الآتي:

$$\begin{pmatrix} 4 & 1 \\ 3 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 5/17 & -1/17 \\ -3/17 & 4/17 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ويمكن أن تكتب بالرموز وكالاتي:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \begin{pmatrix} a_{22}/\Delta & -a_{12}/\Delta \\ -a_{21}/\Delta & a_{11}/\Delta \end{pmatrix}$$

يطلق على الكمية العددية Δ محدد المصفوفة. من الواضح إن معكوس المصفوفة غير معرف إذا كان $\Delta = 0$. في هذه الحالة تسمى بالشاذة Singular. بالنسبة للمصفوفات ذات حجم 3×3 أو أكبر فإن عملية استخراج معكوس المصفوفة يكون أكثر تعقيدا مما تقدم في أعلاه ويستخدم برنامج حاسوب لاحتسابه.

الأشكال التربيعية Quadratic forms

يفترض أن $A_{n \times n}$ مصفوفة مربعة وأن $\underline{x}$ متجه عمودي بطول n فإن التعبير :

$$Q = \underline{x}' A \underline{x}$$

يسمى الشكل التربيعي . وهو عبارة عن عدد ويمكن أن يوضح بالمعادلة البديلة الآتية:

$$Q = \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j$$

حيث أن x_i هو عنصر في الصف i في المتجه $\underline{x}$ و a_{ij} هو عنصر في الصف i والعمود j من المصفوفة A .

القيم الذاتية (المميزة) والمتجهات الذاتية (المميزة) Eigen values and Eigen vectors
نتأمل مجموعة المعادلات الآتية:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \lambda x_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \lambda x_2$$

.

.

.

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = \lambda x_n$$

والتي يمكن كتابتها بصيغة المصفوفات وكالاتي :

$$A \underline{x} = \lambda \underline{x}$$

$$(A - \lambda I) \underline{x} = 0 \quad \text{أو}$$

حيث أن I هي المصفوفة المحايدة ذات حجم $n \times n$ و 0 هو متجه من الأصفار بحجم $n \times 1$. يمكن ملاحظة أن هذه المعادلات يمكن أن تثبت لقيم خاصة من λ ، والتي تسمى بالجذور الذاتية (المميزة) للمصفوفة A . هنالك n من هذه القيم (الجذور) المميزة .
مثال : المصفوفة A الآتية :

$$A = \begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix}$$

لغرض استخراج القيم الذاتية فإن :

$$A\underline{X} = \lambda\underline{X}$$

$$\begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \lambda \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\begin{aligned} A \underline{x}_i - \lambda \underline{x}_i &= 0 \\ (A - \lambda I) \underline{x}_i &= 0 \end{aligned}$$

وبصيغة المصفوفات :

$$\left(\begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

وحسب قاعدة كرامر Cramer's rule والتي تعطي تعبير لحل نظام المعادلات الخطية هذا الحل يكون وحيدا. ويمكن ان يكون حل لهذه المعادلة في حالة واحدة اذا كان المحدد للمصفوفة A يساوي صفر وبعبارة أخرى

$$\begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0$$

ولحل مثل هذا المحدد يتبع الآتي:

$$\begin{aligned} (1-\lambda)(1-\lambda) - (4)(9) &= 0 \\ (1-\lambda)^2 - 36 &= 0 \end{aligned}$$

وبأخذ الجذر التربيعي للطرفين

$$1-\lambda \pm 6 = 0$$

أما

$$\begin{aligned} 1-\lambda_1 + 6 &= 0 \\ -\lambda_1 + 7 &= 0 \end{aligned}$$

القيمة الأولى هي:

$$\lambda_1 = 7$$

أو

$$\begin{aligned} 1 - \lambda_2 - 6 &= 0 \\ -\lambda_2 - 5 &= 0 \end{aligned}$$

القيمة الثانية

$$\lambda_2 = -5$$

وبهذا فان جذور هذه المعادلة هي:

$$\lambda_1 = 7$$

$$\lambda_2 = -5$$

إن هذه القيم تسمى القيم الذاتية أو المميزات للمصفوفة A . ولكل قيمة خاصة λ_i يقابلها متجه خاص x_i للمصفوفة المربعة A .

متجهات الأوساط ومصفوفات التباين

لغرض دراسة أي ظاهرة ما (تجربة معينة)، يتم سحب عينة من مجتمع الدراسة. لغرض تلخيص العينة المسحوبة من مجتمع الدراسة، فانه غالبا ما يعبر عنها بالمعدل والتباين. نفرض ان عينة الدراسة هي عينة عشوائية بحجم n ويتولد عنها قيم متغير العينة $X_1, X_2, \dots, X_n$ فان متوسط العينة هو :

$$\bar{X} = \sum_{i=1}^n x_i / n \quad (1.4)$$

في حين تقدير تباين العينة هو:

$$S_j^2 = \sum_{i=1}^n (x_{ij} - \bar{X})^2 / (n-1) \quad (2.4)$$

مما تقدم أعلاه تطلق تسمية احصاءة على كل من متوسط وتباين العينة، وبالمقابل، فان هذه الإحصاءات هي تقديرات إلى معلمات المجتمع وهي كل من متوسط المجتمع (μ) وتباين المجتمع (σ^2).

وبطريقة مماثلة، فان المجتمع الذي يمتلك أكثر من متغير يسمى مجتمع متعدد المتغيرات Multivariate population والعينات المسحوبة من هذه المجتمعات يمكن تلخيصها والتعبير عنها (أو تمثيلها) بواسطة متجهات المتوسط ومصفوفات التباين، والتي يمكن تعريفها كالآتي: نفرض انه لدينا p من المتغيرات $X_1, X_2, \dots, X_p$ وقيم المشاهدة في العينة هي $X_{1p}, \dots, X_{ip}, \dots, X_{jp}, \dots, X_{np}$ على التوالي. فان متوسط العينة للمتغير j هو

في حين تباين العينة هو:

حيث أن $j = 1, 2, 3, \dots, m$

بالإضافة إلى ذلك تعرف التباين للعينة بين المتغيرين j و k كالآتي:

$$C_{jk} = \sum_{i=1}^n (x_{ij} - \bar{X}_j) (x_{ik} - \bar{X}_k) / (n-1) \quad (3.4)$$

حيث أن :

$k = 1, 2, 3, \dots, m$

والتي تقيس علاقة المتغيرين. فان معامل الارتباط بين المتغيرين j و k هو r_{jk} الذي يعبر عنه بالمعادلة الآتية :

$$r_{jk} = C_{jk} / (S_j S_k) \quad (4.4)$$

S_j : الانحراف المعياري للمتغير j

S_k : الانحراف المعياري للمتغير k

ومن الواضح من المعادلتين (3.4) و (4.4) ان:

$$c_{jk} = c_{kj}, r_{jk} = r_{kj}, r_{kk} = 1$$

يحسب متجه متوسط العينة من المعادلة (1.4):

$$\bar{X} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{pmatrix} \quad (5.4)$$

يمكن القول أن $\bar{X}$ هو مركز العينة. هو تقدير لمتجهات متوسط المجتمع

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad (6.4)$$

وان مصفوفة التباين والتغاير هي:

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{pmatrix} \quad (7.4)$$

حيث ان C ، والتي تسمى مصفوفة تغاير العينة تعكس كمية الاختلاف في العينة والتي تحتوي p من المتغيرات المرتبطة. الذي هو تقدير مصفوفة تغاير المجتمع . تعرف مصفوفة الارتباط بالمعادلة (4.4) وكالاتي:

$$R = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

وتسمى مصفوفة ارتباط العينة. وان كلا المصفوفتين (R و C) يجب ان تكون متماثلتين.

مفاهيم إحصائية Statistical concepts

تلعب المفاهيم الإحصائية دوراً كبيراً في المعلوماتية الحيوية، يوضح الجدول (2.4) التطبيقات النموذجية للإحصاء في علم المعلوماتية الحيوية والتي تمتد من التشخيص السريري إلى تماثل السلاسل وفي الكثير من الحالات تستخدم الأدوات الإحصائية لغرض تخمين خصائص أو معالم المجتمع من خلال سحب عينة عشوائية أو غير عشوائية حسب طبيعة البحث والهدف منه لغرض إجراء الدراسة المطلوبة ، هذه المعلومات (الخاصة بمجتمع الدراسة) غير معروفة ، وبسبب تخمين هذه المعلمة (المعلومات) يحصل خطأ (هو عبارة عن الفرق بين قيمة المعلمة الحقيقية وقيمة التخمين لهذه المعلمة) حيث أن أكثر الإحصائيين يعملون على تقليص حجم هذا الخطأ قدر الإمكان .

جدول 2.4: التطبيقات الإحصائية في مجال المعلوماتية الحيوية

Clinical diagnosis	التشخيص السريري
Descriptive statistics	الإحصاء الوصفي
Equipment calibration	تقويم المعدات
Experimental data analysis	تحليل البيانات التجريبية
Gene expression prediction	التنبؤ بالتعبير الجيني
Gene prediction	التنبؤ بالجين
Genetic linkage analysis	تحليل ارتباط الجين
Nucleotide alignment	اصطفاف النيوكليوتيدات
Population studies	دراسات المجتمع
Protein function prediction	التنبؤ بوظائف البروتين
Protein structure prediction	التنبؤ بتركيب البروتين
Quantifying uncertainty	تقييم الشك
Quality control	السيطرة النوعية
Sequence similarity determination	تحديد تشابه التواليات

فضلاً عن أغراض أخرى لذلك كان من الضروري التطرق إلى القواعد الأساسية للعشوائية و الاحتمالية التي لها علاقة بعلم المعلوماتية الحيوية. على سبيل المثال ، يعد الجنس حالة عشوائية ، وان احتمالية ان يكون كائن ما ذكر مساوي نظرياً إلى احتمالية ان يكون انثى بافتراض عدم وجود قوة خارجية للتأثير في التوازن الطبيعي لحالة الجنس. يوضح هذا المثال الحالة العشوائية لتوزيع الكائن حسب حالة الجنس إلى ذكر وأنثى. تعد حالات التغيرات ، فرصة التقاء الأزواج أو ضغوط البيئة العشوائي والتفسيرات النسبية لأنواع الجينية بين الآباء وأولادهم حالات يمكن تفسيرها إحصائياً.

رموز المجموعة The notion of set

تعرف المجموعة بأنها عبارة عن مجموعة من الأشياء أو العناصر التي تشترك بصفة أو أكثر تميزها عن باقي الأشياء الموجودة في المجموعات الأخرى. يرمز إلى اتحاد المجموعة A مع

المجموعة B بالرمز $A \cup B$ وإلى تقاطع المجموعتين بالرمز $A \cap B$ ويرمز إلى مكمل (متمم) المجموعة A بالرمز A^c .
على سبيل المثال
مثال (1.4) إذا كانت:

$$A = \{ a, b \}$$

$$B = \{ b, c \}$$

$$A \cup B = \{ a, b, c \}$$

$$A \cap B = \{ b \}$$

$$A^c = \{ c, d, \dots, z \}$$

أو بعبارة أخرى

$$A^c = 1 - A$$

مثال (2.4) : إذا كان A و B تواليين وإن التوالي A مكون من (5) خمسة أحماض امينية والتوالي B مكون من (6) أحماض امينية ، وكالاتي:

$$A = \{ A, C, F, G, P \}$$

$$B = \{ G, S, F, A, L, I \}$$

$$A \cup B = \{ A, C, F, G, I, L, P, S \}$$

$$A \cap B = \{ A, F, G \}$$

إن كل العناصر في المجموعة الواحدة هي من النوع نفسه (لها صفات متشابهة) ، فعلى سبيل المثال إذا كانت مجموعة معينة تسمى بالتفاح فأنها تحتوي على أنواع مختلفة من فاكهة التفاح، وأنها تختلف في الألوان والإحجام لكن هذه المجموعة (التفاح) لا تحتوي على فاكهة أخرى مثل البرتقال أو الرمان أو الخ.... كذلك، على سبيل المثال فإن مجموعة A تحتوي على مجموعة مختلفة من الأحماض الامينية $A = \{ L, F, A, C, P, H, G, S \}$ فإن هذه المجموعة تحتوي على أحماض امينية مختلفة في الحجم والخصائص الكيميائية ولكنها ترمز إلى أحماض امينية لبروتين معين يختلف عن أي مجموعة أخرى مثل B (التي تمثل القواعد النتروجينية) . حيث إن $B = \{ T, C, G, A, A, T, C \}$ التي تمثل القواعد النتروجينية العائدة إلى DNA والتي تختلف بأية حال من الأحوال عن المجموعة الأولى A .

أساليب سحب العينات

إن أكثر مشاكل العد تتمثل في صيغة تحديد الاختيارات التي تتمثل بأربعة أنواع أساسية ، المناظرة إلى إحلال أو عدم الإحلال ومرتبة أو غير مرتبة. وهو مرادف إلى مجموعة مختلفة من الكرات موجودة في صناديق مختلفة . وكذلك تماثل أنواع مختلفة من التواليات الخاصة بأنواع مختلفة من الكائنات الحية.

عينة مرتبة مع الإحلال Ordered sampling with replacement

إذا تم سحب عينات متعاقبة بحجم متساوي n وبعدد r من المرات مع الإحلال بعبارة أخرى عند سحب العنصر أو الشيء يتم إعادته بعد الانتهاء من إجراء العملية على هذا العنصر ويمكن إعادة سحبه مرة أخرى في حالة سحب عينة ثانية وهكذا فإن المجموع الكلي:

$$\prod_{i=1}^r n_i = \underbrace{n \cdot n \cdot n \cdots n}_r = n^r$$

بدلالة الكرات ، نفرض ان لدينا صف من الخلايا (الصناديق) قي كل خلية كرة وتم سحب 5 عينات متتالية كل عينة تملك 10 كرات وبعد سحب كل كرة يتم تعويضها بكرة مماثلة ومن النوع والحجم نفسه وبذلك يمكن حساب العدد الكلي كالآتي:

$$\prod_{i=1}^5 10 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 10^5 = 100000$$

إذن هناك 100000 طريقة يمكن بها سحب 5 عينات متتالية بحجم 10 كرات بوجود الإحلال . مثال آخر في هذا المجال سحب 7 عينات متتالية بحجم 5 تواليات مختلفة بوجود الإحلال فإن العدد الكلي هو:

$$\prod_{i=1}^7 5 = 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 = 5^7 = 78125$$

إذن هناك 78125 طريقة يمكن بها سحب 7 عينات متتالية بحجم 5 سلاسل مختلفة بوجود الإحلال .

عينات مرتبة بدون إحلال Ordered sampling without replacement

في هذا النوع من المعاينة يتم سحب العنصر لكن لا يتم إعادته أو تعويضه . وفي هذه الحالة عند عدم تعويض الكرة المسحوبة فإنه سيكون لدينا $(n-1)$ من الخيارات لملي الخلية الثانية (الصندوق الثاني) و $(n-2)$ من الخيارات لملي الخلية الثالثة..... وهكذا. فإذا كان عدد الخلايا مساوي الى عدد الكرات $(r = n)$ فإن هنالك $n!$ ($n! = n(n-1)(n-2) \cdots 1$) من التباديل المختلفة. وبصورة أكثر عمومية إذا كان $r \leq n$ فإن عدد الترتيب هو:

$$P_r^n = \frac{n!}{(n-r)!} \quad (8.4)$$

حيث p يرمز الى التباديل. علما ان $0!$ تعرف على انها مساوية للواحد ، مثال على ذلك اذا سحبت أربع عينات بحجم أربع كرات بدون إحلال فإن عدد التباديل الممكنة هو:

$$4! = 4.3.2.1 = 24$$

أو

$$P_4^4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = 24$$

مثال آخر ، سحبت 5 عينات بحجم 7 تواليات مختلفة بدون إحلال فان عدد التباديل الممكنة هو : ($r=5 < n=7$)

$$P_r^n = P_5^7 = \frac{7!}{(7-5)!} = \frac{7!}{2!} = \frac{5040}{2} = 2520$$

إذن العدد 2520 يمثل عدد التباديل الممكنة لسحب 5 عينات بحجم 7 تواليات مختلفة بدون إحلال (تعويض).

الاختيار العشوائي Random choice

يعني ان جميع الاختيارات متساوية الاحتمال في العينات العشوائية. بعبارة أخرى للعينات العشوائية ذات الحجم الثابتة ، فان جميع العينات تكون متساوية في احتمالية الظهور (الاختيار) ومقدار هذا الاحتمال هو $1/n^r$ في حالة وجود الإحلال و $1/p_r^n$ في حالة عدم الإحلال .

Unordered sampling without replacement العينات غير المرتبة بدون إحلال
نفترض الآن إننا نعيد تكرار العمل في العمليات المذكورة في الحالات السابقة ، لكن بدون الأخذ بنظر الاعتبار الترتيب، بعبارة أخرى اختيار r من العناصر من مجموع n . ولحساب عدد التوافيق الممكنة لاختيار (r من العناصر من مجموع n) هي :

$$C_r^n = \binom{n}{r} = \frac{n!}{r! (n-r)!}$$

يرمز C الى التوافيق ولاحظ ان:

$$\binom{n}{r} = \binom{n}{n-r} \quad ; \quad r = 0, 1, 2, \dots, n ;$$

$$\frac{n!}{r! (n-r)!} = \frac{n!}{(n-r)! (n-n+r)!}$$

وبذلك

$$\frac{n!}{r! (n-r)!} = \frac{n!}{(n-r)! r!}$$

بعبارة أخرى فان اختيار 5 من مجموع 9 مساوي الى اختيار 4 من 9 وكما يأتي:

$$\binom{9}{5} = \frac{9!}{5! 4!} = 126$$

$$\binom{9}{4} = \frac{9!}{4! 5!} = 126$$

العينات غير المرتبة مع الإحلال Unordered sampling with replacement

يعد أحد الأنواع الأربعة الأساسية ، ويمكن وصف الحادثة من خلال عدد الخلايا المشغولة. على سبيل المثال رمي زار $r=8$ من المرات (حيث ان الزار له 6 أوجه) وكانت النتائج كما يأتي: 3, 1, 0, 0, 0, 4 ، وهذا يشير الى ظهور الرقم 1 ثلاث مرات والرقم 2 مرة واحدة والرقم 6 اربع مرات ، بعبارة أخرى يمكن كتابتها كالآتي: 1, 1, 1, 2, 6, 6, 6. 6.

وبصورة عامة فان عدد التكرارات الكلي هو:

$$r_1 + r_2 + \dots + r_n = r$$

وفي المثال أعلاه فان:

$$r_1 = 3, r_2 = 1, r_6 = 4, r = 8$$

الاحتمالات المستقلة والمشروطة Independences and conditional probabilities

قبل الدخول الى تفاصيل الاحتمالات المستقلة والمشروطة، لابد من معرفة ما هو المقصود بكلمة الاحتمال او الاحتمالية التي كثيرا ما تستخدم حتى في الحياة اليومية حيث ترتبط الاحتمالية بوجود او حدوث ظاهرة معينة التي يرغب الباحث في دراستها. على سبيل المثال اذا كان هنالك أربعة ألوان (احمر - اخضر - اصفر - ابيض) وأننا مهتمون بظهور اللون الأحمر (صفة معينة أو ظاهرة معينة) فان احتمال ظهور اللون الأحمر هو:

$$P(r) = \frac{a}{s} = \frac{1}{4}$$

حيث ان r رمز اللون الأحمر

a قيمة النجاح او ظهور الصفة المطلوبة

s عدد العناصر الموجودة في عينة البحث

في حين اذا كانت عينة البحث مكونة من الألوان ابيض ، ازرق ، اخضر، احمر، اسود، اصفر فان احتمالية الحصول على اللون الأحمر هو:

$$P(r) = \frac{a}{s} = \frac{1}{6}$$

وهنا تدرج أهم الخصائص

$$(i) p(A) \in [0, 1]$$

$$(ii) p(\Omega) = 1$$

$$(iii) P\left(\sum_{i=1}^{\infty} A_{ik}\right) = \sum_{i=1}^{\infty} P(A_i)$$

حيث ان Ω عناصر عينة البحث
على سبيل المثال:

$$P(A \cup B) = P(A) + P(B), A \cap B = \emptyset$$

$$P(\Omega) = P(A) + P(A^c) = 1$$

$$P(A^c) = 1 - p(A)$$

وبصورة عامة

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

وفي حالة كون A مستقلة عن B فان

$$P(A \cup B) = P(A) + P(B)$$

مثال (3.4): تتكون عناصر DNA من أربعة قواعد هي: A , C , G , T. المطلوب معرفة مقدار احتمالية كل عنصر من العناصر الرئيسية في التوالي الآتي:

A , C , G , G , A , A , C , T , T , A , C , C , C , T , T , A , A , A , C , C
الحل:

طول التوالي هو 20

إذن $n = 20$

يرمز الى احتمالية ظهور الحدث (ظهور العناصر T و G و C و A) كما يأتي:

$$P(T) = \frac{\text{No. of T}}{\text{Total no. of seq. (n)}} = \frac{4}{20} = 0.2$$

$$P(G) = \frac{\text{No. of G}}{\text{Total no. of seq. (n)}} = \frac{2}{20} = 0.1$$

$$P(C) = \frac{\text{No. of C}}{\text{Total no. of seq. (n)}} = \frac{7}{20} = 0.35$$

$$P(A) = \frac{\text{No. of A}}{\text{Total no. of seq. (n)}} = \frac{7}{20} = 0.35$$

مما تقدم يلاحظ ان جميع الاحتمالات أعلاه (0.2, 0.1, 0.35) هي ضمن الفترة المغلقة [0,1] وبذلك لا يمكن الحصول على قيم اكبر من الواحد الصحيح او اقل من الصفر ، وهذا يفسر الخاصية الأولى (i) اما الخاصية الثانية (ii) فهي:

$$P(\Omega) = P(T,G,C,A) = 0.2 + 0.1 + 0.35 + 0.35 = 1$$

يمكن توضيح الخاصية الثالثة (iii) كالآتي:
 باستخدام الطرف الأيسر يصبح لدينا

$$\begin{aligned} P\left(\sum_{i=1}^4 A_i\right) &= P(A,C,G,T) = \frac{7+7+2+4}{20} \\ &= \frac{20}{20} = 1 \quad \text{————— I} \end{aligned}$$

باستخدام الطرف الأيمن يصبح لدينا

$$\begin{aligned} \sum_{i=1}^4 P(A_i) &= P(A_1) + P(A_2) + P(A_3) + P(A_4) \\ &= P(A) + P(C) + P(G) + P(T) \\ &= 0.35 + 0.35 + 0.1 + 0.2 \\ &= 1 \quad \text{————— II} \end{aligned}$$

يلاحظ ان الطرفين (I) و (II) متساويان .
 كذلك يمكن توضيح الخاصية الآتية:

$$P(A \cup T) = P(A) + P(T) - P(A \cap T)$$

$$= 0.35 + 0.2 - 0$$

$$= 0.55$$

$$P(A^c) = 1 - P(A)$$

$$= 1 - 0.35$$

$$= 0.65$$

أو

$$P(A^c) = P(C) + P(G) + P(T)$$

$$= 0.35 + 0.1 + 0.2$$

$$= 0.65$$

من الأفكار المهمة في الاحتمالية والإحصاء هو موضوع استقلالية الاحتمالات، بعبارة أخرى استقلالية حدوث الحالة A عن حدوث الحالة B . رياضيا يمكن التعبير عن استقلالية الحالتين A و B اذا وفقط اذا

$$P(A \cap B) = P(A)P(B)$$

في حين الاحتمالية المشروطة، بعبارة أخرى هو احتمالية حدوث الحالة A بوجود (على شرط وجود) الحالة B تعرف كما يأتي:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

في حالة الحالتين A و B مستقلتين فان:

$$P(A/B) = P(A)$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(A) P(B)}{P(B)}$$

$$= P(A)$$

كذلك فان:

$$P(A/B) P(B) = P(B/A) P(A)$$

للتثبت من صحة الطرفين نأخذ الطرف الأيسر وكالاتي:

$$\begin{aligned}
 P(A/B) \cdot P(B) &= \frac{P(A \cap B)}{P(B)} \cdot P(B) \\
 &= \frac{P(A) P(B)}{P(B)} P(B) \\
 &= P(A) P(B)
 \end{aligned}$$

ثم نأخذ الطرف الأيمن وكالاتي:

$$\begin{aligned}
 P(B/A) P(A) &= \frac{P(B \cap A)}{P(A)} \cdot P(A) \\
 &= \frac{P(B) P(A)}{P(A)} P(A) \\
 &= P(B) P(A) \\
 &= P(A) P(B)
 \end{aligned}$$

وبهذا فإن الطرف الأيمن يساوي الطرف الأيسر.

إشارة الى المثال (3.4) الخاص بعناصر DNA يمكن توضيح القانون الخاص بالاحتمالية المشروطة وكالاتي:

$$\begin{aligned}
 P(C/G) &= \frac{P(C \cap G)}{P(G)} \\
 &= \frac{P(C) P(G)}{P(G)} \\
 &= \frac{(0.35) (0.1)}{(0.1)} \\
 &= 0.35 = P(C)
 \end{aligned}$$

يلاحظ ان قيمة الاحتمال المشروط $P(C/G)$ كان مساويا الى قيمة $P(C)$ ومقداره 0.35 .

مثال(4.4): يفرض ان الجين A له صورتين (alleles) هي A1 و A2 وان B هو جين آخر له صورتين هما B1 و B2 فان:

$$P(B) = P(B1) + P(B2) = 1$$

وان

$$P(A) = P(A1) + P(A2) = 1$$

نفرض ان

$$P(A1/B1) = 0.8 , P(A2/B2) = 0.7$$

وطالما ان

$$P(A1/B1) + P(A2/B1) = 1$$

فان

$$\begin{aligned} P(A2/B1) &= 1 - P(A1/B1) \\ &= 1 - 0.8 \\ &= 0.2 \end{aligned}$$

وطالما ان

$$P(A1/B2) + P(A2/B2) = 1$$

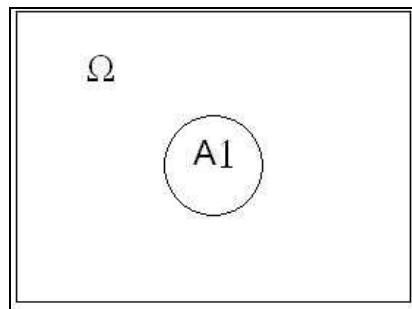
فان

$$\begin{aligned} P(A1/B2) &= 1 - 0.7 \\ &= 0.3 \end{aligned}$$

إذا كان لدينا حوادث (حالات) $A1$ و $A2$ هي أحداث متنافرة (مستقلة)
 $(A_i \cap A_j = \emptyset)$ ، حيث أن $i \neq j$ وفي حالة جمع هذه الحالات تعطينا المجموعة الشاملة
 $(\Omega = \bigcup_{k=1}^{\infty} A_k)$ ، فان

$$P(A1) = P(A1 \cap \Omega)$$

ويمكن التعبير عنها بالشكل (10.4)



شكل 10.4 : المجموعة الشاملة والمجموعة الجزئية

مثال (5.4): بافترض ان المجاميع الموجودة في الشكل 10.4:

$$\Omega = \{ 1, 2, 3, 4, 5, 6 \}$$

$$A1 = \{ 1, 2, 4 \}$$

فان:

$$P(A1) = \frac{3}{6} = \frac{1}{2}$$

$$P(A1 \cap \Omega) = \frac{(A1 \cap \Omega)}{\Omega} = \frac{3}{6} = \frac{1}{2}$$

وبذلك فان:

$$P(A1) = P(A1 \cap \Omega)$$

مثال (6.4): من المثال (4.4)، المطلوب إيجاد الاحتمالات المشتركة لكل من A و B او بعبارة أخرى إيجاد قيم الجدول (3.4) الآتي:

جدول 3.4: إيجاد القيم الاحتمالية لكل من A و B

	P(A1)	P(A2)	المجموع
P(B1)			
P(B2)			
المجموع			

في البدء نحدد الاحتمالات الآتية:

$$P(B1) = 0.3$$

$$P(B2) = 1 - P(B1) = 1 - 0.3 = 0.7$$

$$P(A1) = 0.45$$

$$P(A2) = 1 - P(A1) = 1 - 0.45 = 0.55$$

لاستخراج الاحتمالات المشتركة يتبع الآتي:

$$P(A1 \text{ and } B1) = P(B1) P(A1/B1) = 0.3 \times 0.8 = 0.24$$

$$P(A2 \text{ and } B2) = P(B2) P(A2/B2) = 0.7 \times 0.7 = 0.49$$

$$P(A1 \text{ and } B2) = P(B2) P(A1/B2) = 0.7 \times 0.3 = 0.21$$

$$P(A2 \text{ and } B1) = P(B1) P(A2/B1) = 0.3 \times 0.2 = 0.06$$

بذلك يصبح الجدول (4.4) بالشكل الآتي:

جدول 4.4: القيم الاحتمالية لكل من A و B

	P(A1)	P(A2)	المجموع
P(B1)	0.24	0.06	0.3
P(B2)	0.21	0.49	0.7
المجموع	0.45	0.55	1.0

توزيع ذي الحدين Binomial distribution

عبارة عن تجربة يمكن تكرارها بعدد معين من المحاولات المستقلة وليكن n حيث أن ناتج هذه التجربة تتمثل بنتيجتين فقط ، هما نجاح المحولة (p) أو فشلها (q). فإن احتمال الحصول على x من المحاولات الناجحة ($x < n$) هو :

$$b(k; n; p) = \binom{n}{k} p^k q^{n-k}$$

تعرف هذه الدالة باسم توزيع ذي الحدين، ويمكن إعادة صياغة التعريف أعلاه وكالاتي:
يتوزع المتغير العشوائي K وفق توزيع ذي الحدين إذا كانت دالة الكتلة الاحتمالية له تأخذ الشكل الآتي:

$$P(K) = \begin{cases} b(k; n; p) = \binom{n}{k} p^k q^{n-k} & K=0,1,2,3,\dots,n \\ & 0 < p < 1; q = 1 - p \\ 0 & \text{o. w.} \end{cases}$$

مثال (7.4) : بافتراض توالي بطول 100 حامض اميني واننا مهتمون بدراسة نوع معين من الأحماض الامينية في هذه التوالي وليكن الحامض الاميني الالنين Ala. لمعرفة عدد مرات ظهور الحامض الاميني الالنين Ala في التوالي هو

$$b(k; n; p) = \binom{100}{k} \left(\frac{1}{20}\right)^k \left(\frac{19}{20}\right)^{100-k}, k = 0,1,2,\dots,100$$

علما انه تم تحديد احتمالية ظهور الحامض الاميني الالنين في حالة تساوي الاحتمالات بمعنى ان كل حامض اميني له فرصة الظهور نفسها
تعرف المحاولات المستقلة بالصيغة الآتية:

$$P(AH) = P(A) P(H)$$

والتي تتحقق في حالة كون الحالتين A و H مستقلين. في حين تمثل K عدد مرات ظهور الحامض الاميني Ala في التتابعين ذات الطول 100. فعندما تكون قيمة K مساوية الى الصفر، دل ذلك على عدم ظهور الحامض الاميني Ala في هذه السلسلة، في حين اذا كانت قيمة $K=1$ دل ذلك على وجود حامض اميني A واحد في هذه السلسلة وإذا كانت قيمة $K=2$ دل ذلك على وجود حامض اميني عدد 2 في هذه السلسلة وهكذا.
فاذا أراد الباحث تحديد ظهور حامض اميني من النوع Ala بعدد 5 في السلسلة فان ذلك يكون مساوي الى :

$$b(k; n, p) = b(5; 100, \frac{1}{20}) = \binom{100}{5} \left(\frac{1}{20}\right)^5 \left(\frac{19}{20}\right)^{95}$$

$$\binom{100}{5} = \frac{100!}{5! 95!} = \frac{100.99.98.97.96.95!}{5! 95!}$$

$$= 75287520$$

$$b(5; 100, \frac{1}{20}) = 75287520 \times 0.0000003125 \times 0.00765$$

$$= 0.18$$

وبهذا فان

مثال (8.4): استنادا الى المعطيات في المثال السابق اذا كان المطلوب هو معرفة احتمالية ظهور الحامض الاميني Ala لمرة واحدة فيكون الآتي:

$$b(k; n, p_A) = b(1; 100, \frac{1}{20}) = \binom{100}{1} \left(\frac{1}{20}\right) \left(\frac{19}{20}\right)^{99}$$

$$\binom{100}{1} = \frac{100!}{1! 99!} = 100$$

حيث أن :

$$b(1; 100, \frac{1}{20}) = 100 \times 0.05 \times 0.00623 = 0.0312$$

وبهذا فان

وإذا أريد معرفة احتمالية ظهور الحامض الاميني C لمرة واحدة فيكون الآتي:

$$B(k; n, p_C) = b(1, 100, 1/20) = 0.0312$$

وهكذا بالنسبة لبقية الأحماض الامينية العشرين.

يمكن الإشارة الى تطبيق مهم، فعندما حجم n كبير وقيمة p صغيرة فان $np = \lambda$ ويمكن تبسيط التوزيع بالشكل الآتي:

$$b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda} = P(K; \lambda)$$

الذي يطلق عليه توزيع بواسون Poisson . على أية حال اذا كانت λ ثابتة، فان $\sum p(K; \lambda) = 1$ وبالتالي فان $p(K; \lambda)$ هي احتمالية حدوث K من النجاحات.

مثال (9.4): اذا كان توالي مكون من 150 حامض اميني، وان احتمالية ظهور الحامض الاميني C هو $1/200$ فان احتمالية الحصول على K من المرات من هذا الحامض الاميني هو:

$$P(K; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad K = 0, 1, 2, \dots$$

وان

$$\lambda = 150 \times \frac{1}{200} = 0.75$$

$$p(k; 0.75) = \frac{(0.75)^k e^{-0.75}}{k!}$$

مثال (10.4): استنادا الى المثال السابق، اذا كان المطلوب الحصول على ظهور الحامض الاميني C مرة واحدة، فان الاحتمالية تحسب كما يأتي:

$$P(K; \lambda) = P(1; 0.75) = \frac{(0.75) e^{-0.75}}{1!} = 0.354$$

باستخدام توزيع الثنائي Binomial

$$P(k; n, p) = P(1; 150, \frac{1}{200}) = \binom{150}{1} \cdot \left(\frac{1}{200}\right) \cdot \left(\frac{199}{200}\right)^{149}$$

$$= 150 \cdot 0.005 \cdot 0.474 = 0.355$$

يلاحظ تساوي الناتجين أعلاه.

ان احتمالية $f(k; r, p)$ هي حدوث حالات فشل عددها k قبل الحصول على r نجاح تحسب بالصيغة الآتية:

$$f(k; r, p) = \binom{r+k-1}{k} p^r q^k, \quad k = 0, 1, 2, \dots$$

اذا فقط اذا

$$\sum_{k=0}^{\infty} f(k; r, p) = 1$$

مثال (11.4): اذا كان توالي بطول 100 حامض اميني والمطلوب تحديد احتمالية الحصول على الحامض الاميني A مرة واحدة بعد 6 مواقع (6 حالات فشل). علما ان احتمالية ظهور هذا الحامض الاميني هو $1/20$.

$$r = 1, k = 6$$

$$f(k;r,p) = f(6;1, \frac{1}{20}) = \binom{1+6-1}{6} (\frac{1}{20}) (\frac{19}{20})^6$$

$$= 1 \times 0.05 \times 0.735 = 0.037$$

مثال (12.4): استنادا الى المثال السابق، اذا كان المطلوب هو الحصول على احتمالية ظهور الحامض الاميني A بخمس مرات بعد ست مواقع ، يمكن ان تحسب بالصيغة الآتية:

$$f(k;r,p) = f(6;5, \frac{1}{20}) = \binom{5+6-1}{6} (\frac{1}{20})^5 (\frac{19}{20})^6$$

$$= \binom{10}{6} = \frac{10!}{6!4!} = \frac{10 \times 9 \times 8 \times 7 \times 6!}{6! \times 4 \times 3 \times 2 \times 1}$$

$$f(6;5, \frac{1}{20}) = 210 \times 0.0000003125 \times 0.7351$$

$$= 0.0000482$$

التعاقبات Runs

ان الدراسات الإحصائية المتعلقة بخصائص DNA التي غالبا ما تبدأ بتحديد مجموع أعداد القواعد الأربعة A و C و T و G . تهتم نظرية التوزيع للتعاقبات بالمعلومات الخاصة بترتيب حدوث هذه القواعد. يعرف التعاقب بأنه سلسلة متشابهة من الأحداث ويشير عدد العناصر في التعاقب الى طول هذا التعاقب . ان عدد التعاقبات بالطبع مساوي الى عدد الحروف (الرموز) المتجاورة غير المتشابهة.

مثال (13.4): اذا كان التوالي الآتي:

AAAAAGGGAAAAATTTTTT

يلاحظ الآتي:

يوجد أربعة تعاقبات هي:

- 1- تعاقبين للقاعدة A بطول 5 قواعد.
- 2- تعاقب واحد للقاعدة G بطول 3 قواعد.
- 3- تعاقب واحد للقاعدة T بطول 7 قواعد.

التوزيع الهندسي Hypergeometric distribution

بافتراض نوعين من العناصر a و b ونفترض ان n_1 هو عدد العناصر من النوع a وان n_2 هو عدد العناصر من النوع b حيث ان $n = n_1 + n_2$. اختيار (سحب) عينة من العناصر بعدد r

بصوره عشوائيا من خليط من العنصرين a و b . ما هو احتمال q_k (احتمالية العنصر المطلوب) بأن المجموعة ستحتوي عدد k من العنصر a ؟ فانه من الضروري ستحتوي على عدد $r-k$ من العنصر b ، ويمكن احتساب الاحتمالية كما يأتي:

$$q_k = \frac{\binom{n_1}{k} \binom{n-n_1}{r-k}}{\binom{n}{r}}$$

وتسمى هذه الصيغة بالتوزيع الهندسي

مثال (14.4): توالي من القواعد النتروجينية بطول 100 قاعدة سحبت عينة من القواعد النتروجينية من هذا التوالي بحجم 20 ووجد 12 منها من القاعدة A ، ما هو احتمال القاعدة A ؟

$$q_k = \frac{\binom{20}{12} \binom{100-20}{20-12}}{\binom{100}{20}}$$

$$\binom{20}{12} = \frac{20 \times 19 \times 18 \times 17 \times 16 \times 15 \times 14 \times 13 \times 12!}{12! \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 125970$$

$$\binom{100-20}{20-12} = \binom{80}{8} = \frac{80!}{8! \cdot 72!} = \frac{80 \times 79 \times 78 \times 77 \times 76 \times 75 \times 74 \times 73 \times 72!}{8! \times 72!}$$

$$= 28987537000$$

$$\binom{100}{20} = \frac{100!}{20! 80!} = \frac{100 \times 99 \times 98 \times \dots \times 81}{20! \times 80!}$$

$$= 5.3598337E$$

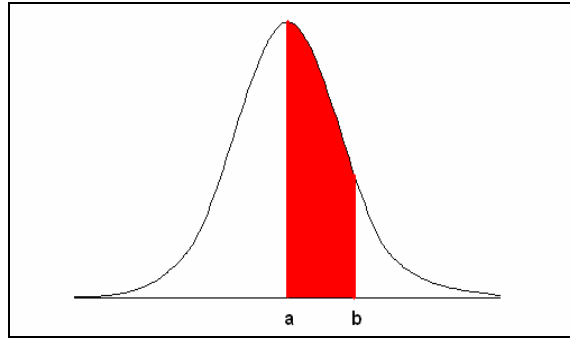
$$q_k = \frac{(125970) (28987537000)}{5.3598337E20}$$

$$= 6.8128233E-6$$

التوزيع الطبيعي Normal distribution

يظهر مصطلح المتغير العشوائي المستمر Continuous random variable عندما يتم التعامل مع الكميات التي تقاس بمقاييس مستمرة. على سبيل المثال عند قياس عمر شخص معين أو الوزن.... الخ. في الجانب العملي غالباً ما يتم تدوير القيم إلى أقرب عدد صحيح، بعبارة أخرى يمكن أن يقال أن المتغير (تحت الدراسة) متغير مستمر في حالة كون القيم التي يتعامل معها هي قيم مستمرة وتحدد بفترات. في هذه الحالة (الحالة المستمرة) يشار إلى الاحتمالية من خلال مفهوم المساحة.

إن المنحنيات المستمرة عبارة عن أشكال بيانية ناتجة عن دوال والتي يشار إليها بالتوزيعات المستمرة. يرمز إلى التوزيعات الاحتمالية المستمرة بالمساحة تحت المنحنى بين قيمتين على سبيل المثال a و b، (شكل 11.4)



شكل 11.4 : المنحنى الطبيعي فيه المساحة تحت المنحنى بين a و b

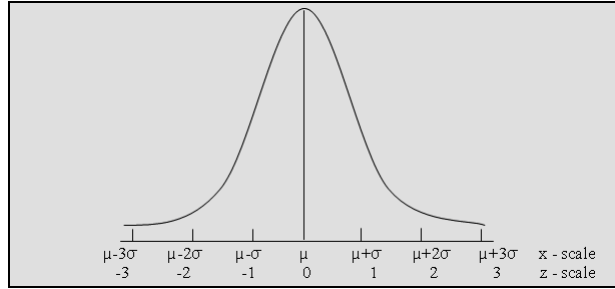
إن التوزيع الطبيعي أو في بعض الأحيان يشار إليه بتوزيع كاوسيان Gaussian distribution له أهمية كبيرة وذلك :

1- له العديد من التطبيقات في الواقع العملي.

2- ان العديد من متوسطات العينات التي لها توزيعات تتبع التوزيع الطبيعي، وذلك من خلال إجراء التقارب بين هذه التوزيعات والتوزيع الطبيعي.
ان دالة الكثافة الاحتمالية Probability density function(pdf) للمتغير العشوائي المستمر X هي عبارة عن صيغة لإيجاد احتمالية x ، وكالاتي:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

حيث ان μ هو المتوسط و σ هو الانحراف المعياري للتوزيع الطبيعي، و π ثابت قيمته 3.14159 تمثل دالة الكثافة الاحتمالية الطبيعية (شكل 12.4) على شكل متمائل حول المتوسط.



شكل 12.4 : توزيع المنحنى الطبيعي وفيه المساحة مقسمة حسب المقاييس الطبيعية والقياسية

ان الميزة المهمة للتوزيع الطبيعي ان المعادلة الرياضية يمكن تحديدها بصورة تامة اذا عرفت قيمتي الوسط والانحراف المعياري .
من الجانب العملي، يتم استخراج المساحة تحت المنحنى للرسم البياني للتوزيع الطبيعي بالاستعانة بجدول قيم z (الجدول ملحق I) . لغرض استخدام هذا الجدول يجب التحويل من الصيغة الطبيعية (التوزيع الطبيعي) الى التوزيع القياسي الطبيعي Standard normal distribution وكالاتي:

$$z = \frac{x - \mu}{\sigma}$$

مقاييس النزعة المركزية Measures of central tendency

غالبا ما تكون هنالك ضرورة من تمثيل مجموعة من البيانات بمقاييس متمثلة برقم مفرد واحد وهذا الرقم يتمركز في منتصف البيانات ليعبر عنها. وان أهم هذه المقاييس هي المتوسط والوسيط والمنوال.

المتوسط والوسيط والمنوال Mean , Median , and Mode

عند وصف البيانات ، يتجه التفكير الى قيمة المتوسط او المعدل او القيمة التي نتوسط البيانات للتعبير عنها. هنالك ثلاثة أنواع من مقاييس النزعة المركزية. المتوسط ($\bar{x}$) وهو عبارة عن مجموع جميع القيم المشاهدة للمتغير (تحت الدراسة) مقسوم على عدد المشاهدات.

مثال 15.4: الجدول (5.4) يمثل قيم انقباض ضغط الدم Systolic blood pressure

116	116	115	110	105	104	100	100
128	128	124	122	120	120	120	116
138	134	130	130	130	130	130	130
144	140	140	140	140	140	140	140
170	170	152	150	150	150	150	144

احسب معدل انقباض ضغط الدم لـ 40 مريض.
المعدل هو:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{40} x_i}{40} = \frac{x_1 + x_2 + x_3 + \dots + x_{40}}{40} \\ &= \frac{100+100+104+\dots+170}{40} = \frac{5256}{40} = 131.4 \text{ mm Hg}\end{aligned}$$

ان الشهرة التي يتمتع بها مقياس المتوسط او المعدل كونه مقياس الوسط او المركز لمجموعة من البيانات ، لم تكن صفة عرضية ، إضافة الى ذلك فهو مقياس بسيط وله العديد من الخصائص الآتية:

- 1- يمكن حسابه لأي مجموعة رقمية من البيانات، وبهذا فانه موجود دائما.
- 2- كل مجموعة رقمية من البيانات لها معدل واحد وواحد فقط ، وبهذا فان هذا المقياس هو وحيد دائما.
- 3- يمكن إجراء معالجات حسابية أخرى او إضافية عليه. مثال ذلك يمكن حساب المعدل العام لمجموعة من المعدلات المختلفة المحسوبة من عدد من المجموعات المختلفة بعبارة أخرى احتساب معدل المعدلات.
- 4- يأخذ في حساباته كل عنصر في البيانات ، وهو بالتالي لا يهمل أية قيمة مهما كانت صغيرة او كبيرة.
- 5- يمكن الوثوق نسبيا في حالة احتساب عدد من المعدلات الخاصة بعدد من العينات المختلفة المسحوبة من مجتمع واحد بان تكون قيمها (قيم المتوسطات) غير متذبذبة او مختلفة اختلافا كبيرا.

مثال (16.4): أعمار سبعة طلاب هي 20 ، 18 ، 19 ، 20 ، 17 ، 19 ، 18 وكان عمر التدريسي الذي يدرسه 53. فان معدل الأشخاص الثمانية هو :

$$\frac{20+18+19+20+17+19+18+53}{8} = \frac{184}{8} = 23$$

يلاحظ ان قيمة المعدل هي 23. وبالمقابل فانه لا توجد أية قيمة من قيم مجموعة البيانات مساوية الى هذه القيمة او حتى مقاربة لها وذلك لتأثرها بالقيم التي تعد شاذة (53 سنة) بالنسبة لقيم أعمار الطلاب المتقاربة .

مثال (17.4): اذا كان المطلوب حساب معدل عمر منتج من المصاييح عددها $N = 40000$ مصباح (N تمثل حجم المجتمع). ومن غير المعقول ان يتم اختيار جميع هذه المصاييح لمعرفة حساب معدل العمر. لهذا يتم سحب عينة ، لاحتساب متوسط العينة ($\bar{x}$) واستخدام هذه الاحصاءة (المقياس) لتقدير قيمة معدل المجتمع (μ). فاذا كانت $n=5$ (حجم العينة) وكان مدة اشتغال المصاييح الخمسة هي : 967 , 949 , 940 , 952 , 922 ساعة اشتغال ، فانه:

$$\bar{x} = \frac{967+949+940+952+922}{5} = \frac{4730}{5} = 946 \text{ ساعة اشتغال}$$

إذن يمكن تقدير (μ) معدل عمر الاشتغال للمجتمع (N) بمقدار 946 ساعة اشتغال. اما في حالة سحب عينة أخرى من المجتمع نفسه بالحجم نفسه ($n=5$) او بأحجام مختلفة فقد تكون قيمة $\bar{x}$ مساوية (لقيمة 946) او قد تكون مختلفة تبعا الى قيم مفردات العينة المسحوبة (العينة الثانية). نفترض انه في حالة ان القيمة الثانية (عمر اشتغال المصباح الثاني) قد سجل خطأ 499 بدلا من 949 وعند حساب قيمة المعدل فانه:

$$\bar{x} = \frac{967+499+940+952+922}{5} = \frac{4280}{5} = 856 \text{ ساعة اشتغال}$$

وبهذا يصبح لدينا خطأ مقداره 90 ساعة اشتغال ($946 - 856 = 90$) عند تقدير معدل عمر اشتغال المصاييح الخاص بالمجتمع.

ولتجنب مثل هذه الأخطاء الناتجة من قيم صغيرة او كبيرة جدا ، يمكن في بعض الأحيان وصف او تحديد قيمة الوسط او المركز لمجموعة من البيانات باستخدام مقاييس إحصائية أخرى . إحدى هذه المقاييس هو الوسيط Median، والذي يعرف بأنه القيم التي تقع في منتصف البيانات في حالة كون مجموعة البيانات فردية او بأنه متوسط قيمتي المنتصف في حالة كون مجموعة البيانات زوجية في حالة ترتيب البيانات تصاعديا او تنازليا.

مثال(18.4): في حالة احتساب معدل عمر اشتغال المصاييح للعينة المسحوبة من مجتمع حجمه $N=40000$. يحسب الوسيط كما يأتي

الخطوة الأولى: ترتيب البيانات تنازليا

967 , 952 , 949 , 940 , 922

الخطوة الثانية: اختبار قيمة المنتصف (لكون عدد البيانات فردي ويساوي 5) وبهذا فان قيمة الوسيط هي 949. وبهذا فان قيمة الوسيط (949) مقاربة الى قيمة الوسط 946 والفرق بينهما 3.

اما في حالة المثال السابق الذي سجلت فيه القيمة الثانية خطأ 499 بدلا من 949. لغرض حساب الوسيط يتبع الآتي:

الخطوة الأولى: ترتيب البيانات ترتيب تنازلي

967 , 952 , 940 , 922 , 499

الخطوة الثانية : اختبار قيمة المنتصف. وهي القيمة 940 ، وبهذا فان قيمة الوسيط الحالية (بعد إدخال القيمة الخطأ) مقاربة الى قيمة الوسيط (القيمة الصحيحة) والفرق بينهما هو $(949 - 940 = 9)$. وأيضا فان قيمة الوسيط الحالية 940 مقاربة الى قيمة الوسط 946 والفرق بينهما هو $6 (946 - 940 = 6)$.

مثال (19.4): لغرض حساب قيمة الوسيط لأعمار ثمانية أشخاص (سبعة طلاب مع التدريسي الذي يدرسهم في المثال (14)) يتبع الآتي:

الخطوة الأولى: ترتيب القيم ترتيبا تصاعديا

17 , 18 , 18 , 19 , 19 , 20 , 20 , 53

الخطوة الثانية: إيجاد قيمة الوسيط وذلك بتحديد موقعه (عدد القيم زوجي) حيث ان:

$$\frac{n+1}{2} = \frac{8+1}{2} = 4.5$$

وبهذا فان قيمة الوسيط تقع بين المرتبتين الرابعة والخامسة ومقدارها:

$$\frac{19+19}{2} = 19$$

وهي قيمة تمثل أغلب قيم البيانات باستثناء عمر التدريسي (التي تعد قيمه شاذة مقارنة مع باقي القيم الأخرى). ان الخاصية الأكثر أهمية في الفرق بين المتوسط والوسيط هي مشاكل التقدير والتنبؤ و... الخ. حيث انه يمكن التعديل على الوسط أكثر من الوسيط في هذه الحالة. وذلك لان قيم الوسيط لعدة عينات مسحوبة من مجتمع واحد عادة تختلف اختلافا كبيرا مقارنة مع قيم الوسط للعينات نفسها.

مقياس آخر، يمكن استخدامه لوصف المنتصف لمجموعة من البيانات هو المنوال Mode الذي يمكن تعريفه ببساطة بأنه القيمة التي تشير للتكرار الأكثر حدوثا. يمتاز هذا المقياس بفائدتين رئيسيتين هما:

1- ان هذا المقياس لا يحتاج الى إجراء حسابات.

2- يمكن حسابه للبيانات الكمية والوصفية معا.

مثال (20.4): القيم التالية، تمثل أعمار مجموعة من الكائنات الحية:

7 , 8 , 9 , 7 , 5 , 10 , 12 , 7 , 9 , 12 , 7

فان قيمة المنوال هي 7 والتي تكررت أكثر القيم (4 مرات).

مثال (21.4): القيم التالية تمثل أعمار مجموعة من الكائنات الحية:

10 , 3 , 2 , 1 , 3 , 4 , 10 , 1 , 3 , 10

ان القيمتين 3 و 10 تكررت أكثر من القيم الأخرى (ثلاث مرات لكل من القيمتين) وبهذا لدينا منوالين هما 3 و 10.

مثال (22.4): القيم التالية تمثل أعمار مجموعة من الكائنات الحية:

9, 7, 8, 6, 4, 10, 11, 14, 1, 0, 5

لا يوجد منوال في هذه الحالة وذلك لعدم تكرار أية قيمة من قيم مجموعة البيانات.

مقاييس الاختلاف Measures of variation

هي واحدة من أهم الصفات والمميزات لأية مجموعة من البيانات ، حيث ان قيم المجموعة ليست متساوية جميعها وبالتالي فإنها قيم متباينة. ان مقاييس النزعة المركزية تصف قيمة مهمة واحدة فقط هي منتصف هذه المجموعة من البيانات. كل منها بالمقابل لا تصف أي شئ عن المميزات او الصفات الأساسية لهذه المجموعة. بعبارة أخرى هنالك حاجة الى وجود طرق لقياس مقدار امتداد البيانات او ابتعادها عن المركز (مركز او متوسط البيانات). يسمى هذا المقياس الذي يقيس هذه المعلومة بمقياس الاختلاف.

مثال (23.4): في إحدى مختبرات التجارب كان معدل درجة حرارة تجربة معينة هي 75 درجة مئوية ، وهي الدرجة المرغوبة لإجراء هذه التجربة. لكن واقع الحال فان قيمة المعدل هذه قد تكون ناتجة من درجة حرارة 35 وعدد كبير من درجات الحرارة المساوية الى 100 وبالتالي فان قيمة المعدل لا يعكس واقع حال ظروف التجربة المطلوبة. إذن لابد من وجود مقياس (مقاييس) لقياس مقدار التذبذب الحاصل في البيانات.

مثال (24.4): بافتراض ان إحدى المستشفيات أخذت النبضات لمريضين في الصباح وعند الظهر وعند المساء وسجلت كما يأتي:

جدول (6.4) عدد نبضات مريضين في ثلاثة أوقات مختلفة ليوم معين

عدد النبضات (ضربة في الدقيقة)			
المساء	الظهر	الصباح	
74	76	72	المريض الأول
59	91	72	المريض الثاني

ان معدل نسب النبضات للمريضين متساوي ومقداره 74 لكن هنالك اختلافاً كبيراً بين قيم المجموعتين. حيث ان نسب نبضات المريض الأول تبدو أكثر استقراراً مقارنة مع نسب نبضات المريض الثاني.

المدى The range

يعد مقياس المدى احد مقاييس التشتت او الاختلافات ويمكن تعريفه بأنه الفرق الناتج بين اكبر قيمة واصغر قيمة لمجموعة من البيانات.

مثال (25.4): تحسب قيمة المدى لقياس نسب نبضات المريضين في المثال (24.4) كالآتي:

$$R1 = 76 - 72 = 4 \text{ المريض الأول}$$

$$R2 = 91 - 59 = 32 \text{ المريض الثاني}$$

وبهذا يلاحظ مقدار الاختلاف الكبير بين قيم المجموعتين من البيانات على الرغم من ان قيمة المتوسط للمجموعتين متساوي.

الانحراف المعياري والتباين The standard deviation and the variance

يعرف مقياس الانحراف المعياري بأنه عبارة عن الجذر التربيعي لمجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوم على حجم العينة. ويمكن التعبير عن هذا المقياس بصيغة الرموز كما يلي:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (9.4)$$

حيث ان:

S: الانحراف المعياري

x_i : قيم المشاهدات $i = 1, 2, 3, \dots, n$

$\bar{x}$: متوسط قيم بيانات هذه المجموعة

n: حجم العينة

وقد طورت المعادلة (9.4) لتصبح بالشكل الآتي:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (10.4)$$

حيث استخدام المقدار $n-1$ بدلا من n للإشارة الى استبعاد قيمة الوسط الحسابي من حجم العينة.

وعند تربيع المعادلة (10.4) نحصل على التباين ، وبصيغة الرموز:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad (11.4)$$

مثال (26.4): احد الباحثين في المجال البايولوجي وجد عدد من الكائنات المجهرية 6 و 12 و 9 و 10 و 9 و 8 في ست نماذج مختلفة ولغرض حساب الانحراف المعياري والتباين يتبع الآتي:

الخطوة الأولى: حساب قيمة المتوسط ،

$$\bar{x} = \frac{6+12+9+10+9+8}{6} = \frac{54}{6} = 9$$

قيمة منتصف عدد الكائنات المجهرية لكل النماذج ، بعبارة أخرى فان لكل نموذج سيكون هنالك عدد متساوي فيها من الكائنات الحية ومقدارها 9.

الخطوة الثانية : إيجاد ناتج المقدار $\sum (x_i - \bar{x})^2$ كما يأتي:

جدول 7.4: احتساب مجموع مربعات القيم عن وسطها الحسابي

$(x_i - \bar{x})^2$	$(x_i - \bar{x})$	X
9	-3	6
9	3	12

0	0	9
1	1	10
0	0	9
1	-1	8
20	0	المجموع

ثم تحسب قيمة كل من الانحراف المعياري و التباين وكالاتي:

$$S = \sqrt{\frac{20}{5}} = 2$$

$$S^2 = (2)^2 = 4$$

معاملات الارتباط Correlation coefficients

في الكثير من البحوث المهمة، يتوجب على الباحث إجراء قياسات تتعلق بعينات تحتوي على زوج من القيم (تتألف بيانات البحث من متغيرين فقط على سبيل المثال X و Y). بعبارة أخرى ان الهدف من إجراء البحث (الذي يحتوي على متغيرين فقط) هو إيجاد العلاقة بين هذين المتغيرين. على سبيل المثال ما هي العلاقة بين وزن الطفل عند الولادة وبين مقدار الزيادة في الوزن للمدة من 70 يوم الى 100 يوم بعد ولادته؟ لغرض الإجابة على هذا السؤال والأسئلة المشابهة يستخدم مقياس الارتباط لتحديد العلاقة بين المتغيرين ومدى قوتها.

مثال (27.4) : يوضح الجدول (8.4) قيم أوزان الأطفال المولودين (X) والزيادة في الوزن للمدة من 70 يوم الى 100 يوم بعد ولادته ، التي يعبر عنها كنسبة مئوية (Y) وقد أخذت البيانات لـ 12 طفل .

جدول (8.4) : أوزان الأطفال حديثي الولادة (X) ونسبة الزيادة في الوزن (Y)

التسلسل (الأطفال)	الوزن (X)	نسبة الزيادة (Y%)
1	112	63
2	111	66
3	107	72
4	119	52
5	92	75
6	80	118
7	81	120
8	84	114
9	118	42
10	106	72
11	103	90
12	94	91

لغرض تمثيل هذه البيانات على شكل رسم بياني أو مخطط بياني يرمز الى زوج البيانات (X, Y) على شكل نقطة في الرسم البياني ان البيانات جميعها لا تقع على خط مستقيم واحد ، لكنها تنتشر حول خط. ان مواقع هذه النقاط تعكس بعض المعلومات حول اتجاه وأيضاً قوة ارتباط البيانات تحت الدراسة (X و Y). فإذا كان اتجاه البيانات يتجه من الجهة اليسرى السفلى الى الجهة العليا اليمنى تكون العلاقة بين البيانات موجبة ، في حين اذا كان اتجاه البيانات من الجهة العليا اليسرى الى الجهة السفلى اليمنى بذلك تكون العلاقة بين البيانات سالبة. وتبدأ العلاقة بين المتغيرين بالضعف شيئاً فشيئاً كلما كان توزيع (انتشار النقاط) البيانات بالاتجاه الأفقي أو قريباً منه (شكل 13.4)



شكل 13.4 : محور اتجاه البيانات

معامل ارتباط بيرسون Pearson's correlation coefficient

بافتراض ان مجموعة من البيانات (المشاهدات) بحجم n (حجم العينة أو المفردات المسحوبة من مجتمع) لمتغيرين (X و Y) فان معامل الارتباط بين المتغيرين هو:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}} \quad (12.4)$$

حيث ان:

$$-1 \leq r \leq 1$$

ويمكن ايضا استخدام الصيغة البديلة التالية لإيجاد قيمة r

$$r = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{[\sum x^2 - (\sum x)^2/n][\sum y^2 - (\sum y)^2/n]}} \quad (13.4)$$

ان الصيغتين في المعادلتين (12.4) و (13.4) يطلق عليها معامل ارتباط بيرسون. وتشير قيمة معامل ارتباط القرية الى الواحد (أو مساوية بالضبط الى الواحد الصحيح) الى ان العلاقة بين المتغيرين (X و Y) علاقة طردية وقوية. والمقصود بالعلاقة انها طردية موجبة ،

هي انه في حالة زيادة احد المتغيرين (على سبيل المثال المتغير X) تقابله زيادة في المتغير الثاني (هنا في هذه الحالة يكون المتغير Y) اما مقدار الزيادة فهي تعتمد على مقدار قيمة معامل الارتباط. في حين تشير قيمة معامل الارتباط القريب الى سالب واحد (او مساوية بالضبط الى سالب واحد) الى ان العلاقة بين المتغيرين (X و Y) علاقة عكسية قوية. والمقصود بالعلاقة انها عكسية قوية، انه في حالة زيادة احد المتغيرين (على سبيل المثال المتغير X) تقابلها نقصان بالمتغير الثاني (في هذه الحالة المتغير Y) ويعتمد مقدار النقصان على قيمة معامل الارتباط r.

وتشير قيم معامل الارتباط في حالة اقترابها من الصفر (او مساوية بالضبط الى القيمة صفر) الى ان العلاقة بين المتغيرين (X و Y) ضعيفة (او عدم وجود علاقة في حالة كون قيمة معامل الارتباط مساوية الى الصفر).

في حالة بيانات الأطفال الرضع حديثي الولادة (مثال 27.4) ، المدرجة في جدول (8.4). ولغرض استخراج قيمة معامل الارتباط r (باستخدام المعادلة رقم (13.4)) من الجدول (9.4) .

جدول (9.4) استخراج قيمة معامل الارتباط

التسلسل	X	Y	X ²	Y ²	XY
1	112	63	12544	3969	7056
2	111	66	12321	4356	7326
3	107	72	11449	5184	7704
4	119	52	14161	2704	6188
5	92	75	8464	5625	6900
6	80	118	6400	13924	9440
7	81	120	6561	14400	9720
8	84	114	7056	12996	9576
9	118	42	13924	1764	4965
10	106	72	11236	5184	7632
11	103	90	10609	8100	9270
12	94	91	8836	8281	8554
المجموع	1207	975	123561	86487	94322

حيث ان X تمثل وزن الطفل ، في حين ان Y تمثل نسبة الزيادة في الوزن للطفل حديث الولادة للمدة من 70 الى 100 يوم.

$$r = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{[\sum x^2 - (\sum x)^2/n][\sum y^2 - (\sum y)^2/n]}}$$

$$= \frac{94322 - (1207)(975) / 12}{\sqrt{[123561 - (1207)^2 / 12][86487 - (975)^2 / 12]}}$$

$$= \frac{94322 - 98068.75}{\sqrt{[123561 - 121404.08][86487 - 79218.75]}}$$

$$= -0.946$$

تؤشر قيمة معامل الارتباط الى وجود علاقة عكسية قوية بين المتغيرين. بمعنى انه كلما كان وزن الطفل (X) كبيرا عند الولادة كلما كان مقدار الزيادة في وزنه (Y) قليلة والعكس بالعكس.

مثال (28.4): تمثل البيانات المدرجة في جدول (10.4) متغير العمر (X) وقراءات انقباض ضغط الدم (Systolic blood pressure) لـ 15 امرأة (Y). المطلوب حساب معامل الارتباط.

جدول (10.4) قراءات انقباض ضغط الدم لـ 15 امرأة

الترتيب	العمر (X)	ضغط الدم (Y)
1	42	130
2	46	115
3	42	148
4	71	100
5	80	156
6	74	162
7	70	151
8	80	156
9	85	162
10	72	158
11	64	155
12	81	160
13	41	125
14	61	150
15	75	165

حيث ان :

x: العمر بالسنين

y: ضغط الدم

لغرض حساب قيمة r تحتسب المؤشرات الموضحة في الجدول (11.4):

جدول (11.4) المؤشرات لاحتساب قيمة r

التسلسل	X	Y	X ²	Y ²	XY
1	42	130	1764	16900	5460
2	46	115	2116	13225	5290
3	42	148	1764	21904	6216
4	71	100	5041	10000	7100
5	80	156	6400	24336	12480
6	74	162	5476	26244	11988
7	70	151	4900	22801	10570
8	80	156	6400	24336	12480
9	85	162	7225	26244	13770
10	72	158	5184	24964	11376
11	64	155	4096	24025	9920
12	81	160	6561	25600	12960
13	41	125	1681	15625	5125
14	61	150	3721	22500	9150
15	75	165	5625	27225	12375
المجموع	984	2193	67954	325929	146260

وباستخدام المعادلة (13.4) وكالاتي:

$$r = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{[\sum x^2 - (\sum x)^2/n][\sum y^2 - (\sum y)^2/n]}}$$

$$\begin{aligned}
&= \frac{146260 - \left[(984) (2193) \right] / 15}{\sqrt{\left[67954 - (984)^2 / 15 \right] \left[325929 - (2193)^2 / 15 \right]}} \\
&= \frac{146260 - 143860.8}{\sqrt{\left[67954 - 64550.4 \right] \left[325929 - 320616.6 \right]}} \\
&= 0.564
\end{aligned}$$

تدل قيمة معامل الارتباط الى وجود علاقة طردية معتدلة بين المتغيرين (X و Y) ، بمعنى اذا كانت هنالك زيادة في احد المتغيرين (X) ستصاحبه زيادة في المتغير الثاني (Y) بمقدار 0.564.

معامل الارتباط اللامعلمي Non parametric correlation coefficient

قد يتجه اهتمام الباحثين في بعض الأحيان الى حساب مقاييس مرافقة ، هي أكثر قوة فيما يتعلق بقيم البيانات . ومثال على هذا النوع من المقاييس هو مقياس سبرمان راس Spearman's Rho . ولغرض حساب هذا المقياس تتبع الخطوات الآتية :

أولاً: ترتيب قيم المتغير X تصاعدياً من اصغر قيمة الى اعلي قيمة وتحديد رتب القيمة. تحدد القيمة 1 الى اصغر قيمة وهكذا وصولاً الى اكبر قيمة تحدد لها القيمة n . يفترض ان R_i تمثل الرتبة التي تقابل القيمة x_i . وبصورة مماثلة بالنسبة للمتغير Y الذي ترتب قيمه بصورة تصاعدياً من 1 الى n لكل قيمه ايضاً.

يفترض ان S_i هي الرتبة التي تقابل القيمة y_i . وفي حالة تساوي قيمتين (او أكثر) من قيم المتغير ، عندها يتم استخدام المعدل لتحديد الرتبة لقيم هذه المتغيرات. ثانياً : حساب صيغة معامل ارتباط سبرمان وكالاتي :

$$p = 1 - \frac{6 \sum (R_i - S_i)^2}{n (n^3 - 1)} \quad (14.4)$$

مثال (29.4) تمثل البيانات المدرجة في الجدول (12.4) حول الأطفال الرضع في المثال (27.4)

جدول (12.4) بيانات الأطفال الرضع ، تمثل X وزن الطفل الرضيع وتمثل R رتبة المتغير X ، وتمثل Y الزيادة في وزن الطفل ، وتمثل S رتبة المتغير Y

التسلسل (الأطفال)	الوزن عند الولادة		الزيادة في الوزن		R	X
	(R-S) ²	R-S	S	Y		
1	49	7	3	63	10	112
2	25	5	4	66	9	111
3	6.25	2.5	5.5	72	8	107
4	100	10	2	52	12	119
5	9	-3	7	75	4	92
6	100	-10	11	118	1	80
7	100	-10	12	120	2	81
8	49	-7	10	114	3	84
9	100	10	1	42	11	118
10	2.25	1.5	5.5	72	7	106
11	4	-2	8	90	6	103
12	16	-4	9	91	5	94
المجموع	560.5					

الخطوة التالية هي تطبيق صيغة معامل الارتباط (14.4) وكما يأتي :

$$p = 1 - \frac{6 \sum (R_i - S_i)^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6 (560.5)}{12 (143)}$$

$$= -0.96$$

تشير هذه القيمة الى وجود علاقة عكسية و قوية بين المتغيرين (Y و X).
يلاحظ ان القيمة المستخرجة عن طريق استخدام معامل الارتباط سبيرمان (-0.96) مقارنة الى
قيمة معامل الارتباط بيرسون (-0.946) وبالاتجاه نفسه ايضا.

معنوية الاختبارات الإحصائية Statistical test significance

هناك الكثير من الحالات التي يدخل فيها موضوع الاختبار ، علي سبيل المثال فأن الطالب يتم اختباره في المدرسة ، وفي المستشفى يتم سحب الدم من الشخص لغرض إرساله الى المختبر لغرض اختباره ، والعربات والشاحنات يتم اختبارها في المعمل لغرض اختبار أدائها لمعرفة سلامتها قبل استعمالها . وهكذا فان من الطبيعي ان يكون الاختبار الإحصائي هو الهدف المطلوب إجراءه على مجموعة من البيانات، تطبق الاختبارات الإحصائية لانها تعطي قرار حول طبيعة مشكلة البحث .

لغرض شرح موضوع الاختبارات الإحصائية ، يستخدم مفهوم الإجراءات الجنائية . حيث ان المجرم بريء حتى تثبت إدانته . وكما هو معلوم فان القرارات التي تصدر من المحكمة علي سبيل المثال في بعض الأحيان قد تكون خاطئة .. وهكذا الحال في حالات الاختبارات الإحصائية ، البعض من نتائجها تكون خاطئة . ان الخطأ الذي يحدث هو نتيجة لوجود فروق معنوية إحصائية (وجود فروق كبيرة بين الظاهرتين او الحالتين تحت الدراسة) ، في حين ان واقع الأمر هو العكس من ذلك . كما هو الحال عند إصدار المحكمة التهمة على شخص بريء ، يدعى هذا الخطأ بالخطأ من النوع الأول Type I error ، حيث انه غالبا ما تبني الاختبارات الإحصائية بحيث ان الخطأ من النوع الأول يمكن ان يحدث ضمن مستويات 1% او 5% او 10% . بعبارة أخرى انه في حالة تكرار تجربة معينة 100 مرة فأن مقدار الخطأ المسموح به الذي يجب ان يحدد مسبقا وقبل البدء بالتجربة مرة واحدة (في حالة مستوى الخطأ المسموح به 1%) وخمسة مرات (في حالة مستوى الخطأ المسموح به 5%) او عشر مرات (في حالة مستوى الخطأ المسموح به 10%) . في حين ان الخطأ من النوع الثاني هو الخطأ في حالة عدم الحصول على فروق معنوية إحصائية في الوقت الذي توجد فيه (في واقع الحال) فروق كبيرة بين الظاهرتين تحت الدراسة كما هو الحال عند إصدار المحكمة قرارا بإطلاق المذنب .

فرضيات الاختبار

عند البدء بإجراء البحوث لغرض حل مشكلة معينة ، تبدأ الإجراءات بتحديد الفرضية الخاصة بالبحث . وفي المجال الإحصائي يطلق على الفرضية المطلوب اختبارها بفرضية العدم Null hypothesis ويرمز لها بـ H_0 . ان تسمية العدم يدل على عدم وجود فروق معنوية او عدم وجود علاقة بين المعالم (تحت الدراسة) . كما هو الحال في القول ان المتهم بريء حتى تثبت إدانته ، مثال على ذلك :

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 - \mu_2 = 0$$

حيث ان H_0 ترمز الى فرضية العدم وان μ_1 و μ_2 هما معلمتي المتوسط لكل من المجتمعين الأول والثاني ، وقد حددت هذه الفرضية على أساس تساوي المعلمتين ($\mu_1 = \mu_2$) (عدم وجود فرق بينهما $\mu_1 - \mu_2 = 0$) ان فرضية الاختبار هي عملية صنع القرار التي تختبر مجموعة (او مجموعات) من البيانات الذي يستند على فرضية العدم H_0 .

في حالة رفض الفرضية الأولى (فرضية العدم H_0) ننتقل لصنع القرار الخاص باختبار مجموعة البيانات تحت الدراسة الى الفرضية البديلة. (البديلة عن فرضية العدم). Alternative hypothesis ويرمز لها بالرمز H_A . تحت الفرضية البديلة ، فان الفروق المتشابهة (الملاحظة في البيانات) هي حقيقة (مثال ذلك، فان $X_1 \neq X_2$ هي ليست من قبيل المصادفة وانما يعود السبب في ذلك الى ان $\mu_1 \neq \mu_2$).
ترفض فرضية العدم في حالة وجود دلائل إحصائية قوية من البيانات (تحت الدراسة) تساند الفرضية البديلة.

الدليل الإحصائي Statistical evidence

غالبا ما تتعلق فرضية العدم بمعلمة او معلمات المجتمع ، (يشير مصطلح المعلمة أو المعلمات الى المقاييس الخاصة بالمجتمع مثل المتوسط " μ " والانحراف المعياري " σ ") ، على أية حال فان دراسة البيانات الكلية للمجتمع اما ان يكون مستحيلا او يكون ذا كلفة عالية لغرض معرفة فيما اذا كانت فرضية العدم صحيحة ام لا . ولهذا السبب فان صنع القرار يستند على سحب عينة من مجتمع الدراسة وان بيانات العينة المسحوبة تلخص من خلال احصاءة او إحصاءات (يشير مصطلح احصاءة أو إحصاءات إلى المقاييس الخاصة بالعينة ، مثل المتوسط الحسابي $(\bar{X})$ والانحراف المعياري (δ)) تستخدم لتقدير المعلمة او المعلمات المتضمنة في فرضية العدم. على سبيل المثال ، اذا كانت فرضية العدم حول μ (مثال ذلك $H_0 : \mu = 10$) فان الموقع المناسب للبحث عن المعلومات عن μ هو $\bar{X}$ في هذا السياق ، فان الاحصاءة $\bar{X}$ تدعى احصاءة الاختبار Test stastic . حيث انه تستخدم لقياس الاختلاف بين البيانات (حيث ان القيمة الرقمية لـ $\bar{X}$ تستخرج من العينة) وما هو المتوقع في حالة كون فرضية العدم صحيحة (بعبارة أخرى اذا كانت $\mu = 10$ صحيحة) ، يطلق على هذه العملية بالدليل الإحصائي . وهي تختلف (قيمة الاحصاءة $\bar{X}$ على سبيل المثال) من عينة الى أخرى (مسحوبة جميعها من مجتمع واحد).

الأخطاء Errors

ان فرضية العدم H_0 قد تكون صحيحة او خاطئة والقرار المتخذ بشأن هذه الفرضية هو قبولها او الرفض ، وبهذا يكون لدينا أربع احتمالات (قرارات) ممكنة (جدول 13.4) اثنان منها قرارات صحيحة هي :

- 1- عدم رفض فرضية العدم H_0 عندما تكون H_0 صحيحة
- 2- رفض فرضية العدم H_0 عندما تكون H_0 خاطئة.

جدول 13.4: الاحتمالات الخاصة بفرضية العدم H_0

المصادقية	القرار	
	عدم الرفض H_0	رفض H_0
فرضية العدم صحيحة	قرار صحيح	الخطأ من النوع الأول (α)
فرضية العدم خاطئة	الخطأ من النوع الثاني (β)	قرار صحيح

وبالمقابل هناك حالتين يكون القرار فيهما خاطئاً هي :

1- الخطأ من النوع الأول: رفض فرضية العدم H_0 عندما تكون صحيحة.

2- الخطأ من النوع الثاني: عدم رفض فرضية العدم H_0 عندما تكون خاطئة.

ان الهدف العام من اختبار الفرضية هو الإبقاء على قيمتي (الاحتمالين) α (الخطأ من النوع الأول) و β (الخطأ من النوع الثاني) صغيرة قدر الإمكان ، (عند اتخاذ قرار بتقليص حجم α (قيمة α) سيزداد حجم β (قيمة β) والعكس بالعكس). في العديد من البحوث يتم تثبيت قيمة α عند قيم معينة (مستويات) (0.01 أو 0.05 أو 0.10) في حين ان قيمة β يتم السيطرة عليها من خلال حجم العينة .

مثال (30.4) : بافتراض ان نسبة التدخين العالمي للرجال هي 25% وأجريت دراسة لتحديد نسبة التدخين للرجال في منطقة الوطن العربي ، والتي يرمز لها بالرمز π (نسبة التدخين للرجال في منطقة الوطن العربي) ، وحددت فرضية العدم بان نسبة التدخين في منطقة الوطن العربي مساوية لنسبة التدخين العالمي والتي يعبر عنها بالرموز كما يأتي :

$$H_0: \pi = 0.25$$

بافتراض ان الباحث قد خطط لسحب عينة بحجم $n=100$ مدخن يقطنون منطقة الوطن العربي ، واعتمد قاعدة صنع القرار الآتية :

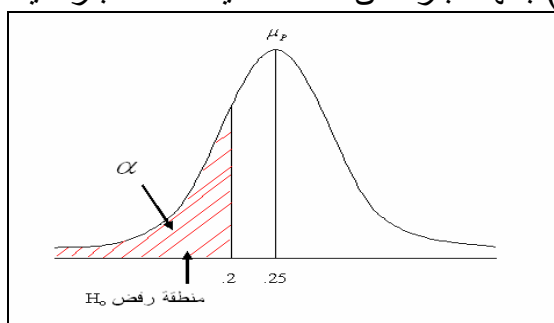
رفض فرضية العدم إذا كانت نسبة التدخين أقل أو تساوي 0.20

حيث ان P_s تمثل نسبة التدخين في العينة المسحوبة حيث يتم الحصول على قيمتها من العينة المسحوبة من المدخنين الرجال القاطنين في منطقة الوطن العربي المطلوب :

أ. استخراج قيمة α

ب. استخراج قيمة β

أ. تعرف ألفا (α) بأنها عبارة عن الاحتمالية الخطأ بفرضية العدم الصحيحة



شكل 14.4 : توزيع المنحنى الطبيعي بوسط حسابي $\mu_p = 0.25$ يظهر فيه قيمة الخطأ من النوع الأول α

بعبارة أخرى $\alpha = P_r (P_s \leq 0.20 / \pi = 0.25)$
 طالما ان حجم العينة $n=100$ وهي كبيرة بحيث يمكن معها تطبيق نظرية النزعة المركزية
 Central limit theorem ، حيث ان توزيع العينة الى P_s هو مقارب الى التوزيع الطبيعي له
 معلمتين هما الوسط والتباين استنادا الى فرضية العدم ، فان :

$$\begin{aligned}\mu_{P_s} &= \pi \\ &= 0.25 \\ \sigma_{P_s}^2 &= \frac{\pi(1-\pi)}{n} \\ &= \frac{0.25(0.75)}{100} \\ &= 0.00188\end{aligned}$$

وباستخدام قاعدة صنع القرار وكما يأتي :

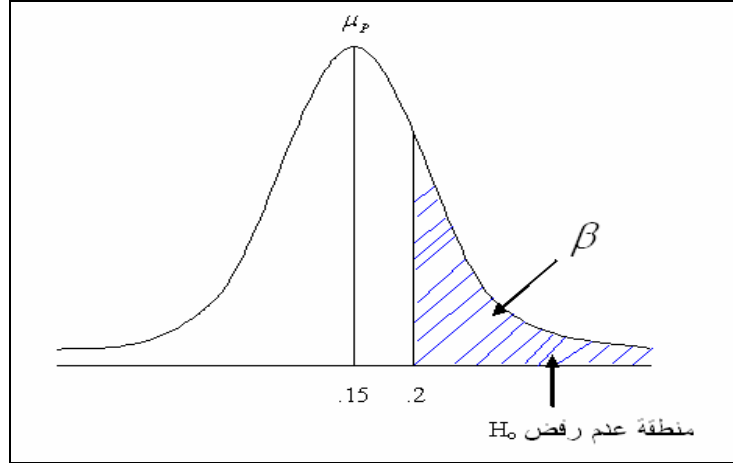
$$\begin{aligned}\alpha &= P_r \left(z \leq \frac{P_s - \pi}{\sigma_{P_s}} \right) \\ &= P_r \left(z \leq \frac{0.20 - 0.25}{0.043} \right) \\ &= P_r (z \leq -1.6) \\ &= 0.123 \text{ or } 12.3 \%\end{aligned}$$

السؤال الذي يبرز لدينا هنا ، هو لماذا تم التقريب الى التوزيع الطبيعي؟ الجواب على هذا السؤال ، كي تتمكن من استخدام جداول هذا التوزيع. السؤال الآخر الذي يبرز ايضا ، هو ما هي فائدة استخدام هذا الجدول؟ ان استخدام الجدول يساعد في اتخاذ القرار الخاص بفرضية العدم.

ب. بافتراض ان الحقيقة هي

$$H_A : \pi = 0.15$$

وان (β) تعرف بأنها الاحتمالية الخاصة بعدم رفض فرضية العدم H_0 الخاطئة (شكل 15.4)



شكل 15.4 : توزيع المنحنى الطبيعي بوسط حسابي $\mu_p = 0.15$ يظهر فيه قيمة الخطأ من النوع الثاني β

بعبارة أخرى

$$\beta = P_r (P > 0.20 / \pi = 0.15)$$

مرة أخرى ، وباستخدام نظرية النزعة المركزية والمتضمنة ان توزيع العينة الخاص الى P_s مقارب الى التوزيع الطبيعي لمتوسط هو :

$$\mu_{P_s} = 0.15$$

وتباين هو :

$$\begin{aligned} \sigma_{P_s}^2 &= \frac{(0.15)(0.85)}{100} \\ &= (0.036)^2 \end{aligned}$$

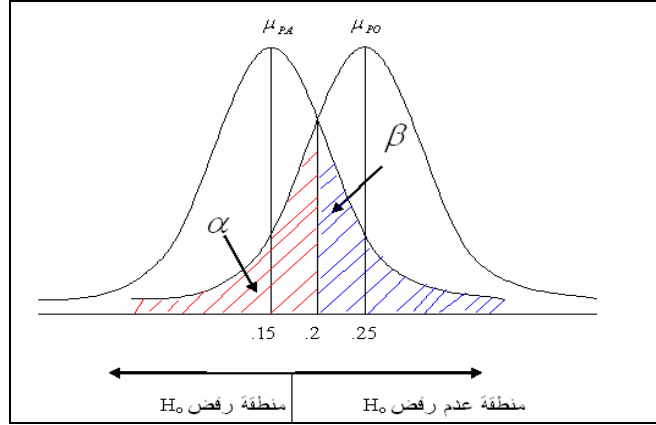
$$\beta = Pr(z \geq \frac{0.20 - 0.15}{0.036})$$

$$= Pr(z \geq 1.39)$$

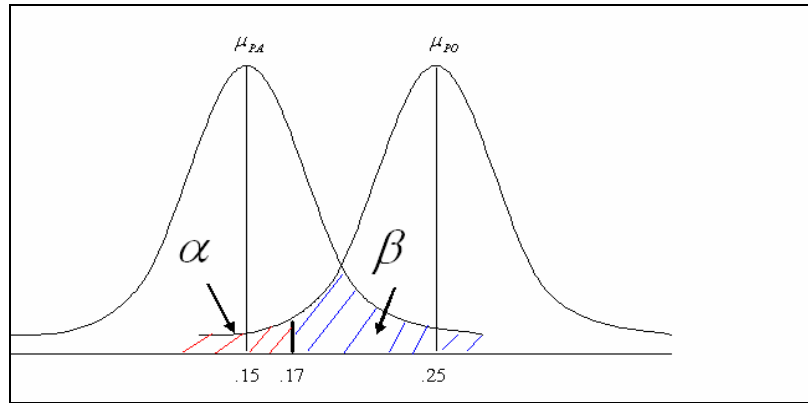
ومن جداول التوزيع الطبيعي نحصل على الآتي:

$$= 0.082 \text{ or } 8.2 \%$$

يمكن تمثيل النتائج التي تم الحصول عليها بيانيا (شكل 16.4) حيث يلاحظ ان β تعتمد على الخاصية البديلة. وفي حالة تغيير قاعدة صنع القرار باستخدام نقطة قطع اصغر (بدلا من نقطة القطع 0.20) ومقدارها 0.17 ، سيؤدي ذلك الى تقليص حجم α (قيمة α) ولكن بالمقابل سيزداد حجم β (قيمة β) (شكل 17.4).



شكل 16.4 : توزيع المنحنيين الطبيعيين بوسطين حسابيين 0.25 و 0.15 معا ، يظهر فيه قيمة الخطأ من النوع الأول (α) والخطأ من النوع الثاني (β) ونقطة الفصل 0.2



شكل 17.4 : توزيع المنحنيين الطبيعيين بوسطين حسابيين 0.25 و 0.15 معا ، يظهر فيه قيمة الخطأ من النوع الأول (α) والخطأ من النوع الثاني (β) ونقطة الفصل 0.17

يلاحظ من الشكل (17.4) انه تم تقليص قيمة نقطة الفصل (من 0.20 إلى 0.17) مما أدى ذلك إلى تقليص منطقة الخطأ من النوع الأول (α) وبالمقابل أدى ذلك إلى زيادة منطقة الخطأ من النوع الثاني (β).

فيما يلي الخطوات المهمة لتحديد آلية إجراء اختبار الفرضية:

1. تحديد صيغة فرضية (العدم) بسيطة.
2. تحديد صيغة الفرضية البديلة ، بحيث انه في حالة رفض فرضية العدم تقبل الفرضية البديلة.
3. تحديد احتمالية الخطأ من النوع الأول.
4. تحديد الأسلوب الإحصائي الملائم لغرض بناء معيار الاختبار لاختبار فرضية العدم التي حددت صيغتها في الفقرة (1) ضد الفرضية البديلة التي حددت صيغتها في الفقرة (2) أعلاه.
5. اتخاذ القرار الخاص بفرضية العدم، رفضها وقبول الفرضية البديلة ، او العكس

هناك أنواع عديدة من اختبارات الفرضيات منها :

1. اختبار الفرضية حول المتوسط.
 2. اختبار الفرضية حول النسبة.
 3. اختبار الفرضية حول التباين.
 4. اختبار الفرضية حول معامل ارتباط بيرسون.
 5. اختبار الفرضية حول الاختلاف بين متوسطين اثنتين.
 6. اختبار الفرضية حول الاختلاف بين نسبتي اثنتين.
- لغرض التعريف أكثر بالمفاهيم الخاصة باختبار الفرضيات سيتم ذلك من خلال المثالين التاليين الذين يجسدان ويمثلان التشابه مع المفاهيم الخاصة بهذا الموضوع.
- مثال (31.4) :** هناك تماثل بين الإجراءات المتخذة من قبل هيئة المحلفين في المحكمة والقرارات المتخذة استنادا الى الاختبار الإحصائي المستخدم.
- في قاعدة المحكمة ، فان مهمة هيئة المحلفين هو تقييم الإجراءات القانونية المتخذة من قبل المدعي العام والدفاع لتحديد فيما اذا كان المتهم بريء أم مذنب. حيث ان قرار هيئة المحلفين قد يكون صحيحا" او خطأ. هنالك نوعين من الخطأ يمكن ان تقع فيه هيئة المحلفين ، هما:
- إثبات التهمة على إنسان بريء (α) او إطلاق سراح المجرم (β) . الجدول (14.4) يوضح التشابه بين الاختبار الإحصائي وقرار هيئة المحلفين:

جدول 14.4: المقارنة بين الاختبارات الإحصائية وقرارات المحكمة

ت	اختبار المعنوية	قاعدة المحكمة
1	فرضية العدم	كل متهم بريء حتى تثبت إدانته
2	تصميم البحث	تحريات الشرطة
3	بيانات	الدلائل و الأدلة
4	القرارات الإحصائية	قرار المحلفين
5	الخطأ من النوع الأول (α)	إدانة المتهم وهو بريء
6	الخطأ من النوع الثاني (β)	تبرئة المتهم وهو غير بريء

توضح هذه المقارنة فكرة مهمة : عندما لا ترفض فرضية العدم ، فانه ليس من الضروري قبولها، وذلك لان نص القرار " غير مذنب " هو مؤشر إلى وجود نقص الأدلة لإثبات بان المتهم هو مذنب وقرار بان المتهم "بريء" هو إحدى هذه الاحتمالات . وهكذا عندما تكون الاختلافات غير معنوية إحصائيا (غير معنوية إحصائيا بمعنى ان الفروق اختبرت من خلال

الاختبارات الإحصائية وكانت نتيجة هذه الاختبارات غير معنوية بعبارة أخرى أن الفروق ليست كبيرة (فهناك احتمالين هما :

1. الاحتمال الأول ، ان فرضية العدم صحيحة .
2. الاحتمال الثاني ، ان فرضية العدم خاطئة ، لكن لا يوجد أدلة كافية من بيانات العينة - المسحوبة - يدعم قرار الرفض (رفض فرضية العدم) (بعبارة أخرى ، ان حجم العينة صغير جدا).

مثال (32.4): (اختبارات التشخيص الطبية)

مقارنة أخرى لمصطلح الفرضية يمكن إجراءه بتطبيق الإجراءات التشخيصية . استنادا الى المشاهدات السريرية او الأساليب المختبرية ، يصنف الأشخاص الى سليم او مريض . هذه الاختبارات قد لا تكون دقيقة او تامة ، فقد يصنف الشخص سليم خطأ بأنه مريض ، في حين ان شخص آخر مريض لا يمكن الكشف عن مرضه . والجدول (15.4) يوضح المقارنة بين الاختبارات الإحصائية والاختبارات التشخيصية:

جدول 15.4: مقارنة بين الاختبارات الإحصائية والاختبارات التشخيصية الطبية.

ت	اختبار المعنوية	إجراء المستشفى
1	فرضية العدم	كل شخص سليم حتى يثبت مرضه
2	تصميم البحث	الاختبارات الطبية المتبعة (الفحوص)
3	البيانات	المواد المسحوبة من الشخص لغرض إجراء الاختبار عليها (على سبيل المثال سحب الدم)
4	القرارات الإحصائية	تشخيص الطبيب (الأطباء)
5	الخطأ من النوع الأول	خطأ موجب، بعبارة أخرى يشخص بأنه مريض وهو سليم
6	الخطأ من النوع الثاني	خطأ سالب، بعبارة أخرى يشخص بأنه سليم وهو مريض

من المتوقع أن تشخص الاختبارات الطبية بصورة صحيحة أي مرض ، وهذا من التوقعات غير الدقيقة . كذلك من الأفكار الأخرى الخاطئة ، هو افتراض أن جميع الاختبارات ، بغض النظر عن المرض ، تكون متساوية الدقة.

مفاهيم أساسية

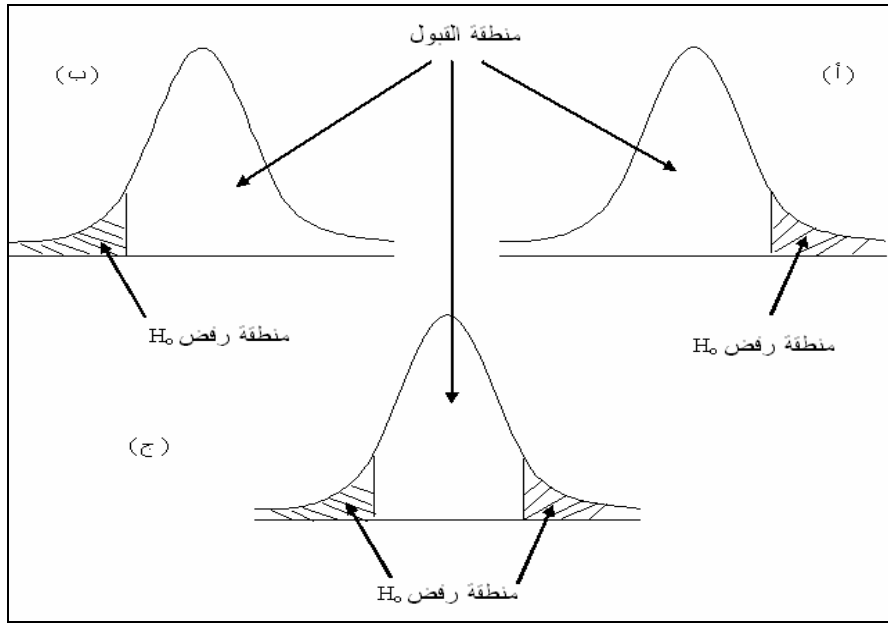
1.منطقة الرفض

تقسم جميع المشاهدات (بيانات العينة) الى منطقتين . فالمنطقة التي تحتوي على القيم التي فيها الاختبار الإحصائي يعطي مؤشرا الى رفض فرضية العدم ، تدعى بمنطقة الرفض . على سبيل المثال ، اذا كانت فرضية العدم كانت حول μ وكالاتي :

$$H_0: \mu=10$$

ان أفضل موقع يمكن معه إجراء الاختبار الإحصائي لفرضية العدم هو استخدام $\bar{x}$ (حيث ان μ هو متوسط المجتمع وان $\bar{x}$ هو متوسط العينة وان استخدام $\bar{x}$ بدلا من μ وذلك لان

الاختبار الإحصائي يطبق على العينة وليس على المجتمع) ، ومن الملاحظ ان H_0 يجب ان ترفض اذا كانت قيمة $\bar{X}$ بعيدة عن 10 (الفرضية الخاصة بقيمة μ) .
في حين أن المنطقة الثانية تدعى منطقة القبول، وتعني القبول بفرضية العدم (شكل 18.4أ،ب، ج)

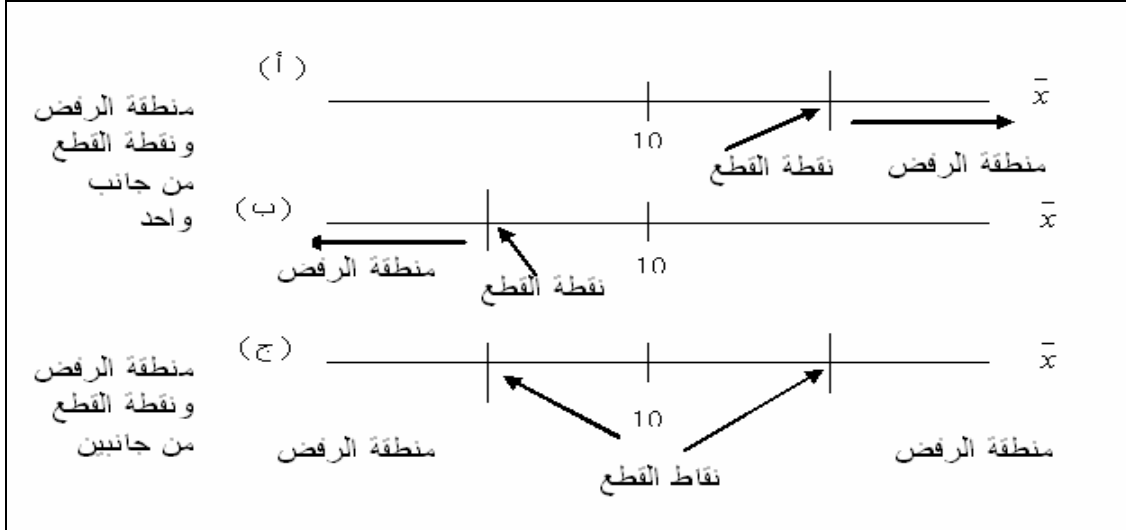


شكل 18.4 : (أ، ب ، ج) توزيع المنحني الطبيعي فيه منطقتي القبول والرفض

2. الاختبار من جانب واحد ضد الاختبار من جانبيين :

مما تقدم ، ينبثق السؤال الأساسي : هل ان الباحث مهتم بالفرق بين $\bar{X}$ عن القيمة 10 باتجاه واحد او باتجاهين؟ فاذا كان الباحث مهتم بتحديد فيما اذا كانت μ تختلف (قيمة μ) اختلافا معنوية (تحدد من خلال الاختبار الإحصائي) عن القيمة 10 في هذه الحالة يفضل استخدام اختبار الاتجاهين ، ذلك أن الباحث لم يحدد نوع الاتجاه (أكبر أم أصغر) (شكل 18.4-ج وشكل 19.4-ج).

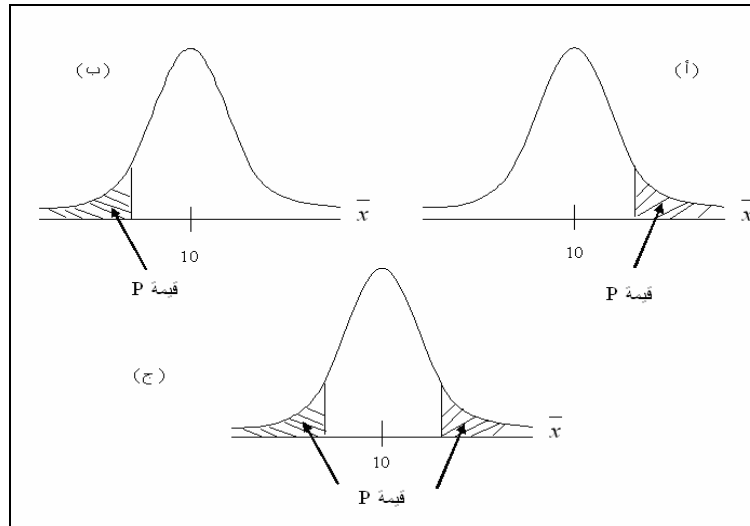
بالمقابل اذا كان الباحث مهتم فيما اذا كانت قيمة μ معنويا اكبر من القيمة 10 ، فيستخدم في هذه الحالة الاختبار من جانب واحد (شكل 18.4-أ وشكل 19.4-أ) وبالعكس اذا كان الباحث مهتم اذا كانت قيمة μ معنويا اصغر من القيمة 10 ، فيستخدم ايضا في هذه الحالة الاختبار من جانب واحد (شكل 18.4-ب وشكل 19.4-ب).



شكل 19.4 : (أ،ب،ج) مناطق الرفض ونقاط القطع

قيم P (P values)

هي عبارة عن مقياس إحصائي يمثل معنوية الحدث استنادا الى التوزيع المحدد. ويحتسب من خلال الاحتمالية التي تدعم الحدث الخاص بفرضية العدم ، وكلما كانت قيمة P صغيرة ، كانت هنالك عدم رغبة اكبر في ان تكون الحادثة (على سبيل المثال $H_0 : \mu = 10$) نتيجة للفرضية العشوائية (فرضية العدم) وبالتالي تكون ذات معنوية إحصائية أكثر.



شكل 20.4 : (أ،ب،ج) توزيع المنحني الطبيعي فيه قيم P

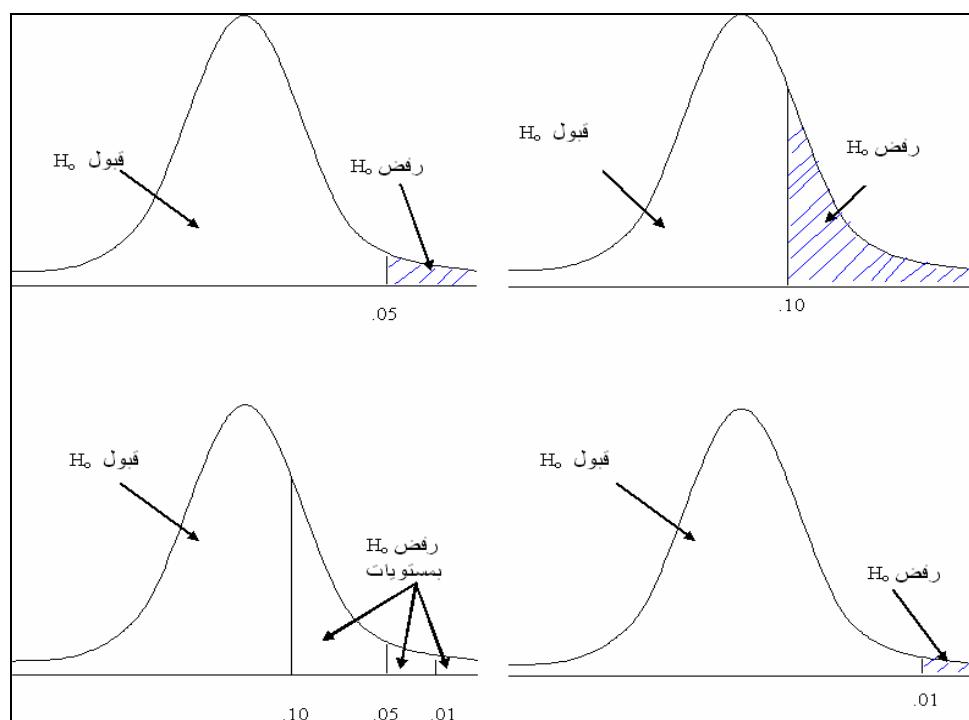
ان قيمة P هي الناتج الرئيس لاختبار المعنوية (الأهمية) الإحصائية ضد فرضية العدم. بعبارة أخرى ان الناتج (الاختبار الإحصائي) يكون معنويا اذا كانت قيمة P صغيرة جدا (قريب من الصفر).

اذا كان الاختبار من جانب واحد ، فان قيمة P ستكون موضحة في الشكلين (20.4-أ) و (20.4-ب) و اذا كان الاختبار من جانبيين ، P موضحة في الشكل (20.4-ج). يمثل المنحني الذي يظهر في الإشكال السابقة توزيع العينة (يمثل التوزيع الطبيعي والتوزيع الطبيعي له معلمتين هما الوسط الحسابي والتباين). واستخدام معيار قيمة P يكون كالآتي :

1. اذا كانت $P < \alpha$ ترفض فرضية العدم H_0 .
 2. اذا كانت $P \geq \alpha$ عدم رفض فرضية العدم H_0 .
- حيث تدرج قيم P كجزء من النتائج وهي تعطي مؤشرات علمية للباحثين أكثر من الحالات الأخرى مثل " القول ان : ترفض فرضية العدم عند مستوى معنوية 0.05 " او " القول ان النتيجة غير معنوية عند مستوى 0.05 ". فعندما تترافق قيمة P مع الاختبار تمكن الباحث معرفة مدى شيوع او ندرة القيمة المحسوبة للاختبار الإحصائي المستخدم في حالة كون H_0 صحيحة. بعبارة أخرى ، يمكن استخدام قيمة P كمقياس للمطابقة بين البيانات (الحقيقة) وفرضية العدم (النظري) ، فكلما كانت قيمة P صغيرة ، كلما كانت المطابقة بينهما اقل (جدول 16.4) ، (انظر شكل 21.4)

جدول (16.4) : تفسير قيم P المختلفة.

قيم P	التفسير
$P > 0.10$	النتيجة (H_0) غير معنوية - عدم رفض فرضية العدم.
$0.05 < P < 0.10$	النتيجة (H_0) ذات معنوية هامشية.
$0.01 < P < 0.05$	النتيجة (H_0) ذات معنوية.
$P < 0.01$	النتيجة هي ذات معنوية عالية / ترفض فرضية العدم.



شكل 21.4 : توزيع المنحنى الطبيعي وفيه قيم مختلفة لـ P

مثال (33.4): بافتراض ان نسبة التدخين العالمي للرجال هو 25% ، ويرغب الباحث في دراسة نسبة التدخين للرجال في الوطن العربي.
فان فرضية العدم تحت البحث هي :

$$H_0 : \pi = 0.25$$

سحبت عينة بحجم $n=100$ من الرجال ، وجد ان عدد $x=15$ منهم مدخنين. هل ان نسبة التدخين π في الوطن العربي تختلف عن تلك الخاصة بنسبة التدخين العالمي؟
طالما ان حجم العينة $n=100$ كبير . لذا يمكن تطبيق نظرية النزعة المركزية ، حيث انها تؤشر ان توزيع المعاينة الخاص بنسبة العينة P_s يقارب الى التوزيع الطبيعي بمتوسط وتباين تحت فرضية العدم هما :

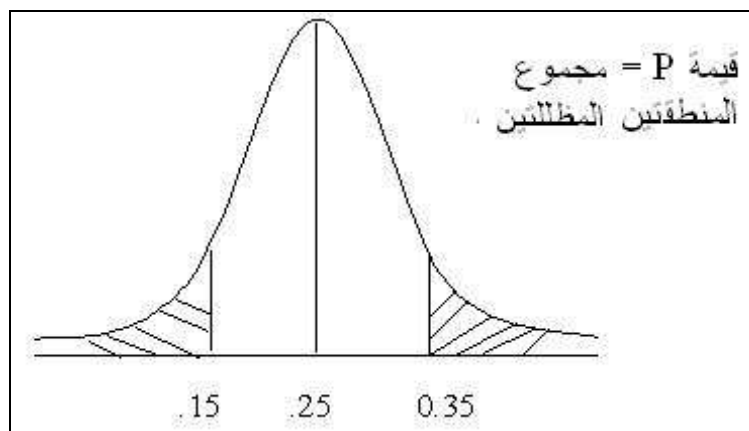
$$\mu_{P_s} = 0.25$$

$$\begin{aligned}\sigma_{P_s} &= \frac{p(1-p)}{n} \\ &= \frac{0.25(1-0.25)}{100} \\ &= (0.043)^2\end{aligned}$$

ان قيمة (النسبة) prob. المشاهدة من العينة المسحوبة هي :

$$\frac{15}{100} = 0.15$$

وهي تمثل فرق مقداره 0.10 (0.25-0.15 = 0.10) عن القيمة الافتراضية (القيمة التي حددت تحت فرضية العدم و مقدارها 0.25). انظر شكل (22.4)



شكل 22.4 : منحنى طبيعي فيه قيمة P

وهكذا ، فإن:

$$\begin{aligned}
 P_{\text{value}} &= \Pr (P \leq 0.15 \text{ or } P \geq 0.35) \\
 &= 2 \times \Pr (P \geq 0.35) \\
 &= 2 \times \Pr (z \geq \frac{0.35 - 0.25}{0.043}) \\
 &= 2 \times \Pr (z \geq 2.33) \\
 &= 2 \times 0.0099 \\
 &= 0.0198 \\
 &\approx 0.02
 \end{aligned}$$

ومن البيانات المعطاة ، فإن الاختلاف بين نسبة التدخين العالمي للرجال ونسبة التدخين في الوطن العربي للرجال هي ذات فروق إحصائية معنوية ($P < 0.05$) بعبارة أخرى ، ترفض فرضية العدم (التي تنص على عدم وجود فروق معنوية) (التي تنص على أن النسبتين متماثلتين) وتقبل الفرضية البديلة (التي تنص على وجود فروق بين النسبتين) . حيث تم اتخاذ قرار الرفض (أو أن الفروق معنوية) استنادا إلى جدول (14.4) .

سلاسل ماركوف Markov chains

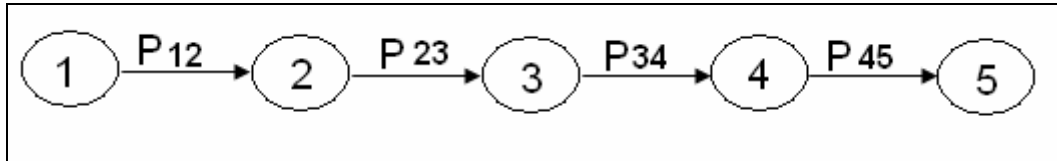
إن أغلب الدراسات التي تتعلق بموضوع الاحتمالية تتعامل مع العمليات التجريبية المستقلة. فإذا كانت سلسلة من التجارب تشكل عمليات تجريبية مستقلة ، فإن النواتج المحتملة لكل تجربة من هذه التجارب لها الفرصة (الاحتمالية) نفسها في الظهور (الحدوث) . في عام 1907 بدأ A. A. Markov دراسته حول أهمية ظهور نوع معين من العمليات ، يؤثر ناتج عمليات سابقة في ناتج العمليات اللاحقة. هذا النوع من العمليات يدعى بسلاسل ماركوف.

سلسلة ماركوف

تعرف سلسلة ماركوف بأنها عبارة عن عملية عشوائية لها خاصية معينة ، وهي إن ناتج العملية التالية (اللاحقة) يعتمد على ناتج العملية الحالية فقط. بعبارة أخرى إن سلسلة ماركوف هي عبارة عن سلسلة من المتغيرات العشوائية $X_1, X_2, X_3, \dots$ والتي لها خاصية ماركوف ، التي هي اعتماد الحالة (ناتج التجربة) التالية (التي يمكن أن تحدث في المستقبل) على الحالة الحالية (التي تحدث الآن) ويمكن أن تكتب بصيغة الرموز كالآتي:

$$P_r (X_{n+1} = N / X_1, X_2 = N_2, \dots, X_n = N_n) = Pr (X_{n+1} = N / X_n = N_n)$$

تشكل القيم المحتملة لـ X_i ($i = 1, 2, 3, \dots, n$) مجموعة قابلة للعد يرمز لها S ويطلق عليها مجال (فضاء) الحالات الخاصة بالسلسلة . بعبارة أخرى يمكن وصف نموذج ماركوف (سلسلة ماركوف) بأنها عبارة عن سلسلة من الأحداث التي يمكن أن تحدث بصورة متتابعة على شكل سلسلة. كل حدث منها يحدد احتمالية الحدث التالي (شكل 23.4). يمكن أن ينظر الى سلسلة ماركوف بأنها عبارة عن عملية انتقال باتجاه واحد من حالة الى الحالة التالية باحتمالية معينة ، والتي تعرف بالاحتمالات الانتقالية مثال على نموذج ماركوف هو تغير إشارة المرور الضوئية ، حيث إن الإشارة الحالية تعتمد على الإشارة السابقة (إن فتح إشارة المرور الخضراء يكون بعد انتهاء إشارة المرور الحمراء ، الذي يفتح بعد انتهاء إشارة المرور الصفراء).



شكل 23.4: سلسلة ماركوف ، تحتوي على سلسلة خطية من الأحداث أو الحالات ، تمثل الأرقام (داخل الدوائر) بالأحداث أو الحالات ، هذه الدوائر موصولة (باسهم) تمثل قيم الاحتمالات الانتقالية.

يمكن تمثيل التواليات البايولوجية (التي يعبر عن هذه التواليات بالحروف) بواسطة سلاسل ماركوف ، حيث يمثل كل حرف حالة وترتبط (هذه الحالات) مع بعضها من خلال قيم الاحتمالات الانتقالية.

هنالك العديد من نماذج سلاسل ماركوف تستخدم (هذه الأنواع) لوصف مجاميع البيانات المختلفة. في كل نوع من نماذج ماركوف ، تحدد حلول رياضية لها. حيث يصف نموذج ماركوف ذو الترتيب - صفر الحالة الحالية باحتمالية مستقلة عن الحالة السابقة ، يطبق هذا

النموذج على التوالي العشوائي الذي فيه كل ثمالة يمكن أن تحدث بتكرار متساوي. في حين يصف نموذج ماركوف ذو الترتيب - الأول الحالة الحالية باحتمالية تعتمد على الحالة السابقة ، وبصورة مماثلة ، يصف نموذج ماركوف ذو الترتيب - الثاني الحالة الحالية باحتمالية تعتمد على الحالتين السابقتين.

على سبيل المثال ، في حالة التشفير لبروتين ، فإن تكرار الترتيب الوحيدة (حدوث ثلاثة حالات مترابطة في آن واحد كما في حالة الشفرة) يجب أن تكون مختلفة عن التوالي العشوائية أو غير المشفرة. هذه العملية (حدوث تكرار في الترتيب أو حدوث عدد من الحالات المترابطة آنيا) يمكن أن توصف من خلال نموذج ماركوف ذو الرتبة - الثانية.

مثال (34.4): في إحدى الدراسات عن حالة الطقس في إحدى الجزر. حيث إن حالة الطقس فيها لا تتوالى فيه الأيام المعتدلة (بدون مطر أو ثلج) ، وفي حالة امتلاكها يوم معتدل واحد فإن احتمالية أن يمر عليها في اليوم التالي يوم ممطر أو يوم مثلج متساوي. وفي حالة ما إذا مر على الجزيرة يوم ممطر أو مثلج فإن هناك احتمالية متساوية إن اليوم التالي يكون ممثلاً ممطر أو مثلج. في حالة وجود تغيير من يوم ممطر أو مثلج إلى يوم معتدل له احتمالية النصف. يمكن تحديد مصفوفة الاحتمالات الانتقالية كالآتي:

		اليوم الحالي		
		R	N	S
P =	R	1/2	1/4	1/4
	N	1/2	0	1/2
	S	1/4	1/4	1/2

حيث إن :

R: يوم ممطر

N: يوم معتدل

S: يوم مثلج

تمثل الصفوف اليوم الحالي فيما تمثل الأعمدة اليوم التالي. كما يلاحظ من المصفوفة انه قد عبر عن العبارة " إن هذه الجزيرة لا تمتلك يومين معتدلين متتاليين " بالاحتمالية صفر في الصف الثاني العمود الثاني. وعبر عن العبارة " وفي حالة امتلاكها يوم معتدل واحد فإن احتمالية أن يمر عليها في اليوم التالي يوم ممطر أو يوم مثلج متساوي " بالاحتمالية 1/2 لكلا النوعين R و S في الصف الثاني. وعبر عن العبارة " وفي حالة ما إذا مر على الجزيرة يوم ممطر أو مثلج فإن هناك احتمالية متساوية إن اليوم التالي يكون ممثلاً ممطر أو مثلجاً " بالقيمة 1/2 في كل من

الصف الأول العمود الأول ، وكذلك في الصف الثالث العمود الثالث. وقد عبر عن العبارة " في حالة وجود تغيير من يوم ممطر أو مثلج إلى يوم معتدل له احتمالية النصف بالقيمة $1/2$ لكل منهما في الصف الثاني العمودين الأول والثالث على التوالي ، كذلك يلاحظ إن مجموع كل صف من صفوف المصفوفة P يساوي 1 ، وهو مجموع كل الاحتمالات الممكنة.

المصفوفة الانتقالية Transition matrix

تشير مدخلات الصف الأول من المصفوفة P من المثال (34.4) الى الاحتمالات للأنواع المختلفة من الطقس لليوم الحالي يتبعه يوم ممطر.

وبصورة مماثلة فإن كل من الصفين الثاني والثالث يشيران الى الاحتمالات المختلفة للأنواع المختلفة من الطقس لليوم الحالي يتبعه يوم معتدل ويوم مثلج على التوالي . تدعى هذه المصفوفة بمصفوفة الاحتمالات الانتقالية أو بالمصفوفة الانتقالية.

أما في حالة السؤال عن إمكانية تحديد احتمالية السلسلة في اليوم الحالي i ، والتي ستكون في الحالة j بعد يومين من الآن ، يرمز لها بالرمز $P_{ij}^{(2)}$. في المثال (34.4) ، يلاحظ أنها إذا كانت ممطرة اليوم فإن الحادثة إنها مثلجة ليومين من الآن هي إحدى الحالات الثلاثة الآتية :

1. إنها ممطرة غدا" ومثلجة بعد يومين من الآن.

2. إنها معتدلة غدا" ومثلجة بعد يومين من الآن.

3. إنها مثلجة غدا" ومثلجة بعد يومين من الآن.

إن احتمالية احتساب الحالة الأولى يكون باحتساب حاصل (ناتج) ضرب الاحتمال الشرطي الذي هو احتمالية (الاحتمال الشرطي) إن غدا ممطرة علما إن اليوم ممطرة و (في) الاحتمال الشرطي الذي هو إن بعد يومين من الآن مثلجة علما إنها ممطرة غدا. وباستخدام قيم المصفوفة P ، يمكن كتابة حاصل الضرب (حاصل ضرب الاحتمالين الشرطيين) كالآتي:

$$P_{11} P_{13}$$

حيث إن:

P_{11} يرمز الى الاحتمال الشرطي إن غدا ممطرة علما إن اليوم ممطرة

P_{13} يرمز الى الاحتمال الشرطي إن بعد يومين مثلجة علما إن غدا ممطرة علما إن الرقم 1 يمثل الجو مطر وان 2 تمثل الجو معتدل وان 3 تمثل الجو المثلج.

يمكن احتساب الحالتين (2 و 3) بالأسلوب نفسه من المصفوفة P .

وهكذا فإن احتمالية إن تكون مثلجة بعد يومين من الآن (اليوم الحالي ممطر) هو:

$$P_{13}^{(2)} = P_{11} P_{13} + P_{12} P_{23} + P_{13} P_{33}$$

بصورة عامة ، إذا كانت سلسلة ماركوف مكونة من r من الحالات فإن الصيغة العامة لها تكتب كالآتي:

$$P_{ij}^{(2)} = \sum_{k=1}^r P_{ik} P_{kj}$$

نظرية : بافتراض أن P هي مصفوفة انتقالية من سلسلة ماركوف. إن المدخل $P_{ij}^{(n)}$ للمصفوفة $P^{(n)}$ يحدد (يعطي) احتمالية أن سلسلة ماركوف ، تبدأ في الحالة S_i وستكون في الحالة S_j بعد n من الخطوات.

مثال (35.4) : (استمرار للمثال 34.4) يمكن دراسة حالة الطقس في الجزيرة لفترات زمنية مختلفة ، على سبيل المثال احتساب القوى المتعاقبة للمصفوفة P 1 و 2 وكالاتي:

$$P^1 = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{bmatrix} .500 & .250 & .250 \\ .500 & .000 & .500 \\ .250 & .250 & .500 \end{bmatrix} \end{matrix}$$

$$P^2 = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{bmatrix} .438 & .188 & .375 \\ .375 & .250 & .375 \\ .375 & .188 & .438 \end{bmatrix} \end{matrix}$$

وعلى سبيل المثال يمكن حساب كل من $P_{13}^{(2)}$ و $P_{11}^{(2)}$ وكالاتي:

$$\begin{aligned}
 P_{13}^{(2)} &= P_{11} P_{13} + P_{12} P_{23} + P_{13} P_{33} \\
 &= (.500) (.250) + (.250) (.500) + (.250) (.500) \\
 &= .375
 \end{aligned}$$

$$\begin{aligned}
 P_{11}^{(2)} &= P_{12} P_{21} + P_{13} P_{31} + P_{11} P_{11} \\
 &= (.250) (.500) + (.250) (.250) + (.500) (.500) \\
 &= .4375 \approx .438
 \end{aligned}$$

وبالطريقة نفسها تحتسب المصفوفات $P^{(3)}$ ، $P^{(4)}$ ، $P^{(5)}$ و $P^{(6)}$ كالآتي:

$$\begin{aligned}
 P^4 &= \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{bmatrix} .402 & .199 & .398 \\ .398 & .203 & .398 \\ .398 & .199 & .402 \end{bmatrix} \end{matrix} \quad P^5 = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{bmatrix} .400 & .200 & .399 \\ .400 & .199 & .400 \\ .399 & .200 & .400 \end{bmatrix} \\ \\ P^6 &= \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{bmatrix} .400 & .200 & .400 \\ .400 & .200 & .400 \\ .400 & .200 & .400 \end{bmatrix} \end{matrix}
 \end{aligned}$$

يلاحظ إن الاحتمالات للأنواع الثلاثة من الطقس R و N و S هي: 0.4 و 0.2 و 0.4 على التوالي في المصفوفة الأخيرة (P^6). هذا النوع من الأمثلة يخص نوع معين من سلاسل ماركوف تسمى بسلسلة ماركوف المنتظمة Regular Markov chain.

مثال (36.4) : (نموذج الجين) إن أبسط نوع من أنواع تجارب الصفات الوراثية لدى الحيوانات تحدث عندما يتم دراسة زوج من الجينات. وكل منهما (الجينين) يحتوي على صورتين ، على سبيل المثال G و g ، وبهذا ستكون توافيق الأنواع المختلفة هي : GG و Gg (وهي جينيا مماثلة الى gG) أو gg.

عند تزاوج حيوانين ، فإن الذرية الناتجة عنهما سيرثان جين واحد من كل من الأبوين ، وإن الفرضية الأساسية هو إن اختيار هذه الجينات يتم بصورة عشوائية ومستقلة الواحد عن الآخر. وبهذه الفرضيات يتم تحديد احتمالية حدوث كل نوع من الجينات في الذرية. وبافتراض إن الاحتمالات ستكون 1/4 للـ GG و 1/2 للـ Gg و 1/4 للـ gg.

وبافتراض إن عملية التزاوج مستمرة بين الحيوانين ، وإن هنالك على الأقل نسل (ذرية) واحد. يتم اختبار النوع الجيني للنسل الى الجيل المتعاقب (التالي) والتي يمكن التعبير عنها بواسطة سلسلة ماركوف.

ويشار الى الأنواع (الحالات) السائدة Dominant والهجين Hybrid والمتحية Recessive بالرموز GG و Gg و gg على التوالي. وان مصفوفة الاحتمالات الانتقالية هي :

$$P = \begin{matrix} & \begin{matrix} GG & Gg & gg \end{matrix} \\ \begin{matrix} GG \\ Gg \\ gg \end{matrix} & \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix} \end{matrix}$$

يلاحظ انه إذا كان الوالدان يمتلكان جينين من النوع السائد GG ستحصل ذريتهما على النوع من الجين نفسه باحتمال 0.5 وعلى النوع الهجين باحتمال 0.5 أيضا. أما إذا امتلك الوالدان جينين هجينين Gg ستحصل ذريتهما على الجينين (النقي) السائدين GG باحتمال 0.25 وعلى النوع الهجين باحتمال 0.5 وعلى النوع المتحي (النقي) باحتمال 0.25 . في حين إذا كان الوالدان يمتلكان جينين من النوع المتحي gg (النقي) في هذه الحالة ستحصل ذريتهما على النوع الهجين باحتمال 0.5 وعلى النوع المتحي gg باحتمال 0.5 .

مثال (37.4) : (تطوير مثال 36.4) بافتراض إن نوع الجين السائد هو الغالب في التجربة ، فان مصفوفة الاحتمالات الانتقالية هي :

$$P = \begin{matrix} & \begin{matrix} GG & Gg & gg \end{matrix} \\ \begin{matrix} GG \\ Gg \\ gg \end{matrix} & \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

يلاحظ إن الجيل الجديد من المواليد لا يمتلك الجين المتحي (النقي) بغض النظر عن نوع الجينات التي يمتلكها الوالدان.

مصفوفات درجات المواقع المحددة (PSSM) Position specific scoring matrices

تعرف هذه المصفوفات بأنها عبارة عن جدول يحتوي على معلومات باحتمالات معينة عن الأحماض الأمينية أو النيوكليوتيدات عند كل موقع من مواقع الاصطفافات المتعددة بدون فجوات للتواليات المختلفة ، يتألف هذا الجدول من صفوف وأعمدة . تمثل أعمدة مواقع الأحماض

الأمينية للاصطفاف المتعدد للتواليات ، في حين تمثل الصفوف أسماء الأحماض الامينية الجدول (17.4).

مثال (38.4) : يمثل الجدول (17.4) اصطفاف 5 تواليات بدون فجوات ولكل توالي منها يحتوي على 6 مواقع وكالاتي:

جدول 17.4 : اصطفاف متعدد بدون فجوات

		المواقع (j)					
الخطوات (i)		1	2	3	4	5	6
	1	A	T	G	T	C	G
	2	A	A	G	A	C	T
	3	T	A	C	T	C	A
	4	C	G	G	A	G	G
	5	A	A	C	C	T	G

لبناء مصفوفة درجات المواقع المحددة يتم بالخطوات الآتية :

- الخطوة الأولى : يتم احتساب التكرارات الأولية لكل ثمالة عند كل عامود (المواقع من 1 الى 6) على انفراد (جدول 17.4) بالمعادلة الآتية :

$$f_{ij} = \frac{\text{No. of occurrence of Residues}}{\text{Total No. of Residues}}$$

حيث إن :

f_{ij} تكرار i و j

i : يمثل رقم الصف وهو $i = 1, 2, \dots, n$ (في المثال $i = 1, 2, \dots, 5$)

j : يمثل رقم العمود وهو $j = 1, 2, \dots, m$ (في المثال $j = 1, 2, \dots, 6$)

في الموقع الأول (يحسب بالنسبة للعمود الأول)

حيث توجد ثلاث ثمالات من النوع A من الثمالات الخمسة الموجودة في الموقع الأول (العمود الأول). في حين توجد ثمالة واحدة من النوعين T و C ولا توجد أية ثمالة من النوع G في هذا الموقع.

$$P_A = \frac{3}{5} = 0.6$$

$$P_T = \frac{1}{5} = 0.2$$

$$P_C = \frac{1}{5} = 0.2$$

$$P_G = 0$$

في الموقع الثاني

$$P_A = \frac{3}{5} = 0.6$$

$$P_T = \frac{1}{5} = 0.2$$

$$P_C = 0$$

$$P_G = \frac{1}{5} = 0.2$$

في الموقع الثالث

$$P_A = 0$$

$$P_T = 0$$

$$P_C = \frac{2}{5} = 0.4$$

$$P_G = \frac{3}{5} = 0.6$$

في الموقع الرابع

$$P_A = \frac{2}{5} = 0.4$$

$$P_T = \frac{2}{5} = 0.4$$

$$P_C = \frac{1}{5} = 0.2$$

$$P_G = 0$$

في الموقع الخامس

$$P_A = 0$$

$$P_T = \frac{1}{5} = 0.2$$

$$P_C = \frac{3}{5} = 0.6$$

$$P_G = \frac{1}{5} = 0.2$$

في الموقع السادس

$$P_A = \frac{1}{5} = 0.2$$

$$P_T = \frac{1}{5} = 0.2$$

$$P_C = 0$$

$$P_G = \frac{3}{5} = 0.6$$

معدل التكرارات الكلي ($f_{i.}$) لكل ثمالة يستخرج من جمع التكرارات (f_{ij}) المقابلة الى تلك الثمالة (مجموع الصف) ثم يقسم على عدد المواقع (m) (في المثال يقسم على 6 مواقع) . ويستخرج وفق الصيغة الآتية :

$$f_{i.} = \frac{\sum_{j=1}^m f_{ij}}{m}, i = 1, 2, \dots, n$$

معدل التكرار الكلي للثمالة A

$$f_{1.} = \frac{0.6 + 0.6 + 0 + 0.4 + 0 + 0.2}{6} = 0.30$$

معدل التكرار الكلي للثمالة T

$$f_{2.} = \frac{0.2 + 0.2 + 0 + 0.4 + 0.2 + 0.2}{6} = 0.2$$

معدل التكرار الكلي للثمالة G

$$f_{3.} = \frac{0 + 0.2 + 0.6 + 0 + 0.2 + 0.6}{6} = 0.27$$

معدل التكرار الكلي للثمالة C

$$f_{4.} = \frac{0.2 + 0 + 0.4 + 0.2 + 0.6 + 0}{6} = 0.23$$

في هذه المرحلة يتم تحويل الاصطفاف المتعدد بدون إحداث فجوات إلى قيم تكرارية يمكن إدراجها في الجدول (18.4)

جدول (18.4) : قيم تكرار الثمالات حسب المواقع المحددة.

الثمالات	المواقع						
		1	2	3	4	5	6
A	معدل تكرار الثمالات	0.6	0.6	-	0.4	-	0.2
T		0.2	0.2	-	0.4	0.2	0.2
G		-	0.2	0.6	-	0.2	0.6
C		0.2	-	0.4	0.2	0.6	-

2-الخطوة الثانية : تحويل قيم التكرارات في الجدول (18.4) الى قيم قياسية وذلك بقسمة قيمة كل تكرار على معدل التكرار الكلي للثمالة والمدرجة في جدول (19.4) وكالاتي:

$$P_{A1} = \frac{0.6}{0.3} = 2, P_{A2} = \frac{0.6}{0.3} = 2, P_{A3} = 0, P_{A4} = \frac{0.4}{0.3} = 1.33, P_{A6} = \frac{0.2}{0.3} = 0.67$$

$$P_{T1} = \frac{0.2}{0.2} = 1, P_{T2} = \frac{0.2}{0.2} = 1, P_{T4} = \frac{0.4}{0.2} = 2, P_{T5} = \frac{0.2}{0.2} = 1, P_{T6} = \frac{0.2}{0.2} = 1$$

$$P_{G2} = \frac{0.2}{0.27} = 0.74, P_{G3} = \frac{0.6}{0.27} = 2.22, P_{G5} = \frac{0.2}{0.27} = 0.74, P_{G6} = \frac{0.6}{0.27} = 2.222$$

$$P_{C1} = \frac{0.2}{0.23} = 0.87, P_{C3} = \frac{0.4}{0.23} = 1.74, P_{C4} = \frac{0.2}{0.23} = 0.87, P_{C5} = \frac{0.6}{0.23} = 2.6$$

جدول (19.4) مصفوفة قيم التكرار القياسية المحددة للثمالات الأربعة ولست مواقع مختلفة.

التمات	المواقع						متوسط تكرار التمثالة
	1	2	3	4	5	6	
A	2.0	2.0	-	1.33	-	0.67	0.30
T	1.0	1.0	-	2.0	1.0	1.0	0.20
G	-	0.74	2.22	-	0.74	2.22	0.27
C	0.87	-	1.74	0.87	2.61	-	0.23

3-الخطوة الثالثة : استخدام اللوغاريتم للأساس 2 لتحويل قيم التكرار القياسية الى قيم لوغاريتمية ، والمدرجة في الجدول (20.4) وكالاتي:

جدول (20.4) : مصفوفة قيم لوغاريتمية المحددة الى التمثالات الأربعة والى ست مواقع مختلفة.

التمات	المواقع					
	1	2	3	4	5	6
A	1.0	1.0	-	0.41	-	-0.58
T	0	0	-	1.0	0	0
G	-	-0.43	1.15	-	-0.43	1.15
C	-0.2	-	0.8	-0.2	1.38	-

4 - خطوة التنفيذ : يمكن استخدام المصفوفة الأخيرة (جدول 20.4) لغرض استخلاص علاقة التواليات المختلفة ضمن المقارنات المتعددة ومقارنتها مع تسلسل منفرد واحد من خلال قيمة كمية تصف هذه المقارنة .

على سبيل المثال ، في حالة البيانات الموضحة في جدول (18.4) ، السؤال الذي يتم طرحه ، ما مدى (مقدار) ملائمة التوالي AACTCG مع بيانات هذا الجدول؟ لغرض الإجابة على هذا السؤال ، تستخدم القيم الاحتمالية للتوالي الخاصة بكل موقع في جدول البيانات ، ثم تجمع هذه القيم (جدول 21.4) حيث ان ناتج الجمع هو القيمة 6.33 . في هذه الخطوة يتم تحديد مدى ملائمة التوالي AACTCG مع بيانات الجدول (20.4) وكالاتي:

جدول (21.4) : مصفوفة درجات المواقع المحددة تشير الخلايا المظلمة الى تسلسل النيوكليوتيدات

		المواقع					
الذمالات		1	2	3	4	5	6
	A	1.0	1.0	-	0.41	-	-0.58
	T	0.0	0.0	-	1.0	0.0	0.0
	G	-	-0.43	1.15	-	-0.43	1.15
	C	-0.2	-	0.8	-0.2	1.38	-

لتحديد درجة الملائمة تجمع قيم الاحتمالية المناظرة للتوالي من الجدول (21.4) وكالاتي:

$$1.0+1.0+0.8+1.0+1.38+1.15 = 6.33$$

ولكون البيانات المدرجة في الجدول (20.4) قد اخذ لها (تم تحويلها باستخدام) اللوغاريتم للأساس 2 ، فان درجة الملائمة هذه يمكن تفسيرها بأنها القيمة الاحتمالية للتوالي بان يتطابق مع البيانات (في الجدول 20.4) بمقدار 80 مرة أكثر ($2^{6.33} = 80$) من أن تتطابق نتيجة للمصادفة ، ومن ثم فان هذا التوالي يمكن تصنيفه ضمن هذه العائلة من التواليات. تعتمد القيم الاحتمالية الناتجة من مصفوفة درجات المواقع المحددة PSSM على عدد التواليات المستخدمة لغرض المقارنة (عند مقارنة التوالي المفرد مع عدد التواليات).

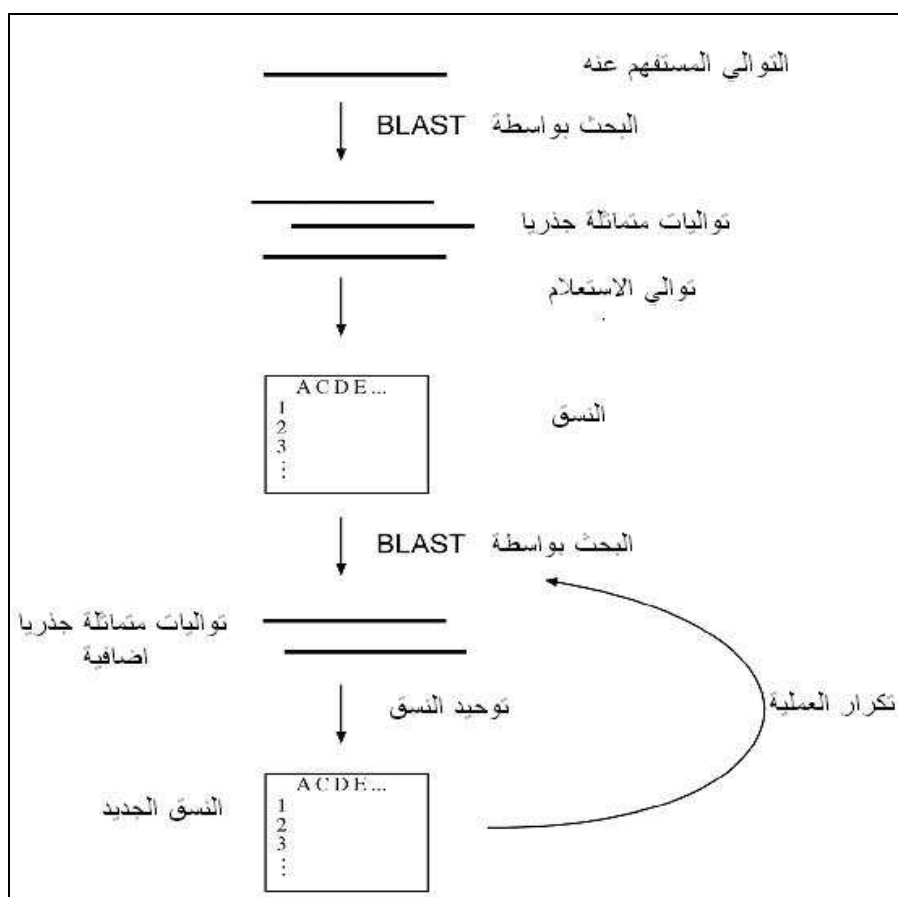
المظاهر الجانبية (النسق) Profiles

تحتوي مقارنة التواليات المتعددة على الفجوات وبأطوال مختلفة . فعندما يتم تضمين المعلومات الخاصة بجزء الفجوة في جدول البيانات الخاص ببيانات التواليات ، يتم إنشاء المظهر الجانبي أو النسق . بعبارة أخرى ، إن النسق هو عبارة عن مصفوفة درجات المواقع المحددة (PSSM) مع (وجود) المعلومات الخاصة بالجزء (المعلومات الخاصة بجزء الفجوة) المتضمنة الإضافات والحذف في التوالي (التواليات).

برنامج PSI-BLAST

يمكن استخدام المظهر الجانبي أو النسق عند البحث في قواعد البيانات لإيجاد التماثل الجذري Homology. إن عملية بناء وتشبيد مصفوفة المظهر الجانبي يدويا يتطلب جهد ووقت وخبرة كبيرة . لهذا السبب يفضل استخدام برنامج لحساب ناتج المظهر الجانبي واستخدامهم للبحث عن

التواليات في قواعد البيانات . مثل هذا البرنامج PSI-BLAST متاح في NCBI يحدد برنامج PSI-BLAST (Position specific iterated BLAST) المظاهر الجانبية للبيانات وينجز أيضا عملية البحث في قواعد البيانات بأسلوب التكرار. في البداية ، يستخدم هذا البرنامج توالي مفرد للاستفسار عنه ، حيث يتم انجاز بحث BLASTP لتحديد مواقع التشابه الأولية. تستخدم الدرجة العالية لبناء اصطفاف التواليات المتعدد ، الذي من خلاله (هذه الاصطفافات) يتكون المظهر الجانبي يستخدم بعد ذلك المظهر الجانبي أو النسق في الجولة (المرحلة) الثانية من البحث لتمييز أعضاء (تواليات) أكثر من العائلة نفسها التي يمكن أن تتطابق مع المظهر الجانبي فعندما يشخص توالي جديد على أنه ضمن مصفوفة المظهر الجانبي (يصنف ضمن العائلة) ، ويتم إضافته (هذا التوالي) إلى الاصطفافات المتعددة السابقة لتوليد مصفوفة مظهر أو نسق جانبي جديدة ، والتي بدورها تستخدم في تكرار البحث في قواعد البيانات. وتستمر هذه العملية لحين الوصول إلى مرحلة عندها لا يوجد توالي يمكن أن يتطابق مع مصفوفة المظهر الجانبي أو النسق.



شكل 24.4 : مخطط عمل برنامج PSI-BLAST استخدام عملية التكرار لتحديد (تشخيص) الشلالات الجذرية

إن الصفة (الميزة) الرئيسية في برنامج PSI-BLAST هو إن مصفوفة المظهر الجانبي أو النسق تنشأ ذاتياً (أوتوماتيكياً) وتعُد (في كل مرة) في كل تكرار متعاقب. مقياس آخر في برنامج PSI-BLAST هو زيادة الحساسية ، من خلال توفير أوزان (ترجيحات) إضافية للثمالات غير الملحوظة (غير المشاهدة) لجعل (مصفوفة) المظهر الجانبي أكثر شمولية. إن الوصول إلى معلومات المظهر الجانبي أو النسق المثلى يجعل برنامج PSI-BLAST حساس جداً عند البحث عن إستراتيجية للكشف عن التواليات ضعيفة التماثل ، لكنها من الناحية البايولوجية ذات تماثل معنوي. على أية حال ، فإن الحساسية العالية لبرنامج PSI-BLAST يمكن أن تؤدي إلى مشكلة أو مأزق وهو انخفاض قدرة هذا البرنامج على تحديد (اختيار) التواليات المتماثلة جذرياً . في حالة الخطأ في إدخال تواليات ليس لها علاقة ، فإن المظهر الجانبي يصبح متحيزاً، هذه العملية تسمح بوجود أخطاء إضافية تضاف إلى الحلقات (العمليات) اللاحقة.

تعرف هذه العملية بانحراف النسق Profile drift وكحل جزئي لهذه المشكلة هو السماح للمستخدم بمراقبة النتائج في كل تكرار ورفض أو استبعاد التواليات المعروفة بعدم وجود علاقة مع تلك الموجودة في قواعد البيانات استناداً إلى المعلومات الإضافية . إضافة إلى ذلك ، يمكن تنفيذ البرنامج لعدد معقول من التكرارات بدلاً من تنفيذه لغاية الوصول إلى التقارب الكامل. إن عدد التكرارات النموذجية لتنفيذ برنامج PSI-BLAST بين ثلاثة إلى خمسة تكرارات وهي كافية لإيجاد التوالي الأكثر تماثلاً جذرياً لتوالي الاستعلام .

سلسلة ماركوف الخفية Hidden Markov chain

في حالة سلسلة ماركوف ، فإن جميع الحالات يمكن وصفها على شكل تتابع خطي (شكل 25.4)، ولكن في بعض الحالات ، فإن عدد من العوامل غير الملاحظة (أو المشاهدة) تؤثر في حسابات (الاحتمالات) الانتقالية للحالة.

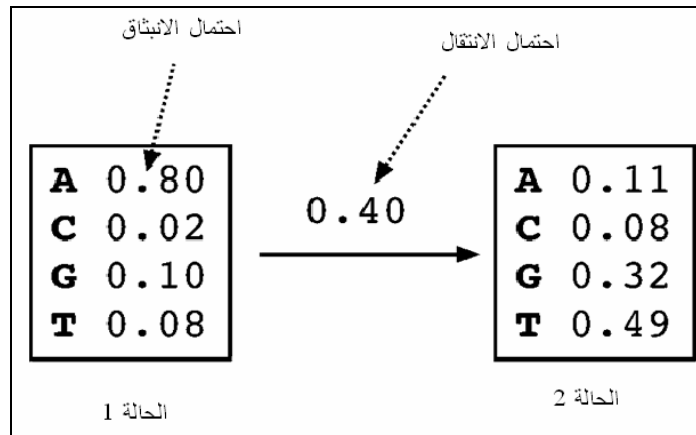
لغرض تضمين مثل هذه العوامل في الحسابات المطلوبة يتطلب ذلك استخدام نماذج أكثر تعقيداً مثل نماذج ماركوف المخفية HMMs . حيث تضم هذه النماذج سلسلتين ماركوف أو أكثر في سلسلة واحدة مؤلفة حالات مشاهدة وأخرى غير مشاهدة (أو مخفية) التي تؤثر في ناتج الحالات المشاهدة . على سبيل المثال ، في الفجوات التي تحدث في عملية الاصطفاف ، هذه الفجوات لا تتطابق مع أية ثمالة وينظر إليها على إنها حالات غير مشاهدة ، وهكذا فإن الفجوات تؤثر في الاحتمالات الانتقالية للحالات المشاهدة .

في نموذج ماركوف الخفي ، كما هو الحال مع نموذج ماركوف العادي ، تنتقل العملية من حالة إلى حالة أخرى باحتمال معين والتي تعرف باسم الاحتمالات الانتقالية . كل حالة من هذه الحالات قد تحتوي على عدد من العناصر والرموز من التواليات النيوكليوتيدية ، فمن المعروف أن هنالك أربعة رموز ممثلة للنيوكليوتيدات هي A و T و G و C في كل حالة . في حين إن التواليات المكونة من الأحماض الأمينية ، تتألف من عشرين رمز لكل حالة . تدعى قيمة الاحتمال الذي يتزامن (يرافق) مع كل رمز في كل حالة باحتمال الانبثاق (الإشعاع ، إطلاق) Emission probability.

لحساب مجموع الاحتمالات لمسار معين من النموذج ، يتطلب ذلك الأخذ بنظر الاعتبار كل احتمال من احتمالات الانتقال والانبثاق والتي تربط الحالات المخفية (غير المشاهدة) مع

الحالات المشاهدة. يوضح (الشكل 25.4) مثال بسيط في كيفية استخدام حالتين في جزء من نموذج ماركوف المخفي لتمثيل (أو توليد) توالي. لغرض تطوير أداء (وظيفة) نموذج ماركوف المخفي الذي يمكن أن يستخدم لتمثيل اصطفايف التوالي أفضل تمثيل ممكن ، فإن النموذج الإحصائي (نموذج ماركوف المخفي) يجب أن يدرّب ، بعبارة أخرى ، هي عبارة عن عملية الحصول على المعلومات الإحصائية المثلى في نموذج ماركوف المخفي. تتضمن عملية التدريب حساب تكرارات الثمالات في كل عمود في الاصطفايف المتعدد التي تبني من مجموعة من التواليات ذات العلاقة (مترابطة). تستخدم القيم التكرارية لتحديد قيم الاحتمالات الانتقالية والانبثاق في النموذج بطريقة مماثلة لبناء مصفوفات درجات المواقع المحددة PSSM . وعندما يتم بناء نموذج ماركوف المخفي المستند على تدريب عدد من التواليات ، عندئذ يمكن استخدامه لتحديد مدى تطابق التوالي غير المعروف مع النموذج.

لغرض استخدام نموذج ماركوف المخفي لوصف اصطفايف التوالي المتعدد بوجود فجوات ، فإن الحرف (الرمز) في الاصطفايف يمكن أن يكون في واحدة من الحالات الثلاثة ، تطابق أو الإضافة أو الحذف. يعبر عن حالة التطابق بأنها حالة مشاهدة ، في حين حالتي الإضافة والحذف يعبر عنها بأنها حالتي غير مشاهدة (أو مخفية) .

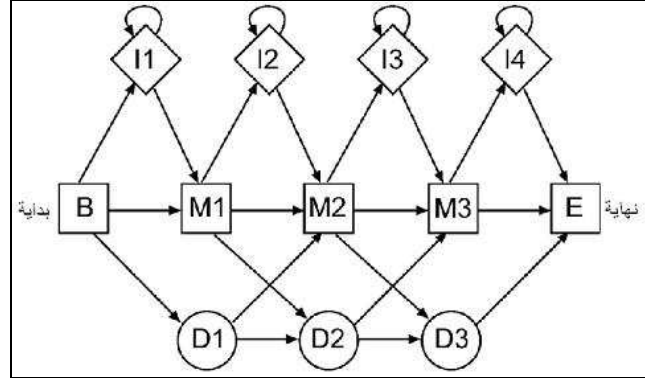


شكل 25.4 : رسم تخطيطي لنموذج ماركوف جزئي لتوالي DNA مع وجود قيم الاحتمالات (الانتقالية والانبثاق)

على سبيل المثال لتوليد التوالي AG (في الشكل أعلاه) ، فإن عملية الحساب في هذا النموذج تحسب من الرمز A من الحالة الأولى إلى الرمز G في الحالة الثانية ، وإن قيمة الاحتمال لهذا المسار هو $0.80 \times 0.40 \times 0.32 = 0.102$ (تمثل قيم الاحتمالية في الحالة الأولى مقدارها 0.8 مضروبة في احتمالية الانتقال 0.4 وقيمة الاحتمالية لـ G في الحالة الثانية مقدارها 0.32) . ومن الواضح هنالك 16 (4×4) مسار مختلف يمكن توليده بين الحالتين الأولى والثانية.

يمكن تمثيل الحالات الثلاثة في نموذج ماركوف المخفي برسم تخطيطي خاص ، حيث يستخدم في هذا الرسم عملية الانتقال من حالة إلى حالة أخرى والانتقال من الجهة اليسرى إلى الجهة

اليمنى بواسطة مسالك (مسارات) مختلفة من خلال النموذج الذي يمثل حسابات جميع الاحتمالات للتطابقات وعدم التطابقات والفجوات لتوليد الاصطفا. إن كل مسار يرتبط مع قيمة احتمالية وحيدة (شكل 26.4).



شكل 26.4 : معمارية نموذجية لنموذج ماركوف المخفي ويمثل اصطفا تواليات متعددة . تشير المربعات الى حالة التطابق (M) والأشكال المعينية تشير الى حالات الإضافة (I) والدوائر الى حالات الحذف (D) وحالتى البداية والنهاية من حالات التطابق B و E على التوالي وتمثل الأسهم مسار العملية . تشير الدوائر أعلى الرسم الى حالة الإضافة والى حالة التكرار الذاتي (التكرار حول حالة النقطة نفسها) ، والتي تسمح بإضافة أي عدد من الثمالات المتطابقة في النموذج ، إضافة الى ذلك توجد حالتى البداية والنهاية

هنالك العديد من الحسابات المحتملة للحالات أو المسارات للانتقال خلال النموذج ، من حالة البداية وصولاً إلى حالة النهاية. كل مسار يولد توالي وحيد ، الذي يتضمن الإضافات أو الحذف وبقيم احتمالية مختلفة . في نموذج ماركوف المخفي قد يكون هنالك مساراً يمثل يولد التابع الأكثر احتمالية لتمثيل الاصطفا الأمثل لعائلة من التواليات . حيث يوضح الشكل (26.4) نموذج ماركوف مخفي ، الاصطفا الذي يبدأ من المربع (B) ينطلق ثلاثة مسارات ، المسار الأول باتجاه المربع التالي (M1) وتمثل حالة مشاهدة ، ومسارين باتجاهين مختلفين أحدهما باتجاه المعين (وتمثل حالة الإضافة I1) والآخر باتجاه الدائرة (وتمثل حالة الحذف D1) وكلاهما تمثل حالتين غير مشاهدتين .

في حين إن المعين الأول فيه ثلاث عمليات ، الأولى ينطلق إليها سهم من مربع البداية (B) (هذه العملية تعني الإضافة) والحالة الثانية ينطلق سهم منها (من المعين) إلى المربع الأول (M1) والحالة الأخيرة هو خروج ودخول سهم من وإلى المعين الأول (نفسه) الذي يمثل إضافة لأي عدد من الثمالات . أما الدائرة (D1) فيها أيضاً ثلاث عمليات ، الأولى ينطلق إليها سهم من مربع البداية (B) ويخرج منها سهمين الأول إلى المربع الثاني (M2) والثاني إلى الدائرة الثانية (D2) وهكذا الحال بالنسبة إلى باقي الأشكال الأخرى وصولاً إلى المربع الأخير الذي هو مربع النهاية (E) .

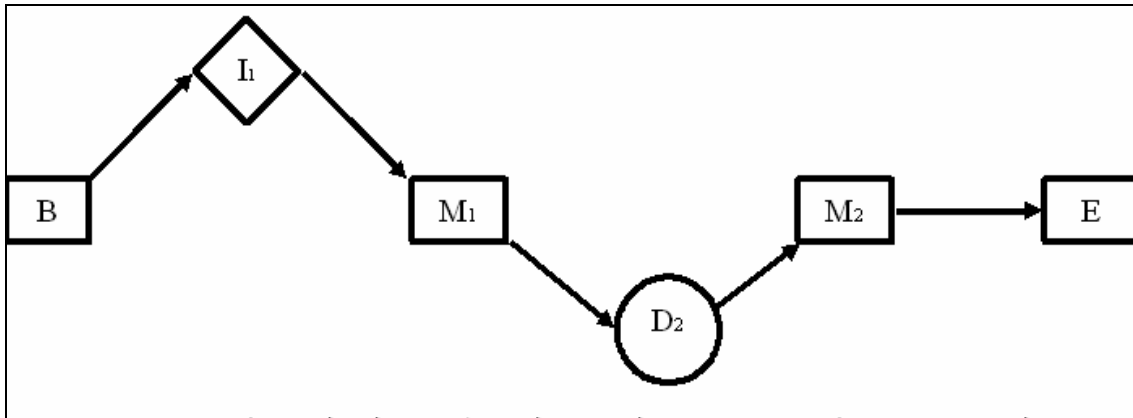
مثال 39.4 : تبدأ عملية الاصطفا من مربع البداية (B) الذي يمثل تطابق ثمالتين هما A من التوالين (شكل 27.4) العملية التالية هو الانتقال إلى إضافة فجوة واحدة لذلك يتم الانتقال إلى المعين الأول ، ثم بعد ذلك يتم الانتقال إلى المربع الأول الذي يمثل عملية تطابق بين ثمالتين

(التطابق بين الثمالتين G شكل 28.4) ، العملية التالية تتضمن الانتقال إلى الدائرة الثانية (D2) لحذف ثمالة (حذف الثمالة T شكل 28.4) ثم الانتقال بعد ذلك إلى المربع الثاني (M2) الذي يمثل عملية تطابق بين ثمالتين (تطابق ثمالتين T شكل 27.4) ثم العملية التالية تتضمن الانتقال مباشرة إلى المربع الأخير (E) بدون عمليتي الإضافة والحذف الذي يمثل تطابق الثمالتين A (شكل 27.4).

Seq. #1	A	G	T	T	A
Seq. #2	A	C	G	T	A
Seq. #1	A	—	G	T	A
Seq. #2	A	C	G	T	A

شكل 27.4 : اصطفاف توالين باستخدام نموذج ماركوف الخفي

بإضافة فجوة واحدة بعد الثمالة الأولى A في التوالي الأول (seq. # 1) (مقابل الثمالة C في التوالي الثاني seq. # 2) للحصول على تطابق ثاني (بعد التطابق الأول للثمالتين A) للثمالتين G في التوالين ثم بعد ذلك حذفت الثمالة T من التوالي الأول للحصول على تطابق بين الثمالتين T في التوالين وأخيرا تم الانتقال إلى الثمالة الأخيرة (الحالة الأخيرة) بدون حذف أو إضافة والحصول على تطابق بين الثمالتين A في التوالين وكما هو موضح في الشكل (28.4).



شكل 28.4 : معمارية نموذج ماركوف الخفي يمثل اصطفاف التوالين الموضحة في الشكل (27.4)

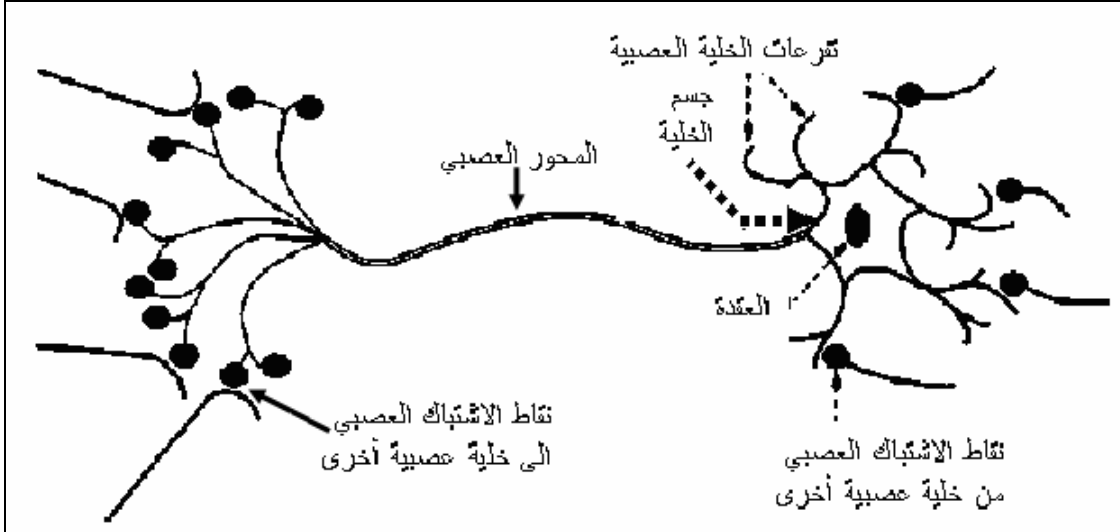
الشبكات العصبية الاصطناعية Artificial neural networks

انبثقت أهمية الشبكات العصبية الاصطناعية من اتجاهين رئيسين ، الاتجاه الأول ، هو طريقة الحساب التي تمتاز بها ، والاتجاه الثاني يعتمد على طريقة الترابط بين أجزائها. إن الشبكات العصبية الاصطناعية ، هو عبارة عن دراسة أو محاكاة سلوك الذكاء لدى الإنسان وكذلك يتعلق أيضا بتنفيذ برامج الحاسوب الذي يظهر (عند تنفيذ البرنامج) ذكاء في سلوكه.

الشبكات العصبية البيولوجية Biological neural networks

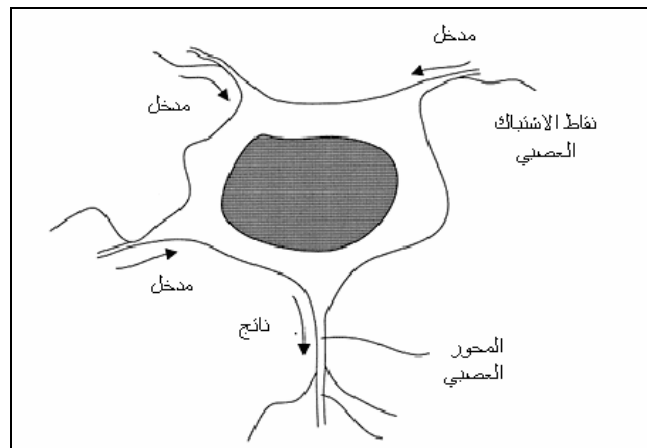
الجهاز العصبي عبارة عن شبكات عصبية معقدة وضخمة . يعد الدماغ العنصر المركزي في الجهاز العصبي . يتألف الدماغ من حوالي 10^{11} من الخلايا العصبية التي تترابط فيما بينها بواسطة العقدة . تتألف العقدة من عناقيد من الخلايا العصبية والتي لها وظائف خاصة ومحددة ، وتعمل العقدة على معالجة المعلومات القادمة من الأجزاء (الأطراف) الحسية قبل إرسالها إلى أجزاء أخرى داخل الشبكة. يتضمن الجزء النهائي من العملية توزيع الإشارات لغرض توليد الفعل ، بعبارة أخرى ، فإن عملية إطلاق الإشارة تبعث أو لا تبعث من قبل العقدة فهي بذلك تعد مركز القرار في إطلاق أو عدم إطلاق الإشارة إلى الأطراف الأخرى.

إن الجهاز الحسي الذي يرتبط بالدماغ يعمل بإتقان عند تحليله للمعلومات الحسية المعقدة إلى المركبات الأساسية التي تمثل المميزات والصفات الأساسية للحاسة . على سبيل المثال ، تقسم العيون والدماغ الصورة المرئية إلى الألوان والكثافة والاتجاه والحركات والمقاييس (الأبعاد) الخ ، هذه المكونات الأساسية لا يتم إعادة بنائها إلى الشكل المرئي الأصلي (الصورة المرئية) لكن بدلا من ذلك يتم إرسالها (المكونات) إلى شبكات فرعية أخرى لإجراء تقييمات أخرى مختارة على هذه المكونات وكذلك إعادة بنائها جزئيا ، بعبارة أخرى يتم اختبار جزء من المكونات المعينة لكي يتم إعادة بنائها. وقد استلهمت فكرة الخلايا العصبية الاصطناعية من طريقة عمل الخلايا العصبية البيولوجية. حيث إن الدماغ وأجزاء أخرى من الجهاز العصبي يحتوي على تنوعا واسعا من الأنواع المختلفة من الخلايا العصبية ، التي تختلف في خصائصها الالكترونية وأعدادها وإحجامها وأنواع الاتصالات .



شكل 29.4 : خلية عصبية حيوية

تتألف الخلية العصبية البايولوجية من ثلاثة مكونات رئيسية : تفرعات الخلية العصبية Dendrites وجسم الخلية Cell body (Soma) والمحور العصبي Axon (شكل 29.4). تستلم تفرعات الخلية العصبية الإشارات القادمة من خلايا عصبية أخرى . كما تتصل المحاور العصبية مع التفرعات العصبية وأسطح أجسام الخلايا بواسطة نقاط الاشتباك العصبي Synapses واستنادا إلى الخصائص الالكترونية التي تمتاز بها أغشية الخلية العصبية فان الإشارات التي تستلمها تفرعات الخلية العصبية تتلاشى سريعا بمرور الوقت وطول المسافة (الحيز) ، وبالتالي فقدان القدرة على تحفيز الخلية العصبية ما لم يتم تقوية الإشارة بإشارات أخرى والتي تحدث في اغلب الأحيان في الوقت نفسه وموقع قريب (شكل 30.4). إن خلية الجسد تعمل على تجميع الإشارات القادمة من التفرعات وكذلك تجميع الإشارات القادمة من نقاط الاشتباك العصبي .



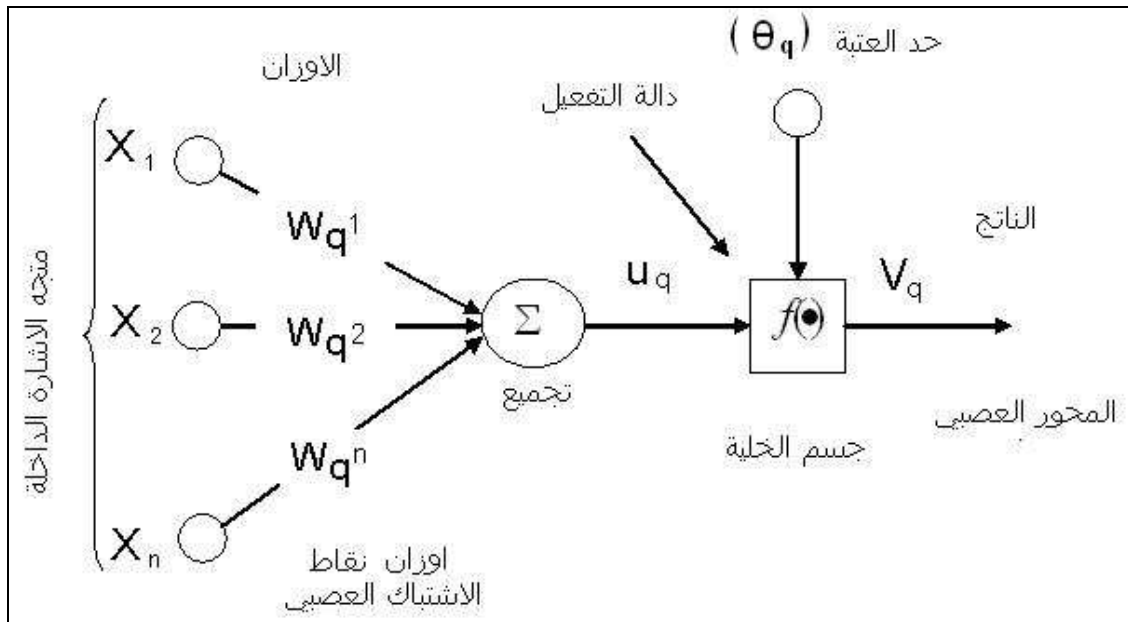
شكل 30.4 : اتجاه سريان الإشارة في الخلية العصبية

فعندما يتوفر أو يتجمع عدد كاف من الإشارات الداخلة إلى الخلية بحيث تكون مساوية أو أكبر من حد العتبة Threshold ، عندها تتولد لدى الخلية العصبية إمكانية إنجاز عمل (القوة الكامنة) ، والتي يمكن أن تطلق الإيعاز (الإشارة أو الأمر) لتنتقل عبر المحور العصبي إلى خلايا أخرى أو إلى خلايا الهدف خارج الجهاز العصبي ، مثال ذلك النقل إلى العضلات . أما إذا كانت مجموع المدخلات لا تصل إلى حد العتبة ، فإن هذه المدخلات سرعان ما تتلاشى وبالتالي لا يتم توليد أية قوة كامنة (شكل 30.4) . وهكذا ، فإنه ينظر إلى إمكانية توليد قوة كامنة أو لا ، وذلك لأن المحور العصبي إما أن يعمل على توليدها أو لا . إضافة إلى ما تقدم ، يعبر عن قوة المدخلات بعدد القوى الكامنة التي يمكن توليدها لكل دقيقة إضافة إلى حجمها .

على سبيل المثال ، فإن حجم وشكل القوة الكامنة الناتجة من مدخل (إشارة) قوي ستكون مساوية إلى حجم وشكل القوة الكامنة الناتجة من مدخل ضعيف . إلا إن المدخل القوي سيولد قوى كامنة خلال الدقيقة أكثر مقارنة مع المدخل الضعيف .

نماذج أساسية من الخلايا العصبية الاصطناعية Basic models of artificial neurons

يعبر عن الخلية العصبية بمعالج العنصر والعقدة ، أو وحدة حد العتبة المنطقية ، وهي عبارة عن وحدة معالجة المعلومات (معالجة الإشارات الداخلة) والتي تماثل (في تركيبها) - بصورة أولية - الخلية البيولوجية .



شكل 31.4 : خلية عصبية صناعية

يوضح شكل 31.4 نموذج خلية عصبية اصطناعية. يتألف نموذج الخلية العصبية من أربعة مكونات أساسية هي :

1- مجموعة من نقاط الاشتباك العصبي مع الأوزان المرافقة لها ، وكما هو واضح بالشكل (31.4) فإن قيم المدخلات إلى نقاط الاشتباك العصبي هي عبارة عن متجه من الإشارات يرمز له $\mathbf{X} \in \mathbb{R}^{n \times 1}$ ، حيث . حيث إن $\mathbf{X}$ هو $\mathbf{X} = [x_1, x_2, x_3, \dots, x_n]$ وان x_j عبارة عن المدخل j ، وان $j=1, \dots, n$ ، وهكذا فإن كل مكون x_j من مكونات المتجه هو مدخل إلى نقطة الاشتباك العصبي j وتتصل بالخلية العصبية q من خلال وزن نقطة الاشتباك العصبي والتي يرمز لها W_{qj} . حيث إن x_j تضرب بوزن نقطة الاشتباك العصبي W_{qj} . وتم إتباع التقليد السائد في اغلب المصادر بتعريف الرموز الذيلية في أوزان نقاط الاشتباك العصبي ، حيث يشير الرمز الذيلي الأول (q) إلى الخلية العصبية في حين يشير الرقم الذيلي الثاني إلى عنصر متجه الإدخال .

2- تعمل أداة التجميع (التي تمثل ملتقى حاصل ضرب المدخلات x_j مع الأوزان المرافقة لها W_{qj}) على إضافة وتجميع جميع الإشارات القادمة إليها ، ذلك إن كل مدخل يضرب بالوزن (بوزن نقطة الاشتباك العصبي) المرافق له وبعد ذلك يتم جمعها ، وبعد أن تجمع هذه المكونات يصدر الناتج u_q والذي يتألف من توليفة خطية ، وذلك لأن u_q هو توليفة خطية من المدخلات مع نقاط (أوزان) الاشتباك العصبي.

3- تعمل دالة التفعيل (Activation function) $f(.)$ الموضحة في الشكل (31.4) على معالجة ناتج (مخرج) الخلية العصبية V_q ، فعندما تكون دالة التفعيل لا خطية فإنها تعمل على الحد من غزارة تدفق ناتج الخلية العصبية V_q . حيث تكون محددة - بصورة طبيعية - ضمن المدى $[0, 1]$ (ثنائي) أو $[-1, 1]$ (ثنائي القطب) . في حالة الارتباطات الداخلية الكثيرة (داخل الشبكة العصبية) ذات معمارية أسلوب بناء متوازي ضخم للخلايا العصبية الاصطناعية فإن اللاخطية (النموذج اللاخطي) - المستخدمة في دوال التفعيل (دوال التفعيل اللاخطية) - يمكن أن تعمل على تحسين العمل في تصانيف الشبكات والتقارب وإمكانية وقابلية التقليل من الضوضاء (الذي يدخل إلى الشبكة من خلال المدخلات) .

4- يطبق حد العتبة Θ_q في العادة بصورة خارجية (شكل 31.4) حيث يعمل على تقليل المجموع التراكمي للمدخل إلى الدالة التجميعية ، وبالتالي فإن Θ_q تطرح من ناتج التوليفة الخطية قبل أن تطبق عملية التفعيل . إن المدخل (القيمة الداخلة إلى الشبكة) إلى دالة التفعيل يمكن أن تزداد قيمتها وذلك بإضافة مقدار التحيز Θ_q (ترمز إلى حد العتبة أو التحيز) إلى u_q ، وذلك لأن إشارة التحيز معاكسة لإشارة حد العتبة . وهكذا فإن العلاقة بين مستوى الفاعلية (مستوى النشاط) (v_q) والذي يستخرج من المعادلة (15.4) :

$$v_q = u_q - \Theta_q \quad (15.4)$$

وناتج التوليفة الخطية (u_q) من الخلية العصبية q يعتمد على قيمة حد العتبة فيما إذا كانت موجبة أم سالبة .

يمكن وصف عملية عمل الخلية العصبية (الموضحة في الشكل 31.4) رياضيا كما يأتي :

$$u_q = \sum_{j=1}^n w_{qj} x_j = w_q^T x = x^T w_q \quad (16.4)$$

حيث إن X هو عبارة عن متجه قيمة المدخلات ، w_q متجه من الأوزان:

$$w_q = [w_{q1}, w_{q2}, \dots, w_{qn}]^T \in \mathbb{R}^{n \times 1}$$

و إن ناتج دالة التفعيل هو:

$$V_q = f(r_q) = f(u_q - \theta_q) \quad (17.4)$$

وبهذا ، وباستخدام المعادلتين (16.4) و (17.4) ، فإن ناتج الخلية العصبية هو :

$$V_q = f \left(\sum_{j=1}^n w_{qj} x_j - \theta_q \right) \quad (18.4)$$

يوضح الشكل (31.4) نموذج بديل للخلية العصبية الاصطناعية ويكتب مستوى الفاعلية (مستوى النشاط) كالآتي :

$$V_q = \sum_{j=0}^n w_{qj} x_j \quad (19.4)$$

يمكن تقسيم عمل الخلية العصبية إلى عمليتين تنجزان في داخل الخلية العصبية ، الأولى هو احتساب مجموع الأوزان لكل المدخلات (الداخلية إلى الخلية العصبية) ، في حين العملية الثانية معاملة ناتج أو مخرجات المجموع مع حد العتبة (انظر المعادلة 18.4).

الهندسة المعمارية للشبكة Network architecture

تحتوي أبسط شبكة عصبية اصطناعية على طبقتين هما ، الإدخال والأخرى طبقة النتائج أو المخرجات (عدم وجود طبقة مخفية) ، تعتمد هذه الشبكة مثل باقي الشبكات الأخرى على الوزن للربط بين (طبقتي) المدخلات و المخرجات . حيث يمثل الوزن النسبة المعنوية (الأهمية المعنوية) للمدخل المعين وأثره في حسابات الناتج (المخرج) . من المهم الأخذ بنظر الاعتبار إن الناتج المتولد (من الشبكة) سيعتمد وبصورة كبيرة على دالة التفعيل المستخدمة (في الشبكة). أما الشبكات الأكثر تعقيدا فإنها تمتلك طبقة (أو أكثر) مخفية حيث تمنح الطبقة المخفية الشبكة العصبية الاصطناعية قابلية على التعلم قوية ، من الناحية التطبيقية تستخدم الشبكات ذات ثلاث طبقات (شكل 31.4) أو أكثر.

إن على أي مصمم للشبكات يجب أن يأخذ العناصر التالية بنظر الاعتبار عند محاولة بناء الشبكة :

1. أفضل قيم بداية (القيم الأولية للأوزان "الترجيحات")
2. عدد الطبقات المخفية
3. عدد الخلايا العصبية في كل طبقة مخفية
4. عدد متغيرات الإدخال
5. قيمة نسبة التعلم
6. قيمة نسبة العزم

7. زمن التدريب أو مقدار التدريب (تحديد عدد التكرارات المستخدمة)

8. نوع دالة التفعيل المستخدمة في الطبقتين المخفية والنتائج

1- الأوزان الابتدائية (المرجحات الابتدائية)

يمتلك كل ارتباط (الذي يربط بين طبقتين) معلمة ترافقه تدل على قوة هذا الارتباط ، والتي تدعى بالوزن . يمكن للشبكة أن تتعلم على النماذج المقدمة من خلال طبقة المدخلات وذلك بواسطة تغيير قيم الأوزان (بطريقة معينة) ، ويمكن الاستفادة من الخبرة المتوفرة في تحديد الأوزان الأولية وغالبا ما يتم تحديد قيم عشوائية صغيرة للأوزان الابتدائية.

2- عدد الطبقات المخفية

إن من الضروريات المهمة التي تواجه المصمم هو في الاستدلال على عدد الطبقات عند تحديد أسلوب تصميم معمارية الشبكة، فعلى سبيل المثال إذا كان نموذج الشبكة العصبية مكون من ثلاث طبقات ، فإنه يحتوي على طبقة مخفية واحدة . أما إذا احتوى نموذج الشبكة العصبية على أربعة طبقات فإنه يحوي على طبقتين مخفيتين . تمكن الطبقة المخفية للشبكة العصبية القدرة على التعميم ، بعبارة أخرى ، تساعد الطبقة المخفية الشبكة العصبية على التعلم ومن ثم اتخاذ القرار. في الجانب العملي ، فإن الشبكة العصبية ذات الطبقة المخفية الواحدة وفي بعض الأحيان ذات الطبقتين المخفيتين تستخدم بصورة واسعة ولها أداء حسنا جدا.

3- عدد الوحدات أو الخلايا العصبية في الطبقة المخفية

تعد عملية تحديد عدد الوحدات (عدد الخلايا العصبية) في طبقة مخفية واحدة من أهم المهامات التي تواجه مصمم نموذج الشبكات العصبية، وأفضل طريقة لتحديد أو معرفة ذلك استخدام نموذج الملائمة المتسلسلة وذلك بإضافة وحدة (خلية عصبية) مخفية واحدة في كل مرة ، بعبارة أخرى يتم إضافة وحدة مخفية واحدة إلى الطبقة المخفية وينفذ عمل الشبكة العصبية ومن ثم تقييم النتائج ، بعد ذلك تضاف وحدة مخفية أخرى إلى الطبقة الخفية ومرة أخرى يجري تنفيذ عمل الشبكة العصبية ومن ثم مرة أخرى تقييم النتائج. بعد ذلك تقارن نتائج التنفيذين (الأول والثاني) ثم يتم اتخاذ قرار بخصوص أفضل النتائج (نتائج التنفيذين الأول أم الثاني) أو يصار إلى إضافة وحدة مخفية (أو أكثر) إلى الطبقة المخفية ومراقبة نتائج التنفيذ ... وهكذا وصولا إلى أفضل النتائج. بحيث أنه عند إضافة وحدة مخفية جديدة لا تؤثر أو تحسن في أداء الشبكة .

في حالة احتواء الشبكة العصبية على عدد قليل من الخلايا العصبية فإن ذلك سيتطلب عمليات حسابية قليلة (عند تنفيذ الشبكة العصبية) وكذلك تحتاج وقت قصير لأداء هذه المهمة . أما إذا احتوت الشبكة العصبية على عدد كبير من الخلايا العصبية (وعدد كبير من الطبقات) فإن ذلك يتطلب استخدام أوزان إضافية أكثر ، وبالتالي يتطلب عمليات حسابية كثيرة وكذلك يتطلب وقت طويل لأداء هذه المهمة. في الواقع إن قابلية تعلم الشبكة العصبية تعتمد على عدد الخلايا العصبية في الطبقة (الطبقات) المخفية ، كما في حالة كون عدد الخلايا العصبية قليلة جدا ، سيكون النموذج (الخاص بالشبكات العصبية) غير مرن بما فيه الكفاية بالنسبة للتعامل مع البيانات (تحت الدراسة). بالمقابل في حالة كون عدد الخلايا العصبية كثيرة جدا فإن النموذج (الخاص بالشبكات العصبية) سيكون غير مطابق للبيانات .

كما إن عدد العقد (الخلايا العصبية) ليس بالإمكان تحديدها مقدماً ، فإن الخبرة التجريبية ضرورية لتحديد هذه المعلمة.

4- عدد المدخلات

في حالة كون عدد المتغيرات المستقلة (موضوع الدراسة) n ، فإنه يجب أن يقابلها (في الشبكة العصبية) عدد n من عقد (الخلايا العصبية) الإدخال في النموذج المقترح . في الجانب التطبيقي يتم اختيار أبسط النماذج والذي يحتوي على عدد قليل من المتغيرات .

5- نسبة التعلم

تصنف نماذج الشبكة العصبية باعتماد أسلوب (طريقة) التعلم أو التعديل التكراري إلى التعلم بالإشراف (الإشراف يكون من قبل الباحث الذي يحدد الهدف أو الأهداف المطلوبة لعمل الشبكة العصبية) أو التعلم بدون إشراف.

في حالة التعلم بالإشراف ، تدرب الشبكة وذلك بتجهيزها بالمدخلات وفي الوقت نفسه توفر لها قيم الناتج (المخرجات) الهدف ، بعبارة أخرى يتم تحديد قيم المدخلات والمخرجات معا . ومن ثم يحدد لاحقا ، الخطأ ، وذلك باحتساب الفرق بين قيم الناتج الهدف وبين قيم الناتج الحقيقية (القيم الناتجة من تنفيذ الشبكة العصبية) ، ثم يتم استخدام خوارزمية التدريب حتى يتم التوصل إلى مستوى الخطأ المسموح (الوصول إلى الفرق المسموح بين قيمة (ناتج) الهدف المطلوب وبين قيمة النواتج الحقيقية (الناتج من الشبكة العصبية)).

أما في حالة التعلم بدون إشراف ، يوجه التعديل التكراري فقط بواسطة متجه الإدخال ، بذلك لا توجد تغذية خلفية من البيئة (عدم وجود قيم (ناتج) الهدف التي تحدد من قبل المشرف) لتحديد فيما إذا كانت نواتج الشبكة هي صحيحة . من ثم ، فإن على الشبكة أن تكتشف المميزات (أو الصفات) الموجودة في البيانات بطريقتها الخاصة . في الواقع ، ان طريقة التعلم بدون إشراف لا تستخدم قيم (ناتج) الهدف ، وإنما هي تستخدم معلومات الإدخال لتصنيف نماذج الإدخال.

تسيطر نسبة التعلم أو معلمات التعليم على النسبة التي عندها يعدل الوزن ويحدث التعليم (تعليم الشبكة العصبية على النماذج) . تحدد نسبة التعلم سرعة تعلم الشبكة العصبية، كذلك فإن نسبة التعلم تحدد حجم الخطوة المطلوبة للاتجاه ناحية الحد الأدنى من دالة الخطأ، حيث إن نسبة التعلم العالي تقود إلى حدوث تذبذب حول الحد الأدنى للخطأ في حين نسبة التعلم القليلة تقود إلى التقليل من تقارب نتائج الشبكة العصبية. إن نسبة التعلم الكبيرة تؤدي إلى تغيرات كبيرة في الأوزان ومن ثم تسرع في تعليم الشبكة العصبية. من ناحية أخرى ، تؤدي نسبة التعلم الكبيرة إلى حدوث تذبذب في الأوزان تسبب معها الوصول إلى حلول غير كافية . الوجه الآخر للحالة ، ان الخطوات الكبيرة (الناتجة من كون نسبة التعلم كبيرة) - لتعليم الشبكة - تؤدي إلى سرعة التقارب لكنها قد تتجاوز الحل أو إنها تتجه في الاتجاه الخاطئ . بالمقابل ، فإن الخطوات الصغيرة (الناتجة عن كون نسبة التعلم صغيرة) تقود إلى الاتجاه الصحيح والمناسب لكنها تحتاج عدداً كبيراً من التكرارات (تكرار عمل الشبكة العصبية وصولاً إلى النتيجة المطلوبة) . وتعتمد عملية تحديد قيمة نسبة التعلم على نوع التطبيق المحدد.

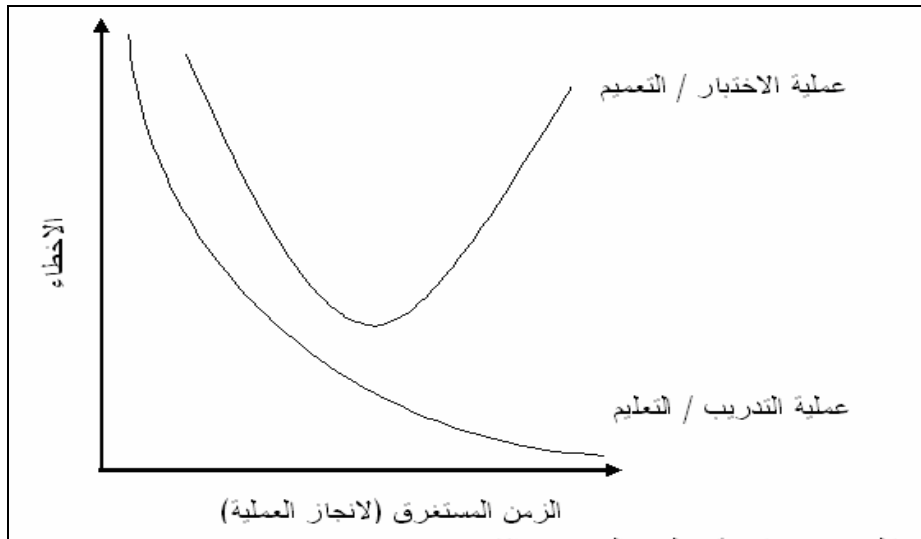
6 - نسبة العزم

إن الهدف من وراء استخدام نسبة العزم هو لغرض الوصول السريع إلى الحد الأدنى للخطأ ، وفي حالة استبعاد هذا الحد أي نسبة العزم فإن عمل (تنفيذ) الشبكة سيحتاج إلى وقت أطول قبل الوصول إلى الحد الأدنى للخطأ.

من الناحية التقنية ، يحدد مقدار العزم كيفية تغيرات الأوزان السابقة (التكرار السابق في الشبكة) استناداً إلى التغيرات التي تحدث في الأوزان الحالية (التكرار الحالي للشبكة). حيث إن استخدام هذا الحد سيجنب حدوث تذبذب في نتائج الشبكة . كذلك تحفز نسبة العزم في الاستمرار في تغيير الأوزان بالاتجاه نفسه وجعل التغيير في الأوزان الحالية لا يعتمد فقط على التغير في الأوزان السابقة فقط ولكنه يعتمد أيضاً على الخطأ الحالي .

7-التدريب

يحدد التدريب (تدريب الشبكة لغرض تعليمها على الأمثلة المقدمة إليها من خلال طبقة الإدخال) استناداً إلى العملية التي من خلالها يتم إدخال البيانات إلى الشبكة العصبية سوية مع قيم النتائج المناظرة لها (بعبارة أخرى ، النتائج - المناظرة - التي تظهر بعد تنفيذ عمل الشبكة على القيم المدخلة إليها من خلال طبقة الإدخال الخاصة بها) بحيث إن الشبكة يمكن أن تعدل الأوزان بطريقة يمكن معها أن تضمن الحصول على أقل خطأ ممكن . يستمر تنفيذ عملية التدريب حتى الوصول إلى أقل خطأ ممكن بين عملية تكرار تدريب الشبكة وبين تكرار عملية الاختبار (شكل 32.4) . أو يستمر تكرار تنفيذ التدريب حتى يتم الوصول إلى نقطة بعدها لا يوجد فيها تحسن في دالة الخطأ استناداً إلى العدد المحدد للأوزان الأولية (الابتدائية) المختارة بشكل عشوائي.



شكل 32.4 : مقارنة عمليات التدريب والاختبار

يتأثر التدريب بعدد من المعلمات مثل اختيار نسبة التعلم وقيم العزوم وعدد الطبقات المخفية المستخدمة.

8- دالة التفعيل

يعد اختيار دالة التفعيل أو دالة السكواش Squashing function أو دالة التنشيط المسماة بالسكمويد Sigmoid activation function (الشبيه بحرف S) من الأمور المهمة في تصميم أية شبكة عصبية . وذلك لأن هذه الدالة لها دور أساسي في التأثير في نتائج الشبكة العصبية. حيث تعمل دالة التفعيل على تحويل - رياضيا - مجموع القيم المدخلة في وحدات الإدخال لغرض تعميمها على الخلايا العصبية في طبقة النماذج. في العادة ، تستخدم الدراسات دالة التفعيل السكمويد Sigmoid activation function للطبقة الخفية والدالة الخطية أو دالة السكمويد لطبقة النماذج (المخرجات).

تتجه اغلب الدراسات - في مجال استخدام الشبكات العصبية - نحو استخدام دوال تحويل السكمويد . إن فائدة هذه الدالة هي في كونها دالة مستمرة (قيمها مستمرة) رتيبة والتي تعطي ناتجين هما حالة الظهور أو عدمه . ويمكن الإشارة إلى أن الشبكة العصبية التي تستخدم دالة السكمويد في الخلايا العصبية المخفية وكذلك في الخلايا العصبية للنماذج تدعى بالشبكة العصبية ذات الإدراك الحسي متعددة الطبقات Multi layer perceptron وتحدد دالة التفعيل السكمويد كما يأتي :

$$g(x) = \frac{1}{1 + e^{(-\alpha x)}}$$

حيث إن :

x: قيمة المدخل إلى الشبكة

α : هي معلمة الميل

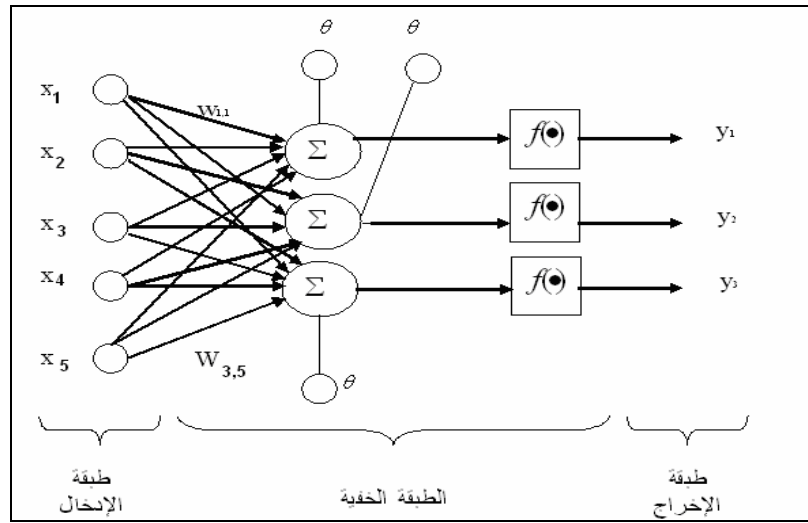
مثال (40.4): توالي من الأحماض الامينية عددها 650 حامض أميني يمثل التركيب الأولي لأحد البروتينات. استخدمت طريقة الشبكات العصبية لغرض التنبؤ بالتركيب الثانوي وكالاتي :

1- **الخطوة الأولى :** ترتيب (التركيب الأولي) توالي الأحماض الامينية على شكل صف (صفوف) كما موضح في شكل (33.4)

1	2	3	4	5	6	7	8	9	10
A	C	G	A	A	A	T	P	L	F
F	F	A	A	A	C	C	C	G	G
L	L	L	P	P	K	K	K	G	A
⋮									
641	642	643	644	645	646	647	648	649	650
A	A	F	S	T	L	I	C	G	C

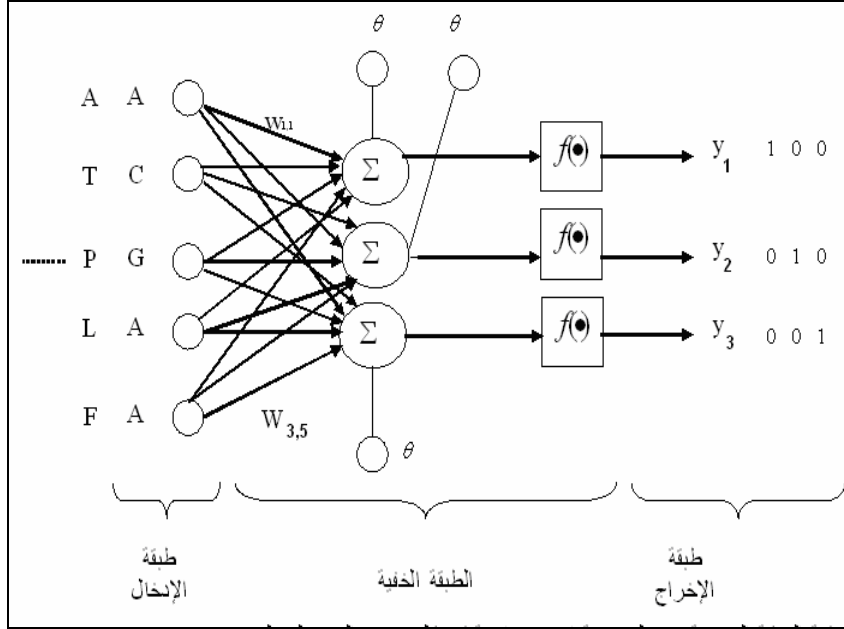
شكل 33.4 : توالي مكون من 650 حامض اميني

2- الخطوة الثانية : تصميم شبكة عصبية ، مكونة من 5 مدخلات (تمثل طبقة الإدخال) وثلاثة عقد مخفية (الطبقة المخفية) وثلاثة مخرجات (طبقة الإخراج) شكل (34.4).



شكل 34.4 : شبكة عصبية مكونة من خمسة مدخلات وثلاثة عقد خفية وثلاثة مخرجات

3- الخطوة الثالثة : تنفيذ البرنامج الخاص بالشبكة العصبية وكما هو موضح في الشكل (35.4) :



شكل 35.4 : تغذية الشبكة العصبية بإدخال خمسة أحماض أمينية في كل مرة والحصول على ثلاث مخرجات

- يتم إدخال (في كل مرة) خمسة ثمالات دفعة واحدة ابتداء من الطرف الاميني N-terminal للتوالي (خمس أحرف وكل حرف منها يمثل حامض أميني) إلى الشبكة عن طريق طبقة الإدخال وبعد تنفيذ البرنامج يتم الحصول على إحدى النواتج الآتية :
- (1,0,0) يرمز إلى حلزون ألفا α -helix ، بعبارة أخرى عند ظهور هذا الناتج من مخرجات الشبكة فيعني ذلك إن الثمالات الخمسة (في تركيبها الأولي وهي ACGAA) هذه (على سبيل المثال) ستكون منطقة الحلزون - ألفا في التركيب الثانوي.
 - (0,1,0) يرمز إلى β -sheet ، بعبارة أخرى ، عند ظهور هذا الناتج من مخرجات الشبكة فيعني ذلك إن الثمالات الخمسة (في تركيبها الأولي وهي ACGAA) هذه (على سبيل المثال) ستكون منطقة صفائح - بيتا في التركيب الثانوي.
 - (0,0,1) يرمز إلى Coil ، بعبارة أخرى ، عند ظهور هذا الناتج من مخرجات الشبكة فيعني ذلك إن الثمالات الخمسة (في تركيبها الأولي وهي ACGAA) هذه (على سبيل المثال) ستكون منطقة التفاف عشوائي في التركيب الثانوي.
- تستمر عملية إدخال الثمالات (خمس ثمالات في كل مرة يتم فيها تنفيذ البرنامج) إلى الشبكة العصبية لغاية انتهاء التوالي والوصول إلى النهاية الكربوكسيلية الطرفية C-terminal

التحليل العنقودي Cluster analysis

تقنية إحصائية تستخدم لتجميع وتصنيف المشاهدات أو الأشياء إلى مجاميع (عناقيد) حيث إن المشاهدات في المجموعة (العنقود) نفسها تكون أكثر تماثلاً مقارنة مع المشاهدات المتجمعة في عناقيد أخرى . بعبارة أخرى ، يمكن تعريف التحليل العنقودي ، بأنه أسلوب من أساليب متعدد المتغيرات يهدف بصورة أساسية لتشخيص وتصنيف الأشياء (المشاهدات) أو المتغيرات بحيث تتجمع كل مفردة مماثلة إلى المفردات الأخرى في العنقود الواحد استناداً إلى معيار محدد مسبقاً . إن الناتج من عملية العنقدة هذه هو الحصول على مجموعة من المشاهدات (الأشياء) داخل العنقود الواحد تمتاز بكونها ذات تجانس عالي داخلي (داخل - العنقود الواحد) وتبدي عدم تجانس عالي خارجي (بين العناقيد المختلفة) وفي حالة نجاح شمولية التصنيف ستكون المشاهدات داخل العناقيد قريبة بعضها من البعض في حالة تمثيلها بيانياً أما المشاهدات في العناقيد المختلفة فتكون متباعدة.

يعد التحليل العنقودي أداة مفيدة لتحليل البيانات في العديد من الحالات المختلفة على سبيل المثال ، الباحثون الذين يجمعون البيانات بالاستبيانات قد يواجهون أعداد كبيرة من المشاهدات والتي هي بدون معنى ما لم تصنف إلى مجاميع ذات معنى . ولذلك يمكن استخدام التحليل العنقودي لانجاز إجراءات تقليص (تصنيف) البيانات بتقليص المعلومات من المجتمع بأكمله إلى معلومات عن مجاميع فرعية صغيرة ومحددة وبهذا الأسلوب يحصل الباحثون على مشاهدات مختصرة يمكن وصفها بأقل خسارة أو ضياع المعلومات .

كذلك يمكن للباحثين الاستفادة من استخدام التحليل العنقودي في حالة الرغبة بتطوير الفرضيات التي تتعلق بطبيعة البيانات . على سبيل المثال ، يعتقد الباحث وجود علاقة بين كمية الضوء مع استهلاك الطعام ، حيث يمكن استخدام التحليل العنقودي لفصل مستهلكي الطعام إلى أجزاء منطقية أو إلى مجاميع مختلفة باعتماد كمية الضوء مقابل استهلاك الطعام.

تطبيق التحليل العنقودي Application of cluster analysis

يمكن تقسيم عملية تطبيق التحليل العنقودي إلى الخطوات الرئيسية الآتية :

1- التجزئة

2- التفسير

3- إيجاد المظهر الجانبي أو النسق

إن مرحلة التجزئة هي عملية تحديد عدد العناقيد التي تتجمع فيها المشاهدات . أما مرحلة التفسير فهي عملية فهم الخصائص لكل عنقود وتحديد الاسم والعنوان الذي يلائم ويعرف طبيعة العنقود في حين يتضمن المظهر الجانبي وصف الخصائص المتعلقة بكل عنقود لتفسير مقدار الاختلاف بين هذه العناقيد .

المرحلة الأولى : تتضمن هذه المرحلة ثلاث محاور يمكن أن تؤخذ بنظر الاعتبار هي:

- 1- كيفية قياس التماثل الداخلي (داخل العنقود أو المجموعة الواحدة) بين المشاهدات .
- 2- كيفية تحديد الإجراء (الخوارزمية) التي يجب استخدامها لوضع المشاهدات المتشابهة في عناقيد أو مجاميع خاصة بها.
- 3- كيفية تحديد عدد العناقيد التي يجب أن تشكل.

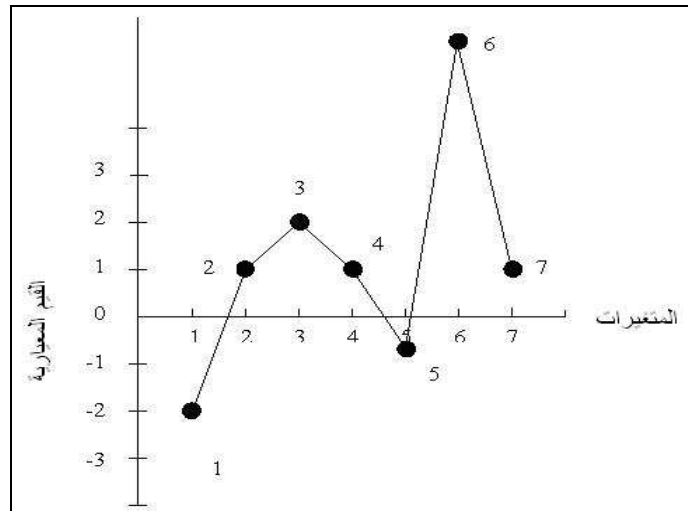
هنالك العديد من الطرق التي استخدمت لغرض الإجابة عن التساؤلات الواردة في المحاور الثلاثة . إلا انه لا توجد طريقة لها القدرة على توفير إجابة محددة للأسئلة الواردة ، هذا من ناحية ، ومن ناحية أخرى فان العديد من هذه الطرق توفر (تعطي) نتائج مختلفة لمجموعة البيانات نفسها.

1- مقياس التشابه Similarity measures

يمكن أن يقاس التشابه الداخلي بين المشاهدات بعدة طرق. إحدى هذه الطرق هو تحديد التقارب (القرب) بين زوج من المشاهدات لغرض تحديد التشابه وطريقة أخرى هو تحديد المسافة أو الاختلاف بين أزواج من المشاهدات. طالما إن المسافة مكملية للتشابه فيمكن استخدام هذه الطريقة للوصول إلى التماثل الداخلي بين المشاهدات.

إن التشابه الداخلي بين المشاهدات يمكن أن ينتج من احتساب معامل الارتباط بين المتغيرات (لمصفوفة البيانات) ، فالارتباط (التشابه) العالي بين مشاهدتين يمثل تماثل بين هاتين المشاهدين ، والارتباط الواطي (القليل) يمثل نقص في التماثل بينهما.

تعد مقاييس المسافات الأكثر استخداما لقياس التشابه بين المشاهدات ولا بد الإشارة إلى إن جميع مقاييس المسافات هي حساسة للقيم الشاذة (المشاهدات التي تختلف اختلافا كبيرا عن باقي المشاهدات الأخرى، بعبارة أخرى لها قيم إما أن تكون كبيرة جدا أو تكون صغيرة جدا) . إن أسهل طريقة للتعرف على هذه القيم (الشاذة) هو برسم البيانات بيانيا كما هو موضح في الشكل (36.4)

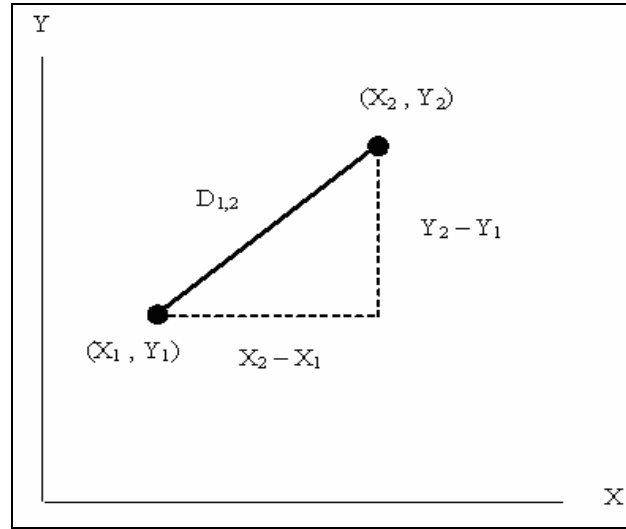


شكل 36.4 : رسم بياني لعدد من المتغيرات . تمثل النقطة رقم 6 قيمة شاذة مقارنة مع باقي القيم

إن مقياس المسافة الاقليدية Euclidean distance يعد الأكثر استخداما لقياس التماثل. يوضح الشكل (37.4) كيفية الحصول على المسافة الاقليدية . بافتراض وجود نقطتين (تحدد النقطة بمحورين ، المحور السيني ويرمز له X والمحور الصادي ويرمز له Y) ويعبر عن النقطة

الأولى (X_1, Y_1) ويعبر عن النقطة الثانية (X_2, Y_2) . يحتسب مقياس المسافة الاقليدية بين النقطتين بالطول (البعد) بين هاتين النقطتين بالمعادلة (20.4) الآتية :

$$D_{1,2} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (20.4)$$



شكل 37.4 : المسافة بين المشاهدين الأولى والثانية

حيث إن $D_{1,2}$ هي المسافة بين نقطتين الأولى والثانية ، ويمكن تعميم الفكرة لأكثر من نقطتين حددت إحداثيات المشاهدين عن طريق درجات خام في الشكل (37.4) في حين تستخدم درجة Z (Z-score) في تحليل النموذج للبيانات المعطاة لغرض احتساب المسافات لإزالة التأثيرات الكاذبة الناتجة من عدم تساوي التباينات للمتغيرات تحت الدراسة . هنالك العديد من الخيارات الأخرى المتاحة إضافة لمقياس المسافة الاقليدية .

تتكون بيانات التحليل العنقودي في العادة من مصفوفة بحجم $n \times p$ ، وبصيغة الرموز فإن مصفوفة البيانات تتألف من p من المتغيرات $(X_1, X_2, X_3, \dots, X_p)$ تمثل اعمدة المصفوفة و n من المشاهدات ، التي تمثل صفوف هذه المصفوفة (جدول 22.4). وتحتسب دالة المسافة الاقليدية Ecludain بالمعادلة (21.4) وكالاتي :

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (21.4)$$

حيث ان X_{ik} هي القيمة الخاصة بالمتغير X_k للملاحظة i وان X_{jk} هي القيمة الخاصة بالمتغير X_k نفسه ولكن للملاحظة j .

جدول 22.4 : بيانات من p من المتغيرات و n من المشاهدات

X_1	X_2	X_3	 X_k	 X_p
X_{11}	X_{12}	X_{13}	 X_{1k}	 X_{1p}
X_{21}	X_{22}	X_{23}	 X_{2k}	 X_{2p}
X_{31}	X_{32}	X_{33}	 X_{3k}	 X_{3p}
.				
.				
X_{i1}	X_{i2}	X_{i3}	 X_{ik}	 X_{ip}
.				
.				
X_{j1}	X_{j2}	X_{j3}	 X_{jk}	 X_{jp}
.				
.				
X_{n1}	X_{n2}	X_{n3}	 X_{nk}	 X_{np}

يتم تغيير البيانات (تحويل البيانات الأصلية إلى قيم معيارية) قبل احتساب مقياس المسافة . بذلك تصبح جميع المتغيرات التي عددها p متساوية في الأهمية في تحديد (احتساب) مقياس المسافة. يتم ذلك بتحويل البيانات لتمتلك (البيانات) متوسط قيمته صفر وتباين مقداره واحد.

مثال (41.4) : إذا كانت $p=2$ (متغيرين) و $n=3$ (ثلاثة مشاهدات) وبياناتها مدرجة في الجدول (23.4) الآتي :

جدول (23.4): قيم لمتغيرين وثلاثة مشاهدات.

المتغير الأول X_1	المتغير الثاني X_2	التسلسل
42	76	1
55	80	2
35	90	3

المطلوب: احتساب المسافات بين المشاهدات

الحل :

الخطوة الأولى : تحويل قيم المتغيرين X_1 و X_2 الى قيم معيارية ، بعبارة أخرى ، استخدام صيغة Z-score للحصول على قيم معيارية للمتغيرين X_1 و X_2 وكالاتي :

أ- احتساب متوسط قيم المتغيرين X_1 و X_2 وكالاتي :

$$\bar{X}_1 = \frac{\sum_{i=1}^3 X_{1i}}{n_1} = \frac{42 + 55 + 35}{3} = 44$$

$$\bar{X}_2 = \frac{\sum_{i=1}^3 X_{2i}}{n_2} = \frac{76 + 80 + 90}{3} = 82$$

ب- احتساب الانحراف المعياري للمتغيرين X_1 و X_2 وكالاتي :

$$S_1 = \sqrt{\frac{\sum_{i=1}^3 (X_{1i} - \bar{X}_1)^2}{n_1 - 1}} = \sqrt{\frac{(42 - 44)^2 + (55 - 44)^2 + (35 - 44)^2}{2}} = 10.149$$

$$S_2 = \sqrt{\frac{\sum_{i=1}^3 (X_{2i} - \bar{X}_2)^2}{n_2 - 1}} = \sqrt{\frac{(76 - 82)^2 + (80 - 82)^2 + (90 - 82)^2}{2}} = 7.211$$

ج- احتساب الدرجات المعيارية للمتغيرين X_1 و X_2 وكالاتي :

$$Z_{11} = \frac{42 - 44}{10.149} = -0.197$$

$$Z_{12} = \frac{55 - 44}{10.149} = 1.084$$

$$Z_{13} = \frac{35 - 44}{10.149} = -0.887$$

$$Z_{21} = \frac{76 - 82}{7.211} = -0.832$$

$$Z_{22} = \frac{80 - 82}{7.211} = -0.277$$

$$Z_{23} = \frac{90 - 82}{7.211} = 1.109$$

وبإدراج قيم كل من Z_1 و Z_2 في الجدول (24.4) وكالاتي:

جدول (24.4): قيم كل من Z_1 و Z_2

المشاهدات	Z_1	Z_2
1	-0.197	-0.832
2	1.084	-0.277
3	-0.887	1.109

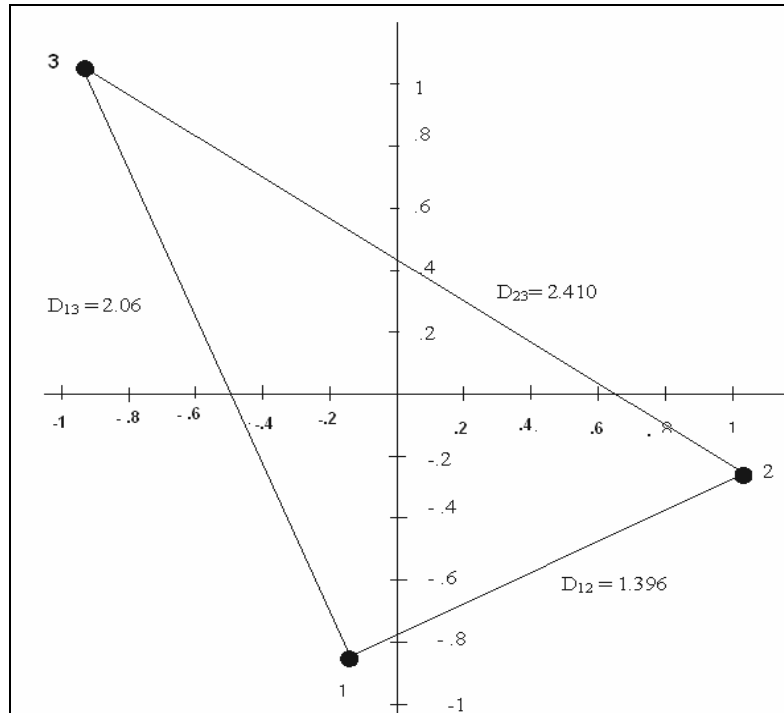
الخطوة الثانية : احتساب المسافات وكالاتي

$$D_{12} = \sqrt{(-0.197 - 1.084)^2 + (-0.832 + 0.277)^2} = 1.396$$

$$D_{13} = \sqrt{(-0.197 + 0.887)^2 + (-0.832 - 1.109)^2} = 2.06$$

$$D_{23} = \sqrt{(1.084 + 0.887)^2 + (-0.277 - 1.109)^2} = 2.410$$

إن اقصر مسافة هي بين النقطتين الأولى والثانية ومقدارها 1.396 في حين إن أطول مسافة كانت بين النقطتين الثانية والثالثة ومقدارها 2.410 (شكل 38.4)



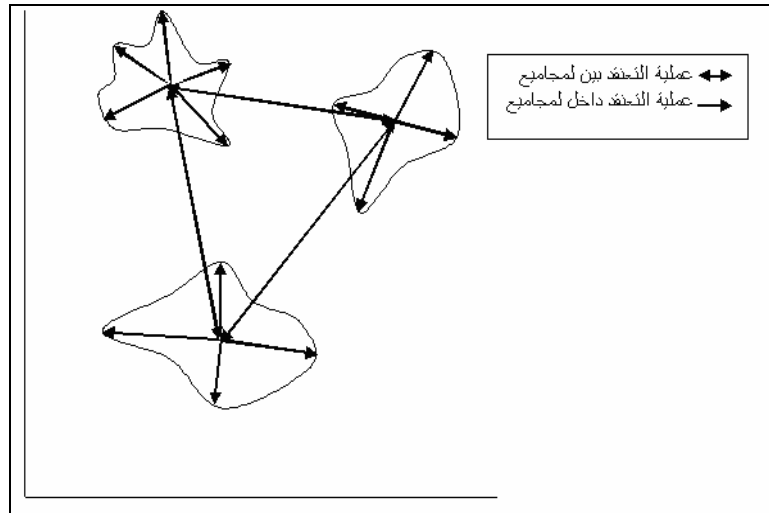
شكل 38.4 : المسافات الثلاثة . المسافة بين النقطتين الأولى والثانية (D_{12}) والمسافة بين النقطتين الأولى والثالثة (D_{13}) والمسافة بين النقطتين الثانية والثالثة (D_{23})

2- خوارزميات العنقدة Clustering algorithms

هنالك العديد من الخوارزميات التي تبحث في إجابة السؤال عن ماهية الطريقة التي يمكن استخدامها لوضع المشاهدات المتمثلة في مجاميع أو عناقيد مختلفة ؟ بعبارة أخرى ، ما هي خوارزمية العنقدة الأكثر ملائمة لوضع المشاهدات المتمثلة في مجاميع أو عناقيد مختلفة؟ للإجابة عن هذا السؤال هناك المئات من برامج الحاسوب التي تستخدم خوارزميات مختلفة متاحة والبعض منها في طور التطوير ، وبعد المعيار الأساسي لجميع هذه الخوارزميات ، هو محاولة تعظيم الاختلافات بين العناقيد استنادا إلى الاختلافات الموجودة (الحاصلة) بين المشاهدات . انظر شكل (39.4).

إن أغلب خوارزميات العنقدة الشائعة الاستخدام يمكن أن تصنف إلى صنفين رئيسيين هما :

- العنقدة الهرمية Hierarchical
- العنقدة اللاهرمية Nonhierarchical



شكل 39.4 : مخطط بياني يوضح العنقدة بين وداخل المجاميع

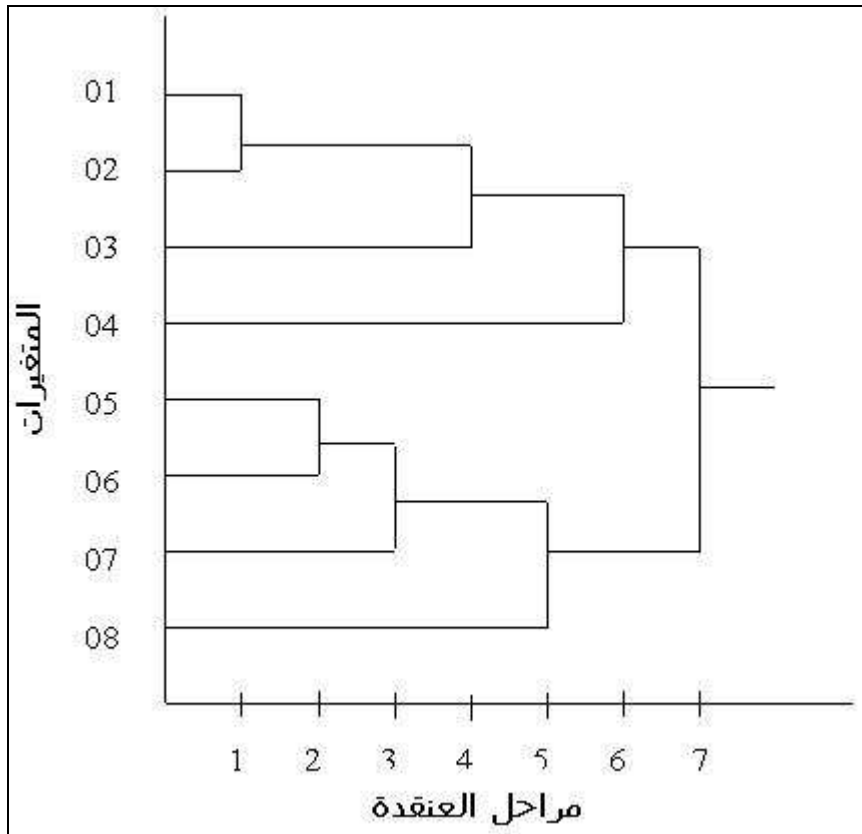
عمليات العنقدة الهرمية

تتضمن عملية العنقدة الهرمية ، بناء هرم أو بناء شكل (هيكل) يشبه الشجرة. هنالك نوعين من عمليات العنقدة الهرمي هما التكتل والانقسام (الانقسام). في طريقة التكتل ، تبدأ العملية بان جميع المشاهدات أو الأشياء تكون منفردة في عناقيد خاصة بها ، بعبارة أخرى فإن كل عنقود يحتوي على مفردة أو مشاهدة واحدة فقط . في الخطوات اللاحقة تتحد (تتجمع) اقرب مشاهدتين (مفردتين) في عنقود واحد ، نتيجة لذلك يقلص عدد العناقيد ويكون النقلص بعدد واحد في كل خطوة من خطوات العنقدة ، وصولا إلى الخطوة الأخيرة التي تعمل على ضم المفردة أو المشاهدة الأخيرة كما موضح في الشكل (40.4).

إما في حالة عملية الانقسام ، فإن العملية تتجه باتجاه معاكس لطريقة التكتل، عندئذ يطلق على هذه العملية (المعاكسة بالاتجاه) بعملية الانقسام (التجزئة) حيث تبدأ العملية بعنقود كبير يضم

جميع المشاهدات ، ثم بعد ذلك ، في الخطوات اللاحقة تبدأ المشاهدات (العناصر) الأكثر اختلافا بالانفصال عن العنقود الرئيس وتتضم إلى عنقود اصغر (في الحجم) آخر وتستمر هذه العملية لحين الوصول إلى المرحلة التي تصبح فيها كل مشاهدة (شيء) في عنقود منفرد خاص به. في شكل (40.4) وعند تطبيق طريقة التكتل فان اتجاه عملية التكتل يكون من اليسار إلى اليمين ، في حين طريقة الانفصام أو الانفصال يكون اتجاه سير العملية من اليمين إلى اليسار . إن الكثير من حزم البرامج تستخدم طريقة التكتل ، التي تمتلك عدد من القواعد هي :

- 1- الترابط الوحيد
- 2- الترابط التام
- 3- معدل الترابط
- 4- طريقة Ward
- 5- طريقة المركز



شكل 40.4 : التركيب الهرمي لعملية العنقدة يمثل اتجاه عملية العنقدة من اليسار إلى اليمين (من المرحلة رقم "1" إلى المرحلة رقم "7") طريقة التكتل. في حين اتجاه عملية العنقدة من اليمين إلى اليسار من المرحلة رقم "7" إلى المرحلة رقم "1" (تمثل طريقة الانفصام

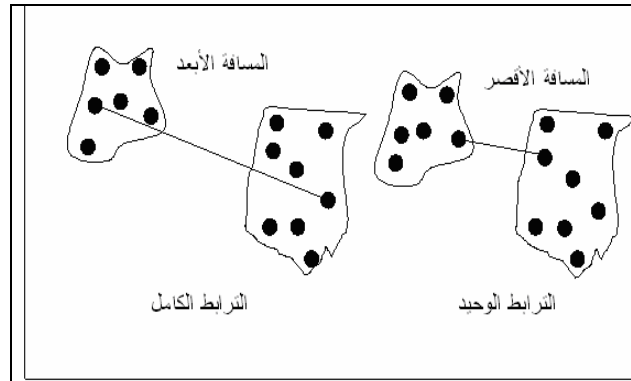
الترابط الوحيد Single linkage : يعتمد أسلوب الترابط الوحيد على المسافة الأقل. حيث يتم البحث عن اقصر (أقل) مسافة بين مشاهدين ووضعها في أول عنقود. ثم يتم البحث بعد ذلك عن اقصر مسافة تالية (بعد المسافة الأولى) ، فإما أن يتم إضافة المشاهدة إلى العنقود الأول ليتشكل عنقود يحتوي على مشاهدات جديدة (وهو الاحتمال الأول) أو (وهو الاحتمال الثاني) إن مشاهدين (آخرتين) يجتمعان في عنقود جديد. حيث تتعقد في المرحلة (1) (انظر شكل 40.4) المشاهدين (1 و 2) ليتكون بذلك أول عنقود يحتوي على مشاهدين فقط في هذه المرحلة ، ثم تتعقد بعد ذلك (في المرحلة رقم 2) المشاهدين (5 و 6) ليتكون بذلك العنقود الثاني والذي يحتوي على مشاهدين فقط، وفي المرحلة التالية (مرحلة رقم 3) تتعقد المشاهدة (7) مع كل من المشاهدين (5 و 6) ضمن العنقود الثاني ، بعد ذلك (في المرحلة رقم 4) تتعقد المشاهدة (3) مع المشاهدين (1 و 2) ضمن العنقود (المجموعة) الأول ، وفي المرحلة التالية (المرحلة رقم 5) تتعقد المشاهدة (8) ضمن العنقود الثاني ، في حين تتعقد المشاهدة (4) ضمن العنقود الأول في المرحلة التالية (المرحلة رقم 6) ، وأخيرا (في المرحلة رقم 7) يتوحد كلا العنقودين ليشكلان مجموعة واحدة.

الجوار الأقرب Nearest – neighbor approach

إن المسافة بين أي عنقودين هي عبارة عن اقصر مسافة بين أي نقطة في العنقود الأول إلى أي نقطة في العنقود الثاني ، ويدمج عنقودين عند أية مرحلة بواسطة المسافة الأقصر التي تربط بينهما. ويمكن أن تحدث المشاكل عندما لا يتم تحديد العناصر بصورة دقيقة.

الترابط الكامل Complete linkage

تشابه طريقة الترابط الكامل طريقة الترابط الوحيد باستثناء إن معيار العنقدة يعتمد على أكبر (أعظم) مسافة . لهذا السبب تسمى أيضا بطريقة الجوار الأبعد Furthest– neighbor approach إن سبب تسمية طريقة الترابط الكامل يعود إلى إن جميع المشاهدات في العنقود تتربط مع بعضها البعض بالاعتماد على ابعد (أطول) مسافة أو التماثل الاقل . يوضح شكل (41.4) المسافة الأقصر بين المشاهدات والمسافة الأبعد بين المشاهدات والتي لا تمثل التماثل الحقيقي بين المشاهدات.



شكل 41.4 : مقارنة بين مقياس للمسافات للترابط الوحيد والترابط الكامل

إن استخدام المسافة الأقصر بين المشاهدات (في عملية العنقدة) تدل على أن المجموعتين متماثلتين جداً ، في حين أن استخدام المسافة الأبعد (الأطول) بين المشاهدات (في عملية العنقدة) تدل على أن المجموعتين غير متماثلتين .

معدل الترابط Average linkage

تبدأ عملية العنقدة (في هذه الطريقة) بأسلوب طريقتي الترابط الوحيد والترابط الكامل نفسها ، إلا أن معيار العنقدة يعتمد على معدل المسافة بين المشاهدات في أحد العناقيد مع تلك المشاهدات الموجودة في عنقود آخر. في مثل هذا الأسلوب لا يتم استخدام القيم المتطرفة ، كما في الترابط الوحيد أو الترابط الكامل ، وأن عملية العنقدة تعتمد على جميع المشاهدات إضافة إلى الترابط بين زوج من المشاهدات (شكل 42.4) . تعمل طريقة معدل الترابط على تكوين العناقيد ذات التباينات (الاختلافات بين المشاهدات) القليلة ، إذ يحسب معدل الترابط AL كالآتي :

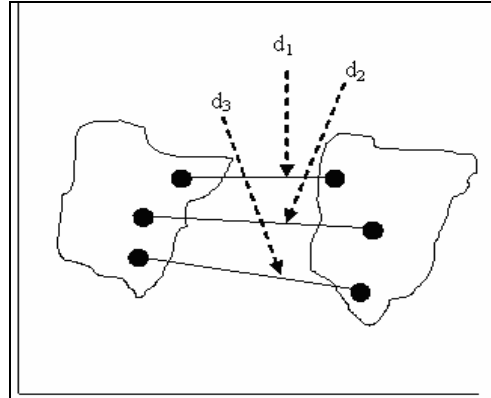
$$AL = \frac{d_1 + d_2 + d_3}{3}$$

حيث أن AL تمثل معدل الترابط (معدل الطول) بين المشاهدات.

d_1 : المسافة الأولى.

d_2 : المسافة الثانية.

d_3 : المسافة الثالثة.



شكل 42.4 معدل الترابط

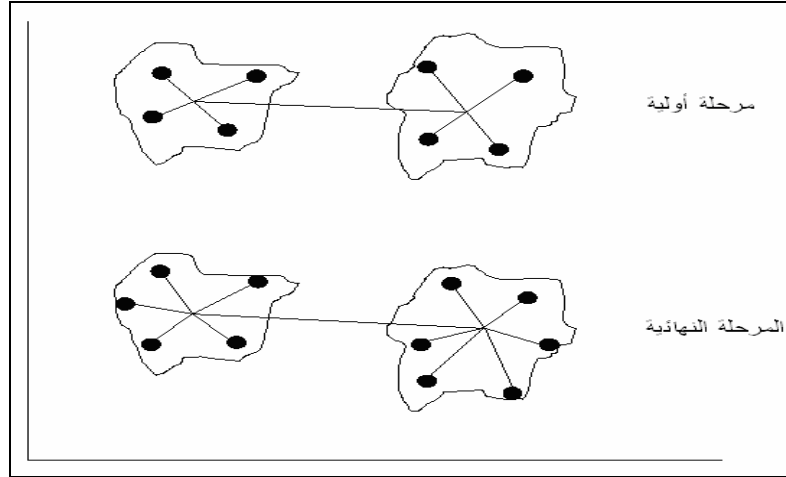
طريقة وورد Ward method

تعتمد هذه الطريقة في احتساب المسافة بين عنقودين على مجموع المربعات بين العنقودين مقسمة على عدد المتغيرات . تعمل هذه الطريقة على تكوين العناقيد بأقل عدد من المشاهدات .

طريقة المركز Centroid method

تعتمد هذه الطريقة على احتساب المسافة بين معدل العنقودين ، وفي كل مرحلة من مراحل العنقدة يتم احتساب معدل جديد للمشاهدات (في داخل العنقود) وأن هذا المعدل يتغير في كل

مرحلة (عند إضافة مشاهدة إلى العنقود) من مراحل العنقدة. (شكل 43.4) . تكمن فائدة هذه الطريقة في كونها أقل الطرق تأثراً بالقيم الشاذة (المتطرفة) مقارنة مع باقي الطرق الهرمية.



شكل 43.4 : مراحل مختلفة من عملية العنقدة باستخدام طريقة المركز

عمليات العنقدة اللاهرمية (غير الهرمية) Nonhierarchical clustering procedures تعمل هذه الطريقة بأسلوب مغاير (على النقيض من) لطريقة العنقدة الهرمية. حيث إنها لا تتضمن عملية بناء تركيب الشجرة ، وبدلاً من ذلك ، فإن عملية العنقدة فيها تبدأ باختيار مركزاً "العنقود (أو البذرة) ، ومن ثم تحدد جميع المشاهدات التي تقع ضمن مسافة (مجال) معينة (سبق وأن تم تحديدها كمعيار في عملية العنقدة) في العنقود. تعرف طريقة العنقدة اللاهرمية بطريقة العنقدة بعدد K من الأوساط (K-means clustering) . تستخدم طريقة العنقدة اللاهرمية واحدة من الطرق الثلاثة الآتية :

أولاً- طريقة المعيار التسلسلي Sequential threshold procedure، التي تبدأ باختيار بذرة لأحد العناقيد ومن ثم تحدد جميع المشاهدات التي تقع ضمن مسافة (المجال) معينة (سبق وأن تم تحديدها كمعيار لعملية العنقدة) في العنقود الأول ، ثم يتم اختيار بذرة جديدة أو بذرة أخرى للعنقود الثاني ، ومن ثم تحدد جميع المشاهدات التي تقع ضمن مسافة (مجال) معينة في العنقود الثاني ، ومن ثم تحدد بذرة (جديدة أو أخرى) للعنقود الثالث وهكذا تستمر العملية لحين عنقدة جميع المشاهدات في داخل العناقيد ولا يوجد هناك إمكانية تحديد بذرة (جديدة) لأي عنقود إضافي.

ثانياً- طريقة المعيار المتوازي Parallel threshold procedure، تعمل هذه الطريقة بأسلوب مغاير لطريقة المعيار التسلسلي ، وفيها يتم تحديد عدة بذور للعناقيد المختلفة حيث يكون واحد عند بداية عملية العنقدة ومن ثم يتم تحديد المشاهدات التي تقع ضمن المسافة (المجال) الأقرب للبذرة ضمن العنقود المحدد (بواسطة معيار المسافة هذه). في بعض هذه الطرق يمكن أن تبقى بعض المشاهدات غير منعقدة (خارج العناقيد) إذا كانت خارج المسافة أو المسافات المحددة كمعيار للعنقدة.

ثالثاً- طريقة المعيار الأمثل Optimizing threshold procedure، تعد الطريقة مثلى والتي تماثل الطريقتين السابقتين فيما عدا إنها تسمح بإعادة تحديد المشاهدات إلى العناقيد الأخرى استناداً إلى معيار أمثل عام .

يمكن الآن طرح التساؤل الآتي : أي الطريقتين يجب استخدامها ، طريقة العنقدة الهرمية أم طريقة العنقدة اللاهرمية ؟

إن لكلا الطريقتين مزايا وخصائص تختلف عن بعضها البعض، في السابق كانت تقنيات العنقدة الهرمية أكثر شهرة ، إذ تمتاز بكونها سريعة إضافة إلى ذلك أنها تحتاج إلى زمن أقل للعنقدة. لكنها يمكن أن تضلل بسبب حدوث توليفات غير دقيقة ممكن أن تؤدي إلى نتائج مصطنعة (كاذبة) . ولغرض تقليص حدوث هذا الاحتمال يمكن للباحث أن يستبعد القيم الشاذة (غير المرغوب فيها) قبل إجراء عملية العنقدة الهرمية.

في حين تبدو طريقة العنقدة اللاهرمية أكثر تقبلاً . فإذا كانت هناك إمكانية في تطوير عملية اختيار البذرة (البذرات) يمكن أن تصبح هذه الطريقة أكثر استخداماً ، إذ يمكن استخدامها في حالة كون حجم البيانات كبير.

3-تحديد عدد العناقيد

إن المسألة (المشكلة) الرئيسة التي تواجه جميع طرق العنقدة هو كيفية اختيار عدد العناقيد؟ هناك العديد من المعايير لتحديد هذا العدد ، ولكن لا يوجد معيار قياسي خاص بها. يعد معيار المسافة بين العناقيد الذي يتم تحديده في كل مرحلة من المراحل معياراً مناسباً وان يحدد الباحث (المحلل) خيار التوقف عندما تزداد المسافة عن حد (مقدار) معين . أو يمكن (من خلال تجربة سابقة) تحديد عدد من العناقيد على سبيل المثال كأن تكون مجموعتين أو ثلاثة أو أربعة أو أن يتم تحديد مدى من العناقيد على سبيل المثال كأن تكون بين ثلاثة إلى ست عناقيد. ومن ثم إجراء عملية العنقدة واختيار أفضل بديل بعد أن يتم تقييم جميع الاحتمالات الممكنة.

المرحلة الثانية : التفسير

تتضمن مرحلة التفسير تحديد أسماء أو عناوين تصف طبيعة العناقيد. وكمثال على هذه المرحلة ، عنقدة بيانات في دراسة شملت خمسة متغيرات : العمر ، فصيلة الدم ، حالة التدخين ، الجنس ، نوع الإصابة بمرض ما، وقد أخذت البيانات لخمسين شخص (مشاهدة) ، وبعد إجراء عملية العنقدة تجمعت البيانات في ثلاثة مجاميع. يطلق على المجموعة الأولى بمجموعة الأشخاص غير البالغين والمجموعة الثانية مجموعة الأشخاص البالغين والمجموعة الثالثة والأخيرة مجموعة كبار السن والشيوخ ، وبهذا فقد تم تحديد أسماء أو عناوين تصف طبيعة العناقيد الثلاثة .

المرحلة الثالثة : المظهر الجانبي تحديد النسق Profiling

تتضمن مرحلة تحديد المظهر الجانبي وصف ميزات وصفات كل عنقود من العناقيد لغرض تفسير الاختلافات الحاصلة بين المجاميع .

فعلى سبيل المثال (في المثال السابق) إن المجموعة الأولى تمتاز بان الإصابة بالمرض تشمل الأشخاص غير البالغين من الذكور الذين يحملون فصيلة دم O من المدخنين بهذا المرض ، في حين إن المجموعة الثانية (مجموعة الأشخاص البالغين) ليسوا مصابين بهذا النوع من المرض

على اختلاف أصناف الدم وللإناث والذكور والمدخنين أو غير المدخنين. أما المجموعة الثالثة (مجموعة كبار السن) تمتاز بأن الإصابة تشمل الأشخاص كبار السن من الإناث الذين يحملون صنف دم B من المدخنين بهذا المرض.

الفصل الخامس

تحليل البيانات Data analysis

تطبيق عمليات التحليل
تحديد تشابه التواليات
أداة بحث الاصطفاف الموضوعي الأساسي BLAST
برنامج FASTA
اصطفاف التراكيب
العلاقات التطورية
المعلوماتية الحيوية والجينوم
الخيطة والخيطة البديلة Splicing and alternative splicing
مشاكل ومثالب حدس الجينات
إيجاد جينات البكتريا
الوراثة اللاجينية Epigenetics
تقنيات تضخيم DNA والمعلوماتية الحيوية
الجوانب التطبيقية الأخرى للمعلوماتية الحيوية في مجال الجينومات
خرائط وأطالس الجينومات
المعلومات الحيوية والبروتينات
المعلوماتية الحيوية و RNA

تحليل البيانات Data analysis

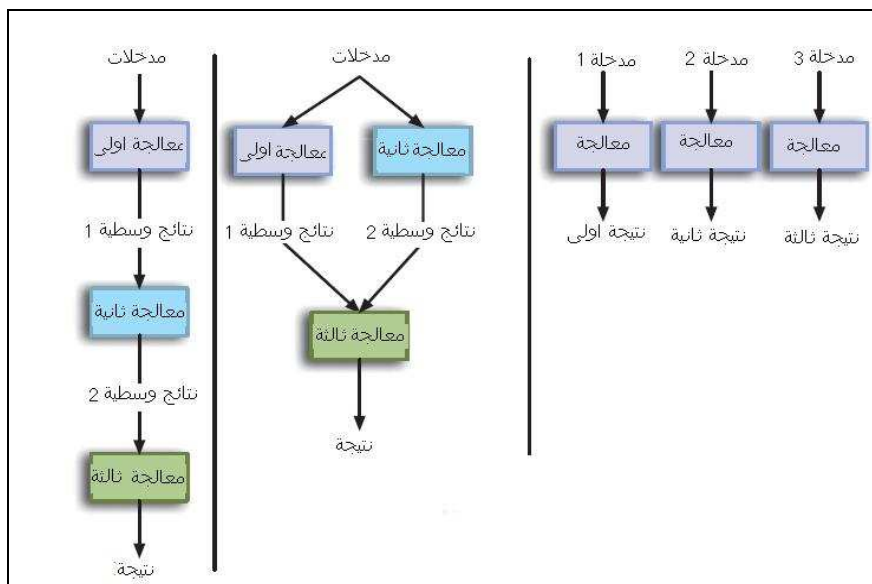
تحليل البيانات هو التطبيق الحقيقي الذي يحدث في مجالات المعلوماتية الحيوية ولكن قبل البدء بهذا الموضوع لابد من توضيح وربما إعادة تأكيد بعض الحقائق الحيوية التي يمكن ان تشوش على تحليل البيانات وربما في بعض الأحيان الوصول الى استنتاجات خاطئة.

بدءاً فالأنظمة الحيوية أنظمة معقدة جداً وترتبط الجزيئات والتراكيب الخلوية فيما بينها بشبكات معقدة جداً، قد يقسمها البعض الى شبكات عامة Global networks وأخرى شبكات صغيرة او ثانوية Minor networks ولكن هذا لا يحل المشكلة للفهم فهي لا تزال مرتبطة مع بعضها، وتزداد الأمور تعقيداً بان الأنظمة الحية او الخلايا لا تكون بمعزل عن المحيط الخارجي اي ان الخلايا تخضع للقوانين الكيماوية والفيزيائية على حد سواء، وبذا يمكن القول ان الشبكات الحيوية Bionetworks هي الأعداء على وجه البسيطة، ولكن ما لا يدرك كله لا يترك جله ومن هنا نشأت محاولات الدراسة لفهم هذه الأنظمة.

المشكلة الأخرى والناجمة عن تعقيد الأنظمة الحيوية هي ان المواد الوراثية او الطبقة الحاكمة من مستويات الخلية تتمثل بأربعة الى خمسة حروف هي القواعد النروجينية A,C,G,T,U ، في حين ان التطبيقات في المستوى الثاني يكون 20 الى 21 حامض أميني، كما ان القضية المحورية التي افترضت في بدايات دراسة علم الحياة الجزيئي والمتمثلة بان مكون الجينوم او DNA ينتسخ ليكون مكون النسخ المتمثل بـ RNA يحول الى مكون بروتيني أصبحت لا تلاؤم ما معروف الآن حول الأنظمة الحيوية وتدخل مؤثرات الوراثة اللاجينية او الوراثة الخارجية ، كما ان هناك ما يقطع تسلسل القضية المحورية وهو التوقف عند إنتاج RNA وعدم تحوله الى بروتينات كما في التواليات غير المشفرة ونظراً لكون القواعد النروجينية هي في حدود 4-5 والحوامض الامينية 20-21 فضلاً عن الاختلاف في الشفرات الوراثة وتفضيلها من قبل الأحياء وانحراف البعض منها عن السياق العام مثل إنتاج السستئين السيليني من شفرة وقف كل هذه الجوانب صعبت من وضع برامج سهلة ومباشرة لعمليات التحليل، وكذلك انقسام الحوامض الامينية الى مجاميع مثل القطبية والحامضية والكارهة للماء وغيرها وإمكانية إحلال بعض الثمالات مكان بعض من المجموعة نفسها دون تأثير كبير في وظائف البروتينات الناتجة.

تطبيق عمليات التحليل

معظم التحليلات هي سبل تنفيذية Execution pipeline لمعلومات يقوم المستخدم بإعدادها او الحصول عليها من مراجع أكاديمية او مصادر تجارية واختيار طريقة التحليل تؤثر بشكل كبير في نوعية النتائج التي يتم الحصول عليها. وتتمثل عملية تنفيذ طرق التحليل بشبكات متداخلة الخطوات، في كل خطوة يتم أداء جزء من المهمة مثل إجراء الحسابات على البيانات المدخلة وغيرها من العمليات. والشكل التالي (1.5) يوضح بعض الطرق التنفيذية المستعملة:



شكل 1.5 : طرق تنفيذ تحليل البيانات

فيلاحظ ان عمليات المعالجة تكون متسلسلة في الجزء الأيسر وفيها ان الخطوة الثانية (مثلا) لا يمكن ان تبدأ إلا بعد اكتمال الخطوة الأولى لتكون معطياتها مجالا لتطبيق الخطوة الثانية ... وهكذا الحال مع الخطوة الثالثة ، في حين ان الطريقة الثانية (الجزء الأوسط) يلاحظ ان العمليات تسير بشكل متزامن في البداية لتعطي معطيات للمرحلة الأخيرة ثم إعطاء النتائج. ويستخدم الحاسوب في تنفيذ كل عملية وهذا التزامن يمكن ان يؤدي الى بطيء الحاسبة والتنفيذ. اما النوع الثالث فتكون المدخلات فيه متعددة (الجزء الأيمن) وتسير العمليات بشكل متواز وهي الطريقة الأكثر استعمالا في بعض مجالات المعلوماتية الحيوية كما في البحث عن الموائمة Matching للبروتينات او الجينات في قواعد المعلومات عند استعمال برنامج BLAST.

تحديد تشابه التواليات

نظرا للتعقيدات المذكورة أعلاه ، واستناد عمليات التنبؤ بالجزئيات الحيوية في اغلب الأحيان على عمليات إيجاد التشابه، وفي هذه الحالة ولكبر كمية المعلومات المشمولة بالدراسة لابد من تدخل طرق الإحصاء ووضع المصفوفات لغرض ترجيح صحة النتائج والتي تبقى على هامش ولو قليل من الشك نظرا لتعقيد الأنظمة الحيوية.

الاصطفاف Alignment

يمثل اصطفاف زوج من التواليات Pairwise sequence alignment أبسط حالاتها. وتعد مقارنة التواليات أحد أهم التحاليل المستخدمة في المعلوماتية الحيوية، فهي الخطوة الأولى في اتجاه تحليل الهيكل والوظيفة للتواليات الجديدة . وفيها يتم البحث عن رموز ثملات الحوامض الامينية المتماثلة في التواليات ذات العلاقة او رموز قواعد DNA . وتعد عملية اصطفاف زوج من التواليات الأساس في البحث في قواعد البيانات عن التشابه للتوالي Similarity ، وكذلك الأساس في عمليات الاصطفاف لعدة تواليات Multiple sequence alignment. وتوفر عملية اصطفاف التواليات إمكانية الاستدلال على العلاقة بين التواليتين تحت الدراسة. فاذا

أشترك التواليان بدرجة معنوية من التشابه أو التماثل Identity فانه من المستبعد للغاية ان يكون التطابق عشوائياً، وهذا يعني ان التواليين يمكن ان يكونا منحدرين من أصل تطوري مشترك (التماثل الجذري) Homologous ، ولكن يكون هناك احتمال لعدم التطابق في بعض المواضع ناتجا عن عمليات استبدال الحوامض الامينية (في حالة البروتينات) وهنا يمكن ترجيح احتمال انحدار التواليان من أصل مشترك ولكن تباعدا بحيث أصبحت العلاقات الجذرية غير قابلة للتمييز على مستوى التوالي، وفي هذه الحالة فان الكشف عن المسافة الخاصة بالعلاقات التطورية يجب ان يتم باستخدام طرق أخرى سيتم التطرق اليها في موقع آخر.

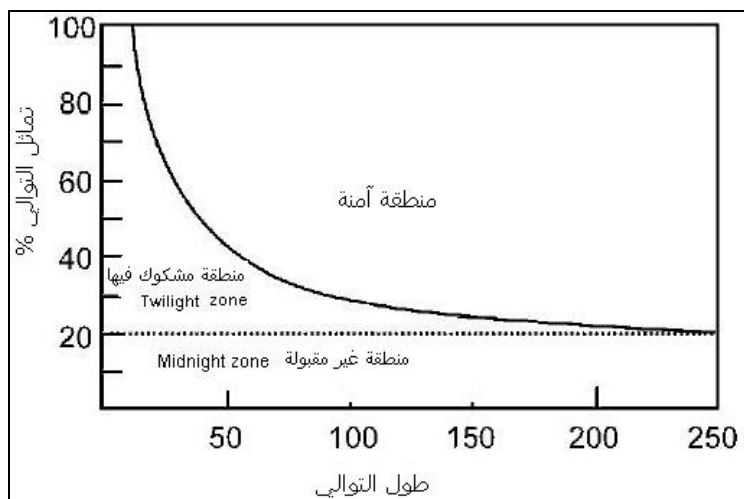
التماثل الجذري والتشابه والتماثل للتواليات

يعد التماثل الجذري للتواليات Sequence homology أحد الأسس المهمة في تحليل التواليات، فعندما ينحدر تواليان من أصل مشترك فيمكن القول إنهما يمتلكان تماثلاً "جذرياً" أو يشتركان في جد مشترك، في المقابل يمثل مصطلح تشابه التوالي نسبة اصطفاف وموائمة ثملات الحوامض الامينية التي هي متشابهة في الخصائص الفيزيوكيميائية مثل الحجم والشحنة وكرهها للماء، لذا كان من المهم التمييز بين مصطلح التماثل الجذري والمصطلح المتصل به وهو تشابه التوالي Sequence similarity وكذلك تماثل التوالي Sequence identity اي ان الثملات تكون هي نفسها في التواليين وليس من المجموعة، والحقيقة ان هناك خلط في استعمال هذه المصطلحات.

فالتماثل الجذري يستدل به من علاقة الأسلاف المشتركة المأخوذة من مقارنة التواليات ويكون التواليان بدرجة عالية من التشابه. اما مصطلح التشابه فهو نتيجة مباشرة من المشاهدات الناتجة من اصطفاف التواليات. ويمكن تحديد كمية تشابه التواليات باستخدام النسب المئوية في حين ان التماثل الجذري للتواليات هو حالة نوعية، فعلى سبيل المثال يمكن القول ان التواليين يشتركان بنسبة 40% من التشابه ومن غير الصحيح القول ان التواليين يشتركان بنسبة 40% من التماثل الجذري فهما اما ان يكون هنالك تماثل جذري بين التواليين او لا يكون. وعلى العموم اذا كان مستوى تشابه التواليان عالياً بما فيه الكفاية يمكن ان يستنتج منه ان هناك علاقة تطورية مشتركة، ولو انه في بعض الأحيان تكون العلاقة غير واضحة دائماً، ويعتمد الجواب على نوع التوالي تحت الدراسة وعلى طول التواليات.

وما ذكر آنفاً يمكن ان يوضح ان تواليات النيوكلويتيدات تتألف من 4 أحرف (وهي عدد القواعد النتروجينية) ومن ثم فان التواليات التي ليست بينها صلة تمتلك احتمالية فرصة التطابق بنسبة 25% ($0.25 = 1/4$) على الأقل نتيجة للمطابقة العشوائية، في حين تمتلك تواليات البروتين المكونة من عشرين حامض أميني التي ليست بينها علاقة احتمالية تطابق يصل الى 5% ($0.05 = 1/20$) من التطابق نتيجة للمصادفة العشوائية ، وفي حالة استخدام الفجوات Gaps فان النسبة تزداد لتصل الى 10-20%. ويعد طول التوالي عاملاً مهماً جداً، فالتواليات القصيرة تمتلك فرصة اكبر في التطابق الذي يعزى الى المصادفة العشوائية ولذلك وجب وضع حد فاصل Cutoff للتواليات القصيرة عند الاستدلال على علاقات التماثل الجذري مقارنة مع التواليات الطويلة.

فعلى سبيل المثال لو تم صف تواليين بطول 100 حامض أميني فان التشابه بنسبة 30% أو أعلى يدل على ان التواليين يمتلكان تشابه مقارب. وعند تمثيل الأهمية الإحصائية تظهر في الآتي:



شكل 2.5 : التمثيل الإحصائي لدرجات التشابه

فان مستوى التطابق بين 20-30% يؤشر الى ان علاقة التشابه تقع ضمن المنطقة غير المؤكدة والتي يطلق عليها منطقة الغروب أو الغبش Twilight zone اذ يتداخل فيها التشابه الضئيل مع تداخل التواليات نتيجة للعلاقات العشوائية. اما عندما تكون نسبة التشابه اقل من 20% فان نسبة عالية من التواليات تظهر بأنها تواليات لا علاقة بينها، وتظهر ممثلة في الشكل بما يطلق عليه منطقة الظلام أو غير مقبولة Midnight zone وهي نتائج لا يمكن الاعتماد عليها. ولكن هذا التمثيل لا يمكن اعتباره مقياساً دقيقاً لتحديد العلاقات بين التواليات خاصة في المنطقة غير المؤكدة (منطقة الغروب) ولذلك احتاجت هذه النتائج الى طرق إحصائية أكثر صرامة لتحديد علاقات التشابه لتحديد المعنوية الإحصائية والتي سيتم التطرق اليها لاحقاً.

يتم استخدام التماثل والتشابه بصورة مترادفة في التواليات النيوكليوتيدية ولكن هذه تكون مختلفة في توالي البروتينات ، اذ يشير تماثل التواليين الى النسبة المئوية لوجود الحوامض الامينية نفسها في التواليين، في حين يشير التشابه الى النسبة المئوية لاصطفاف الحوامض الامينية من الزمرة ذاتها اي التي تمتلك خصائص فيزيوكيماوية متشابهة والتي تكون أكثر استعداداً للاستبدال بعضها مع البعض الآخر دون تأثير كبير.

وهناك طريقتان لاحتساب التماثل والتشابه للتواليات تتضمن :

الأولى: استخدام أطوال كلا التواليين.

وبلغة الرموز تستخدم هذه الطريقة الصيغة الآتية:

$$S = [(L_s \times 2) / (L_a + L_b)] \times 100$$

حيث ان S تمثل النسبة المئوية لتشابه تواليان a , b

L_s عدد الأحماض الامينية المصطفة المتشابه الخصائص

L_a و L_b المجموع الكلي لأطوال لكلا التواليين على انفراد .

ويحتسب تشابه التوالي (I%) بطريقة مماثلة وكما يأتي:

$$I = [(L_i \times 2) / (L_a + L_b)] \times 100$$

حيث أن L_i هو عدد الأحماض الامينية المتماثلة.
الثانية: تعتمد على النسب المئوية لأعداد من الثمالات المتماثلة والمتشابهة منسوبة الى المجموع الكلي للثمالات في التوالي الأقصر.

تعتمد هذه الطريقة في الحساب النسبة المئوية لكل من تماثل وتشابه أعداد الأحماض الامينية على مجموع الأحماض الامينية للتوالي الأقصر وباستخدام الصيغة:

$$S = [(L_s \times 2) / (L_a + L_b)] \times 100$$

$$I(S)\% = L_{i(S)} / L_a \%$$

حيث أن L_a طول التوالي الأقصر من التوالين تحت الدراسة .

مثال : بافتراض التوالين التالية والمراد احتساب :

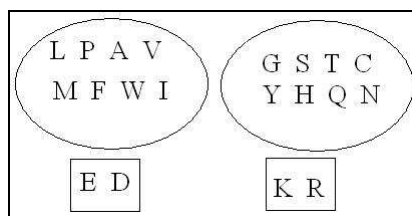
أ - التشابه Similarity لهما .

ب - التماثل Identity لهما .

(a) Seq 1: AAAVPCTLIMLIRSCGAAGCFTIL

(b) Seq 2: CRSCRHAPSRMLAACGAPV

وفق مواصفات الحوامض الامينية التي تم تقسيمها الى عدة مجاميع كما في المخطط الآتي:



أ - يمكن احتساب التشابه من المعادلة الآتية :

$$S = [(L_s \times 2) / (L_a + L_b)] \times 100$$

$$S = [(7 \times 2) / (20 + 25)] \times 100$$

$$S = 31.1\%$$

او باحتسابها وفق المعادلة القياسية وكالاتي :

$$S = (L_s / L_a) \times 100$$

$$S = (7 / 20) \times 100$$

$$S = 35\%$$

ب - يمكن احتساب التماثل Identity من المعادلة الآتية :

$$I = [(L_i \times 2) / (L_a + L_b)] \times 100$$

$$I = [(2 \times 2) / (20 + 25)] \times 100$$

$$= 8.9\%$$

او باستعمال المعادلة القياسية وكالاتي :

$$\begin{aligned} I &= (L_i / L_a) \times 100 \\ &= (2 / 20) \times 100 \\ &= 10\% \end{aligned}$$

مثال : بافتراض ان التوالين التالية والمراد حساب

أ - التشابه Similarity لهما .

ب - التماثل Identity لهما .

(a) Seq 1 :AAAAPCTLIMLIRSCGAAGCFTIL

(b) Seq 2: CRSCRHAPSRMLAACGAPVALMFPA

وفق المخطط أعلاه لمجاميع الحوامض الامينية يمكن

أ - حساب التشابه وفق المعادلة الآتية

$$\begin{aligned} S &= [(L_s \times 2) / (L_a + L_b)] \times 100 \\ S &= [(10 \times 2) / (25 + 25)] \times 100 \\ &= 40\% \end{aligned}$$

او حسابها وفق المعادلة القياسية الآتية :

$$\begin{aligned} S &= (L_s / L_a) \times 100 \\ &= (10/25) \times 100 \\ &= 40\% \end{aligned}$$

ب - يمكن حساب التماثل من المعادلة

$$\begin{aligned} I &= [(L_i \times 2) / (L_a + L_b)] \times 100 \\ &= [(2 \times 2) / (25 + 25)] \times 100 \\ &= 8\% \end{aligned}$$

او حسابها وفق المعادلة القياسية الآتية :

$$\begin{aligned} I &= (L_i / L_a) \times 100 \\ &= (2 / 25) \times 100 \\ &= 8\% \end{aligned}$$

طرق اصطفاف زوج من التواليات

الهدف العام من صف زوج من التواليات هو إيجاد أفضل زوج من التواليات يمتلكان اكبر عدد ممكن من تشابه الحوامض الامينية وتمثل إحدى التواليات توالي الاستعلام المراد البحث عن توالي يشبهه في قواعد البيانات Subject sequence والتي تكون معلومة التركيب والوظيفة. ولذلك يتم صف التوالين وتكون عملية الاصطفاف على نوعين:

- الاصطفاف الشامل Global alignment وهو الأكثر استعمالاً.
- الاصطفاف الموضعي Local alignment .

في الحالة الأولى يتم صف التواليتين الذي يفترض ان يكونا متساويين في الطول، وينفذ الاصطفاف من بداية التوالي الى نهايته لإيجاد أفضل تلاؤم ممكن لهما عبر الطول الكلي للتوالي. وهذه الطريقة أكثر تطبيقاً لصف تواليتين متقاربتين في العلاقة ومتساويان في الطول. وفي حالة التواليات المتباعدة او غير المتساوية في الطول فان الطريقة غير قادرة على إيجاد أفضل النتائج لانها لا تستطيع تمييز المناطق الموضعية ذات التشابه العالي بين التواليتين، وطريقة الصف الشامل موضحة في الشكل 3.5 :

Seq 1: AFRDF - NQYYSSIKRS GSIQ
Seq 2: PLKLFİ DQYYSSİKRT MG-H

شكل 3.5 :اصطفاف التوالي الشامل

ويشير (:) الى تماثل الثمالات في حين(.) يشير الى تشابه الثمالات اي ان الثمالات ليست نفسها ولكن من زمرة واحدة في الحوامض الامينية. اما في حالة الاصطفاف الموضعي فلا تحتاج التواليات ان تكون متساوية الطول ، والطريقة تعتمد الى إيجاد المناطق الموضعية التي على مستوى عالي من التماثل دون الأخذ بنظر الاعتبار اصطفاف بقية المناطق في التوالي، وتستخدم هذه الطريقة لصف التواليات غير متساوية الطول والطريقة موضحة في شكل 4.5 :

Seq 1: NQYYSSIKRS
Seq 2: DQYYSSİKRT

شكل 4.5 : اصطفاف التوالي الموضعي

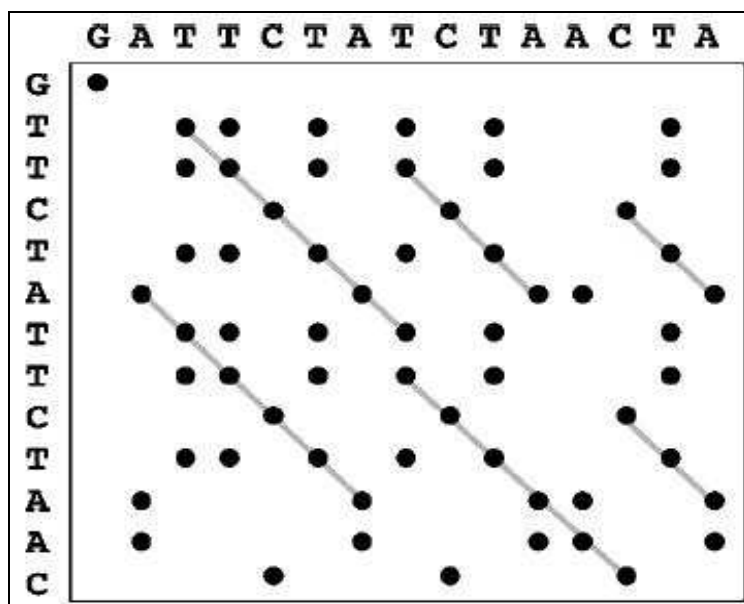
خوارزميات الاصطفاف Alignment algorithms

خوارزميات الاصطفاف الشامل والموضعي تكون متشابهة في الأساس وتختلف في إستراتيجية تحديد الامتلية في اصطفاف الحوامض الامينية المتماثلة، وكلا النوعين يمكن ان يعتمد على واحدة من الطرق الآتية:

- طريقة مصفوفة التنقيط Dot matrix method
- طريقة البرمجة الديناميكية Dynamic programming method
- طريقة الكلمة Word method

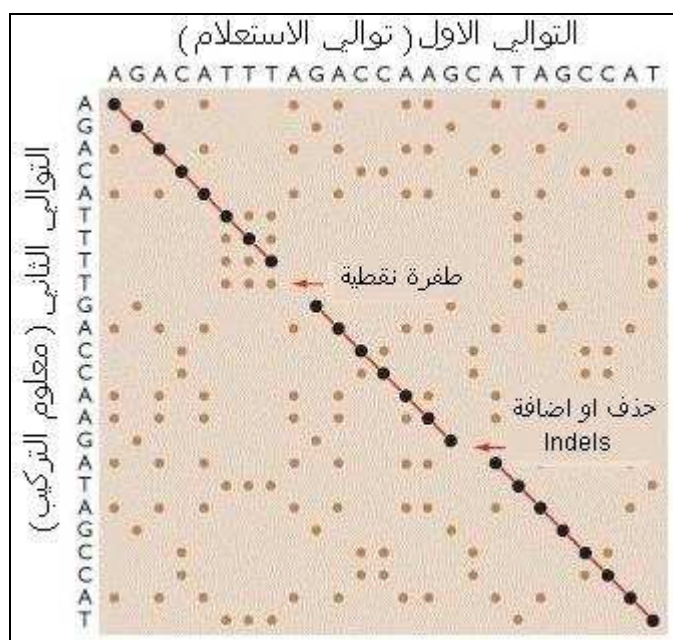
طريقة مصفوفة التنقيط : وتعد من الطرق المهمة في صف التواليات وتعرف ايضا بطريقة خارطة التنقيط Dot plot map. والطريقة تصويرية تستخدم لغرض المقارنة بين تواليتين في مصفوفة ذات بعدين، وفيها يكتب التواليان المراد المقارنة بينهما على المحورين السيني والصادي من المصفوفة. ويتم المسح لكل قاعدة لأحد التواليان مع ما يماثلها من القواعد في

التوالي الآخر، وعند وجود تطابق لأحد القواعد فإنه يحدد بنقطة على الخارطة أو المصفوفة، وفيما عدا ذلك يترك الموضع شاغراً، وعند تحديد مناطق التطابق لكلا التواليين تظهر خطوط من نقاط تصل بينها بشكل خطوط قطرية متجاورة والتي تشير إلى اصطفاف التواليات كما موضح في الشكل 5.5 :



شكل 5.5 : طريقة مصفوفة التنقيط للاصطفاف

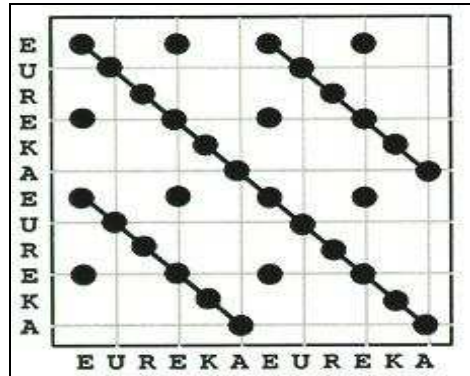
وطريقة مصفوفة التنقيط للاصطفاف الموضحة في الشكل (6.5) لمقارنة التواليين ، فتشير الخطوط على القطر الرئيس إلى اصطفاف التواليين أي تماثلها ، وتشير الخطوط فوق وتحت القطر الرئيس إلى التكرار لكلا التواليين ، وعند ظهور قطع في الخط القطري دل ذلك على وجود حالات إضافة أو حذف، أما الخطوط القطرية المتوازية داخل المصفوفة فتمثل المناطق المتكررة من التوالي. كما موضح في الشكل 6.5 :



شكل 6.5 : ظهور القطع وعدم التواصل نتيجة لحدوث الطفرات او عمليات الحذف او الإضافة

وتعاني طريقة التنقيط من التشويش والالتباس خاصة عند مقارنة تواليات طويلة او ما يسمى بـ High noise level في اغلب خرائط التنقيط، اذ ان النقاط تتوزع على جميع الخارطة تقريبا وبالتالي يمكن ان تحجب التشخيص الحقيقي للاصطفاف وتظهر هذه المشكلة بشكل كبير في حالة صف تواليات DNA وذلك لوجود 4 أحرف فقط وعليه فان كل ثمالة او قاعدة لها فرصة التلاؤم مع ثمالة أخرى في التوالي الثاني بنسبة 4:1 . وللتخلص من هذه الضوضاء يكون من الأفضل استخدام طريقة الترشيح او التصفية ويكون ذلك باستعمال نافذة Window مكونة من امتداد من الثمالات محددة الطول بدلا من استعمال ثمالة واحدة لعملية المسح والتحري عن التشابه. وعند تطبيق نظام التصفية هذا فان النافذة المتحركة او المنزلقة ستقوم بمسح التواليتين لمقارنة الثمالات وستحدد نقطة فقط عند تساوي المنطقة وتلاؤمها مع الامتداد او الحجم المحدد بالنافذة بشكل منضبط تماما وبذا تقلل من عدد النقاط الكبير الذي يمكن ان ينتج عن استعمال حرف او ثمالة واحدة، وبعض الأحيان تدعى هذه النافذة بـ Tuple، ويمكن التحكم بحجم النافذة للحصول على نمط واضح لتوائم التواليات وذلك لان زيادة عدد الحروف في هذه النافذة يؤدي الى تقليل الحساسية وبالتالي فقد العديد من التوائم في عملية الصف.

وتستخدم الطريقة لتحديد التكرار الداخلي للعناصر المكونة للتوالي نفسه وعندها يكون هناك قطر رئيس يمثل التطابق التام، وعند وجود المكررات مثلا في توالي ما يلاحظ ظهور خطوط قصيرة متوازية فوق وتحت الخط او القطر الرئيس موضحة في الشكل 7.5:



شكل 7.5 : ظهور المكررات في عمليات الاصطفاف

وعند تمثيل التطابق بالبرامج التصويرية تظهر بشكل مستوي كما موضح في نتائج البرنامج Clustal X لصف البروتينات (شكل 8.5) :

human	---MEEPQSDPSVEP-PLSQETFS	20
monkey	---MEEPQSDPSIEP-PLSQETFS	20
mouse	MTAMEESQSDISLEL-PLSQETFS	23
rat	---MEDSQSDMSIEL-PLSQETFS	20
xenopus	---ME-PSSETGMDP-PLSQETES	19
chicken	---MA-EEMEPLLEPTVEFMDLW-	19
	* . : :: : :	

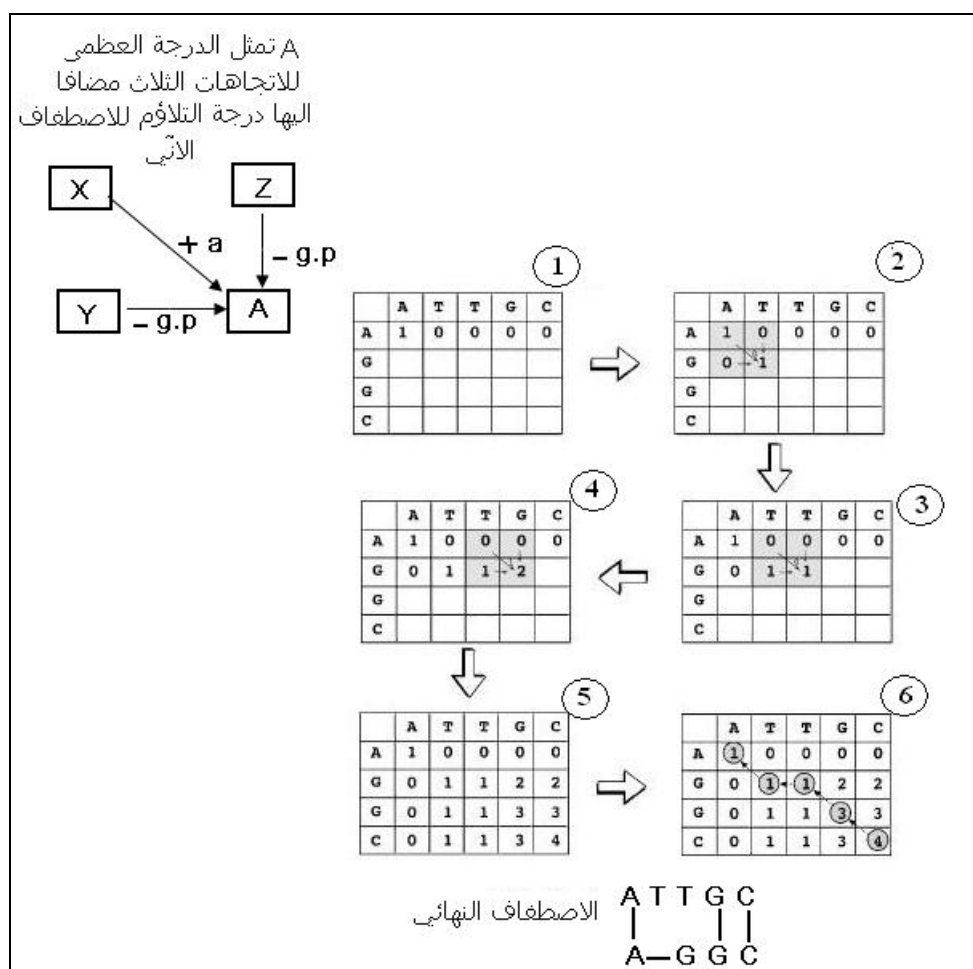
شكل 8.5 : نتائج صف تواليات باستعمال برنامج Clustal X

ولطريقة مصفوفات التنقيط تطبيقات في التعامل مع الجينومات حيث تستعمل في تحديد المكررات في الكروموسومات التي تظهر بشكل خطوط متوازية بالطول نفسه وكذلك تستعمل في تحديد التراكيب الثانوية للحوامض النووية وذلك بالكشف عن التواليات المكملية.

طريقة البرمجة الديناميكية Dynamic programming method

تحدد طريقة البرمجة الديناميكية الاصطفاف الأمثل بواسطة تطابق تواليين لجميع الأزواج المحتملة للحروف بين التواليين، وهي بالأساس مماثلة لطريقة مصفوفة التنقيط في أنها تنشأ شبكة من الاصطفاف ذات البعدين، وتعمل على إيجاد الاصطفاف بطريقة كمية وذلك بتحويل مصفوفة التنقيط الى درجات (أرقام) لحساب التطابق او عدم التطابق بين التواليات ومن خلال البحث في مجموعة من أعلى الدرجات في هذه المصفوفة يمكن الحصول على أدق اصطفاف. وتعمل الطريقة أولاً على بناء مصفوفة ذات بعدين تمتد التواليات على إحداثيات المصفوفة كي يمكن المقارنة بينهما. ويكون تطابق الثمالات في المصفوفة له درجة معينة وتحتسب الدرجة

مرة واحدة لكل صف، وتبدأ عملية تحديد الدرجات عند الصف الأول لأحد التواليات الذي يستخدم للكشف على طول التوالي الآخر ثم يتبعها الصف الثاني، ويتم احتساب درجات التطابق عند الكشف عن الصف الثاني بالأخذ بنظر الاعتبار الدرجة المسجلة للصف الأول وتوضع الدرجة الأفضل في الزاوية اليمنى السفلى من المصفوفة الوسطية وتكرر عملية احتساب الدرجات حتى يتم مليء جميع خلايا المصفوفة كما موضح في الشكل 9.5 :



شكل 9.5 : طريقة البرمجة الديناميكية في الاصطفاف

يوضح الجدول الأول في الزاوية اليسرى العليا من الشكل تحديد القيمة (1) في الصف الأول منه لتطابق الثمالتين (A) في كلا التواليتين ، ويشير كلا الجدولين الثاني والثالث (في الشكل أعلاه) على التوالي الى تحديد القيمة (1) في السطر الثاني من الجدول للثمالات (غير المتمثلة) المتقابلة في كلا التواليتين . ويشير الجدول الرابع الى تحديد القيمة (2) في السطر الثاني التي تقابل الثمالة G في كلا التواليتين . ويشير الجدول الخامس الى امتلاء الجدول بالقيم المقابلة لاصطفاف الثمالات ، وأخيرا يشير الجدول السادس والأخير الى تحديد مسار (اتجاه)

الاصطفاف من الزاوية اليسرى السفلى التي تحتوي على القيمة (4) في الجدول التي تقابل الشمالية C من التوالي الأول (العمود) والشمالية C من التوالي الثاني (الصف) والانتقال الى الزاوية العليا اليمنى التي تحتوي على القيمة (3) في الجدول التي تقابل الشمالية G من التوالي الأول (العمود) والشمالية G ايضا من التوالي الثاني (الصف) ، ثم الانتقال الى الزاوية العليا اليمنى التي تحتوي على القيمة (1) في الجدول التي تقابل الشمالية T من التوالي الأول (العمود) والشمالية G من التوالي الثاني (الصف) ، والانتقال بصورة أفقية الى اليسار وعلى المحور نفسه الى الخلية المجاورة التي تمتلك القيمة (1) ايضا والتي تقابل الشمالية T من التوالي الأول (العمود) في حين انها لا تقابل شمالية من التوالي الثاني (الصف) كون الشمالية G سبق وان حددت في الخطوة السابقة لذلك يحدد خط (-) للدلالة على وجود فجوة في هذا التوالي ثم الانتقال الى الزاوية (الأخيرة) العليا اليسرى التي تمتلك القيمة (1) والتي تقابل الشمالية A من التوالي الأول (العمود) والشمالية A من التوالي الثاني (الصف) . لذا فالشكل (9.5) يوضح تراكم الدرجات على طول القطر ومن الزاوية اليسرى العليا الى الزاوية السفلى اليمنى وبعد تثبيت الدرجات في المصفوفة بصورة كاملة يتم إيجاد المسار الذي يمثل أفضل اصطفاف وتطابق وذلك من الرجوع الى الخلف وباتجاه معاكس لطريق بناء واحتساب المصفوفة في الخطوة السابقة، وذلك بالبداية من الزاوية اليمنى السفلى للمصفوفة باتجاه أصل المصفوفة في الزاوية اليسرى العليا ، ويحدد أفضل مسار للتطابق الذي يحصل على أعلى مجموعة الدرجات ، وفي حالة اذا تطابق او تساوي مجموعين فيتم اختيار أحد الطريقين عشوائياً ليمثل أفضل اصطفاف. وعند بناء مسار الاصطفاف فانه يمكن ان يتجه أفقياً او عمودياً عند نقطة معينة فهو يمكن ان يكون ناتجاً عن حصول فراغ او فجوة او حصول حالة إقحام او حذف في أحد مكونات التوالي.

جزاءات الفجوة Gap penalties

غالباً ما يتضمن الحصول على أفضل اصطفاف بين تواليين الأخذ بنظر الاعتبار الفجوات Gaps التي تمثل بالإضافة (الإقحام) او الحذف في مكونات التواليات، والملاحظ ان عمليات التطور الطبيعية نادراً ما تكون معتمدة على حذف او إقحام أحد الثمالات ولكن عمليات الاستبدال خاصة في حالة الحوامض الامينية تكون هي الشائعة. ولهذا فعند الأخذ بنظر الاعتبار الفجوات تكون عمليات الاصطفاف أكثر صعوبة وتطبق وفق قيود حسابية تعكس ندرة حدوث الإقحام او الحذف لمكونات التواليات في عملية التطور. ولذلك توضع قيم جزاءات لهذه الفجوات ويكون تحديد هذه القيم عشوائياً وذلك لعدم وجود نظرية للتطور لتحديد بدقة القيمة او الكلفة الخاصة بالإضافة او الحذف لمفردات التواليات. وإذا حددت قيم الجزاءات بمستوى واطئ فان هذا يؤدي الى زيادتها بإعداد كبيرة وضخمة وبالتالي السماح حتى للتواليات المتباعدة والتي ليس بينها علاقة من التطابق بدرجات تشابه عالية، اما عند تحديد قيم عالية للجزاءات فعندها يكون من الصعوبة ظهور الفجوات وبالتالي لا يمكن تحقيق الاصطفاف المعقول وهذا ايضا غير واقعي، ومن الدراسات التجريبية للبروتينات الكروية طورت قيم الجزاءات التي تبدو مناسبة لأكثر الاصطفافات وقد تم اعتمادها في أكثر البرامج الخاصة بالاصطفافات. ومن جهة ثانية يمكن النظر الى الفرق في الكلفة بين فتح فجوة او توسيع فجوة موجودة ومن المعلوم ان الحالة الأخيرة تكون هي الحالة الأسهل وبالتالي فان فتح فجوة يجب ان يحدد بقيمة جزائية أعلى مقارنة مع الفجوة الموجودة.

وهذا يعود أساسا الى الحالة المنطقية فعند حدوث حالات الإقحام أو الحذف فإن عدة ثملالات متجاورة يمكن ان تقحم أو تحذف سوية وقيم الجزاءات هذه تسمى جزاءات الفجوات المرتبطة Affine gap penalties ولذلك كان لابد من تهيئة قيم محددة مسبقا لفتح أو توسيع الفجوات، ومن هذه القيم مثلا يكون جزاء فتح فجوة (-12) وتوسيع فجوة موجودة أصلا يكون (-1) وبالتالي يمكن حساب قيمة الجزاء الكلية (W) التي تكون بشكل علاقة خطية مع طول الفجوة ويمكن حسابها من المعادلة الآتية:

$$W = \gamma + \delta \times (K-1)$$

حيث γ تمثل قيمة جزاء فتح الفجوة في التوالي، δ تمثل قيمة جزاء توسيع الفجوة، K تمثل طول الفجوة.

وفضلا عن هذا النوع من الجزاءات المرتبطة ، هناك نظام آخر يستخدم جزاء الفجوة الثابتة الذي يحدد درجات متماثلة لكل موقع من الفجوات بغض النظر عن الحالة اي سواء كانت عملية فتح فجوة أو توسيع فجوة موجودة، وعلى أية حال فإن هذا النوع من المخطط هو أقل واقعية مقارنة مع النوع الأول. وفي العادة يتم إهمال الفجوات التي تقع في أطراف التواليات نظرا لأن التواليات ذات التماثل الجذري الحقيقي تكون ذات أطوال مختلفة لذلك يتم إهمالها لتجنب الحصول على اصطافات غير واقعية.

مثال

من الخوارزميات التقليدية والشائعة لإجراء عمليات الاصطفاف الشامل هي خوارزمية Needleman-Wunsch ، في هذا المثال يفترض دراسة التواليتين الآتيتين:

GATCGATTCGAATCGATT (1) توالي رقم

GATCGAAATCGATT (2) توالي رقم

المطلوب : احتساب جزاء الفجوة

الحل : صف كلا التواليتين للحصول على أفضل تماثل للتوالي تستخدم الفجوات وكالاتي :

GATCGATTCGAATCGATT : التوالي (1)

GATCGA - - - AATCGATT : التوالي (2)

بعد ذلك تستخدم معادلة احتساب الفجوات وكالاتي :

$$\omega = \gamma + \delta \times (k-1)$$

$$\gamma = -12$$

$$\delta = -1$$

$$k = 4$$

وبهذا فإن

$$\omega = -12 + [(-1) \times 4]$$

$$= -16$$

مقدار جزاء الفجوة

البرمجة الديناميكية للاصطفاف الشامل

الطريقة التقليدية لهذا النوع من الاصطفاف هي خوارزمية Needleman-Wunsch إذ يمكن الحصول على أفضل اصطفاف وعلى طول التواليتين الكامل، إذ يجب ان يمد كلا التواليتين من

البداية وحتى النهاية للحصول على أعلى مجموع من الدرجات وبعبارة أخرى فإن مسار الاصطفاف ينطلق من الزاوية اليمنى السفلى للمصفوفة الى الزاوية العليا اليسرى. ان المشكلة في التركيز على الحصول على أعلى او أعظم درجة لاصطفاف التوالي بكامله يؤدي الى خطورة فقدان أفضل تماثل موضعي. ولذلك كان الاصطفاف الشامل إستراتيجية ملائمة لصف تواليين متقاربة في الطول، اما في حالة اختلاف طول التواليات فإن الطريقة ليست هي المثلى.

البرمجة الدائيميكية لاصطفاف الموضعي

ليس من السهل تحديد مدى تباعد التواليين لغرض إخضاعهم لعمليات الصف الشامل كما ان اشتراط تساوي طول التواليين يمثل جانب آخر من المشكلة. وفي هذه الحالة يكون من الأفضل تحديد مناطق التشابه في التواليين وقد يكون هذا ذو معنوية اكبر مقارنة مع إيجاد تطابق جميع مفردات او ثملات التواليين. وتعد خوارزمية Smith-Waterman أول تطبيق للبرمجة الدائيميكية في الاصطفاف الموضعي. وفي هذه الخوارزمية تحدد الدرجات الموجبة للثملات المتطابقة وصفر لعدم التطابق ولا تستخدم الدرجات السالبة، وتستخدم عملية المسار الخلفي Tracing-back بصورة مماثلة للبرمجة الدائيميكية المذكورة أعلاه. ان مسار الاصطفاف قد يبدأ وينتهي على طول القطر الرئيس، اذ تبدأ عند الموقع الأعلى درجة ويستمر قطريا والى اليسار حتى يصل المربع الذي فيه القيمة صفر وتضاف عند الضرورة الفراغات، وغالبا ما تستخدم جزاء الفجوات المرتبطة . وبالتالي يتم الحصول على قطع اصطفاف مثلى لها أعلى الدرجات وكما هو الحال مع الاصطفاف الشامل فإن القيمة النهائية تتأثر باختيار أنظمة الدرجات المستخدمة (التي سيتم ذكرها لاحقا).

ان الهدف من عمليات الاصطفاف الموضعي الحصول على أعلى درجة للمواقع المعنية وهذا قد يكون على حساب أعلى درجة ممكنة لاصطفاف الطول الكلي. وهذه الطريقة قد تكون مناسبة لصف التواليات المتباعدة او التواليات الحاوية على دومينات متعددة التي قد تكون من أصول مختلفة. وهناك عدة مواقع على شبكة الانترنت وبرامج تتعامل مع هذا النوع من الاصطفاف. وعمليات اصطفاف زوج من التواليات لا تعطي نتائج مثلى ولكن تكون دون الحالة المثلى بقليل وتبقى في هذه الحالة حتى عند تعديل بعض المؤشرات والسبب حدوث طفرات في مناطق مختلفة من الجزيئات الحيوية التي يصعب تعديلها.

الاصطفاف شبه الشامل Semiglobal alignment

وتسمى ايضا Glocal وفي هذه الحالة يتم البحث عن بداية ونهاية التواليات لإجراء عمليات الاصطفاف عليها، وهذه تكون للتواليات المتداخلة اي تكون نهاية توالي Downstream متداخل مع بداية توالي آخر Upstream كما في حالة جينومات بعض الفيروسات وفي هذه الحالة تكون عملية الاصطفاف الشامل او الموضعي غير ملائمة ، لان الاصطفاف الشامل سيمتد الى ما بعد مناطق التداخل ، في حين ان الاصطفاف الموضعي قد لا يشمل مواقع التداخل. وتستعمل للعمليات التي لا تحتاج الى دقة عالية مثل البحث في قواعد البيانات التي تكون ذات تشابه عالي مع توالي الاستعلام.

وعلى العموم فانه في البرمجة الدائيميكية القياسية بأنواعها فإن درجة كل موقع من القواعد النتروجينية تكون مستقلة عن تماثل القواعد المجاورة ولذلك فإن تأثير التراص Stacking effect لا يؤخذ بنظر الاعتبار ولكن مثل هذه يمكن التغلب عليها بتحويل البرنامج.

مصفوفات الدرجات Scoring matrices

وضع المصفوفات يسهل دراسة التواليات سواء كانت تواليات حوامض أمينية أو قواعد نتروجينية في DNA فمثلا لو أريد صف التوالين الآتية:

ATTCAGAACCATTAA
TTCAGTAACCAAT

فان ذلك سيكون صعبا حتى مع هذين التوالين البسيطة لذلك تستعمل المصفوفات لغرض جسر الهوة بين البيانات الخام أو الأولية وترتيبها لغرض تسهيل عمليات اصطفاها. وهناك أنواع من المصفوفات فمصفوفة الدرجات وتسمى أيضا بمصفوفة الاستبدال اشتقت من التحليل الإحصائي لبيانات استبدال الحوامض الأمينية عند صف تواليات ذات ارتباط عالي.

اما بالنسبة لمصفوفات الدرجة الخاصة بتواليات النيوكليوتيدات فتكون بسيطة نسبيا. وفيها تعطى قيم موجبة أو درجة عالية لحالة التطابق وقيم سالبة أو درجات واطئة لحالة عدم التطابق. وهذا التحديد يعتمد على فرضية ان تكرار التغيير يكون متساوي لجميع القواعد النتروجينية، وعلى أية حال فان هذه الفرضية قد لا تكون واقعية، وتشير الشواهد العملية على ان حالة الاستبدال المتجانس بين البيورينات أو بين البريميدينات هي الأكثر تكرارا من حالة الاستبدال المتغاير بين مجموعتي القواعد النتروجينية أي استبدال قاعدة بيورينية محل قاعدة برييميدينية وبالعكس، ولذلك كانت هناك حاجة الى إيجاد نموذج إحصائي أكثر تعقيدا يأخذ بنظر الاعتبار احتمالات مختلفة تعكس نوعي الاستبدالين.

وتعد مصفوفات الدرجات الخاصة بالأحماض الأمينية أكثر تعقيدا وذلك لأنها يجب ان تعكس الخصائص الفيزيوكيميائية للأحماض الأمينية وكذلك الاحتمال الأرجح لاستبدال حامض أميني معين من خلال التماثل الجذري للتواليات. ان استبدال حامض أميني معين مكان حامض أميني آخر من زمرته أي مشابه له في الخصائص الفيزيوكيميائية يكون أكثر احتمالا وقبولا من استبدال حوامض من زمر مختلفة الصفات، اذ ان استبدال الحوامض الأمينية من زمرة أو مجموعة واحدة يوفر احتمالية الاحتفاظ بالوظائف الأساسية والتركيبية ولكن احتمالية تعطيل الوظائف وإرباك التركيب هي أيضا واردة ، ولكن الحالة الأخيرة يتم التغاضي عنها أثناء عمليات الانتخاب التطوري وذلك فعند حصولها ستؤدي الى إنتاج بروتينات غير فعالة . ومن المشاهدات المسجلة احتمالية الاستبدال بين الحوامض الأمينية الحلقية مثل الفنيل-النين، والتايروسين والتربتوفان دون التأثير الكبير في تركيب أو وظيفة البروتينات. ولكن هناك حالات خاصة في حالة استبدال البرولين أو الكلايسين لما لهذين الحامضين من تراكيب خاصة وان كانا في زمرة واحدة فهي ثملات لا قطبية صغيرة .

مصفوفات الدرجات الخاصة بالحوامض الامينية

استنبطت مصفوفات الاستبدال للحوامض الامينية وتتكون من 20×20 لتعكس الاحتمال الأقوى لاستبدال الحوامض الامينية. وهناك نوعان أساسيان من مصفوفات استبدال الحوامض الامينية، يعتمد الأول على قابلية التبادل للشفرة الوراثية (نظرا لوجود بعض الحوامض الامينية التي يشفر لها بأكثر من شفرة والتي قد تصل في أقصاها الى ستة شفرات) للحوامض الامينية او خواص الحوامض الامينية ، في حين اشتق النوع الثاني من الدراسات التجريبية لاستبدال الحوامض الامينية . وعلى الرغم من ذلك فان الطريقتين المختلفتين تتوافقان الى مدى معين، الا ان الطريقة الأولى المعتمدة على الشفرات الوراثية وخصائص الحوامض الامينية أظهرت دقة أقل مقارنة بالطريقة الثانية والتي تعتمد على معاينة الاستبدال الحقيقي للحوامض الامينية عند استعمال عدداً من البروتينات ذات العلاقة، لذا حصلت الطريقة الثانية على قبول أكثر في تطبيقات اصطفايف التواليات .

وتشمل المصفوفات التجريبية مصفوفات Point Accepted Mutation (PAM) والأصح Accepted point mutation ولكن لسهولة اللفظ استعملت الأولى. و BLOSUM وقد اشتقتا من الاصطفايفات الحقيقية لتواليات ذات تشابه عالي بواسطة تحليل احتمالية استبدال الحوامض الامينية. وفي هذه الاصطفايفات يمكن تطوير نظام الدرجات بإعطاء درجات عالية لاستبدال الحوامض الامينية الأكثر حدوثاً في حين تعطى درجات واطئة لاستبدال الحوامض الامينية القليلة الحدوث او النادرة .

وفي مصفوفات الاستبدال فان الدرجة الموجبة تعني تكرار استبدال الحامض الاميني في مجموعة البيانات نتيجة للتماثل الجذري أكثر من حدوثها نتيجة المصادفة العشوائية ولذلك فالدرجة الموجبة تمثل استبدال الحوامض الامينية المتطابقة او ذات التشابه العالي (اي تكون في الزمرة نفسها). اما الدرجة صفر فهي تمثل حالة الاستبدال الذي يحدث بالصدفة العشوائية . وتشير الدرجات السالبة الى ان تكرار استبدال الحامض الاميني في بيانات التواليات المتماثلة جذريا هو اقل حدوثاً نتيجة المصادفة العشوائية وهذا يحدث طبيعياً بين الحوامض الامينية غير المتشابهة. تطبق مصفوفات الاستبدال الصيغ اللوغاريتمية لوصف احتمالية استبدال الحوامض الامينية. والقيم المحولة الى الصيغ اللوغاريتمية تسمى ايضا درجات اللوغاريتم الشاذة Log-odds scores او نسب اللوغاريتم الشاذة Log-odds ratios ، وتمثل نسب اللوغاريتم لتكرار الطفرات مقسومة على احتمالية الاستبدال العشوائي الذي يحدث بالصدفة. وصيغ او تحويل الأرقام الى لوغاريتمات اما ان يكون للأساس 10 او للأساس 2.

مصفوفات PAM

أنشأت مصفوفات PAM من قبل العالمة M. Dayhoff ولذلك تدعى بعض الأحيان باسمها، اذ صنفت اصطفايفات 71 مجموعة من تواليات بروتينية متقاربة جداً، ثم أنشأت بعد ذلك مصفوفات أخرى من PAM اعتماداً على الانفراج او الابتعاد التطوري Evolutionary divergence بين التواليات من العنقود Cluster او المجموعة نفسها. وتمثل الوحدة الواحدة في هذه المصفوفات 1% من التغيير في مواقع الحوامض الامينية.

يتضمن بناء المصفوفة PAM1 من الاصطفايف الكامل على طول التواليات ومن ثم بناء أشجار العلاقات التطورية، ثم تشتق درجة PAM لأزواج معينة من الحوامض الامينية من عملية

الخطوات المتعددة المتضمنة حساب التطفر النسبي الذي يمثل عدد التغييرات الحاصلة نتيجة الطفرات في التوالي الأصلي (الجد المشترك) لثمالة الحامض الاميني مقسوما على عدد ثمالات الحامض الحاصلة في عملية الاصطفاف ثم تعديل تكرار استبدال الثمالة المتوقعة من الصدف العشوائية وتحويل الأرقام الناتجة الى قيم لوغاريتمية للأساس 10 لقيم الطفرات المعدلة مقسومة على تكرار الثمالة المعنية ، وتقرب القيم الناتجة الى اقرب عدد صحيح وتدخل في مصفوفة الاستبدال التي تعكس الاحتمال الأرجح لاستبدال الحوامض الامينية ، وبذلك تنتهي عملية حساب درجة اللوغاريتم الشاذة Log-odds score . وبعد تجميع كل احتمالات الاستبدال للطفرات الحاصلة للحامض الاميني تكون مصفوفة PAM (20×20) قد اكتملت. وتشير القيم الموجبة في المصفوفة الى حدوث تكرار في الاستبدال أكثر من المتوقع بين الاستبدالات التطورية الثابتة، في حين تشير القيم او الدرجات السالبة الى حدوث الاستبدالات اقل من التكرار المتوقع . ومصفوفات PAM الأخرى ذات الأعداد المتزايدة للتواليات الأكثر تشعباً او المتباعدة تنتج من عمليات الضرب للـ PAM1 فمثلا المصفوفة PAM80 هي ناتج حاصل ضرب المصفوفة PAM1 في نفسها ثمانين مرة. وقد تم حساب عدد مصفوفات PAM ومعدل حدوث الطفرات ونسب تماثل التواليات ووضعت القيم الموضحة في الجدول 1.5 :

جدول 1.5 : مصفوفة PAM

مصفوفة PAM	% معدل الطفرات المشاهدة	% تماثل التواليات
0	0	100
1	1	99
30	25	75
80	50	50
110	40	60
200	75	25
250	80	20

ويلاحظ من الجدول ان مصفوفة PAM250 تناظر 20% من تطابق الحوامض الامينية او تمثل حصول 250 طفرة لكل 100 ثمالة ولذلك تستخدم للتواليات المتشعبة او المنفرجة ، في حين ان مصفوفات PAM ذات الأرقام الواطئة هي الأكثر ملائمة لاصطفاف التواليات ذات العلاقة القريبة. ويوضح الشكل التالي (10.5) القيم المستنتجة لمصفوفة PAM250 لاستبدال الحوامض الامينية.

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	12																				C
S	0	2																			S
T	-2	1	3																		T
P	-3	1	0	6																	P
A	-2	1	1	1	2																A
G	-3	1	0	-1	1	5															G
N	-4	1	0	-1	0	0	2														N
D	-5	0	0	-1	0	1	2	4													D
E	-5	0	0	-1	0	0	1	3	4												E
Q	-5	-1	-1	0	0	-1	1	2	2	4											Q
H	-3	-1	-1	0	-1	-2	2	1	1	3	6										H
R	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6									R
K	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5								K
M	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6							M
I	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5						I
L	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6					L
V	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4				V
F	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9			F
Y	0	-3	-3	-5	-3	-5	-3	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10		Y
W	-6	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

PAM250 matrix, log-odds representation

شكل 10.5 : مصفوفة PAM 250

مصفوفات BLOSUM

يلاحظ ان مصفوفات PAM توفر الفرصة لاصطفاف التواليات المتقاربة لذلك فعند بناء مصفوفة PAM فان الاستبدال المباشر للمشاهد للحوامض الامينية هو في مصفوفة PAM1 والتي اعتمدت على مجموعة صغيرة نسبيا من التواليات المتقاربة بشدة ، ولذلك في مثل هذه المصفوفات تكون الاختبارات الإحصائية لاصطفاف التواليات الأكثر تباعدا غير متاحة. ولأجل ذلك تم تطوير مجموعة جديدة من مصفوفات الاستبدال وهي مصفوفات القطاعات **Blocks** **Substitution Matrices** (BLOSUM) المشتقة اعتمادا على المشاهدات المباشرة لكل إمكانية استبدال للحامض الأميني في عمليات اصطفاف التواليات المتعددة والتي استعمل فيها أكثر من 2000 نموذج من الأنماط الثابتة للحوامض الامينية تمثل 500 مجموعة من التواليات البروتينية . هذه الأنماط تدعى القطاعات **Blocks** والتي يبلغ طولها اقل من 60 حامض أميني بدون فجوات ويحسب تكرار استبدال الحوامض الامينية في هذه القطاعات للحصول على جدول رقمي او مصفوفة استبدال القطاعات. وبدلا من استخدام دالة الاستقراء فان مصفوفات BLOSUM هي في الواقع نسب تمثل تطابق التواليات المختارة لبناء المصفوفات. فعلى سبيل المثال تشير المصفوفة BLOSUM62 الى ان التواليات المستخدمة في بناء المصفوفة تشترك بتطابق معدله 62%. وتم بناء مصفوفات أخرى اعتمادا على مجاميع من التواليات التي تختلف في مستويات تماثلها. ويلاحظ ان ترتيب الأرقام تصاعديا او تنازليا يكون معاكس لمصفوفات PAM فيشير رقم مصفوفة BLOSUM الواطئ الى درجة عالية من التباعد بين التواليات.

وتشتق درجة مصفوفة BLOSUM الخاصة بزواج من الحوامض الامينية من نسبة لوغاريتم او تكرار استبدال للثمالة مقابل الاحتمالية المتوقعة للثمالة. والقيم الشاذة للوغاريتم Log- odds تؤخذ للأساس 2 بدلا من الأساس 10 كما هو معمول به في مصفوفات PAM، وتقرب القيم الناتجة الى أقرب عدد صحيح وتدخل الى مصفوفة الاستبدال. وكما هو الحال مع مصفوفات PAM فان القيم الموجبة والسالبة تشير الى حدوث عمليات استبدال بتكرار أكثر او اقل من المتوقع أثناء عمليات التطوير ويوضح الشكل التالي (11.5) قيم مصفوفة BLOSUM62:

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

شكل 11.5: مصفوفة BLOSUM 62

المقارنة بين مصفوفتي PAM و BLOSUM

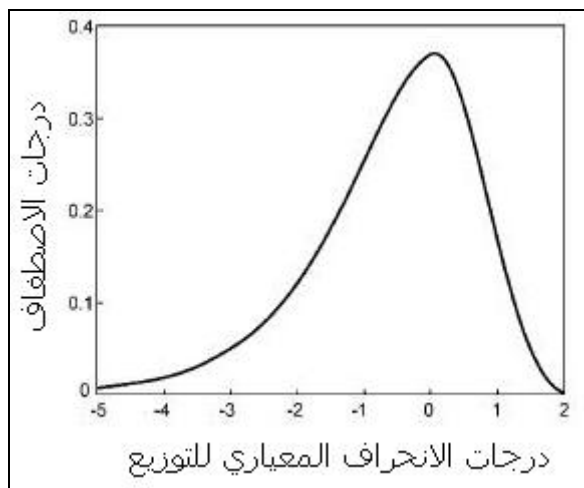
هناك عدد من الاختلافات بين المصفوفتين، والاختلاف الرئيس هو ان مصفوفات PAM باستثناء PAM 1 اشتقت من نماذج عمليات التطور في حين ان الأخرى تتكون من قيم بنيت كليا على المشاهدات المباشرة ولذلك فقد تكون ذات دلائل تطور اقل مقارنة بمصفوفات PAM ولذا غالبا ما تستعمل مصفوفات PAM لبناء أشجار العلاقات التطورية. وعلى اي حال فان قيم الاستقراء الرياضي المستعملة في مصفوفات PAM قد تكون اقل واقعية للتواليات المتباعدة. ان مصفوفات BLOSUM اشتقت كليا من عمليات الاصطفاف الموضعي لقطاعات التواليات الثابتة في حين ان مصفوفة PAM1 تعتمد على الاصطفاف الشامل لطول التوالي الكامل بما يحتويه من مناطق ثابتة وأخرى متغيرة ، لذلك تكون مصفوفات BLOSUM أكثر تفضيلا عند البحث في قواعد البيانات وإيجاد الدومينات الثابتة في البروتينات.

وتشير الاختبارات التجريبية العديدة الى ان مصفوفات BLOSUM تتفوق على PAM من حيث دلالات دقة الاصطفافات الموضعية، وهذا التفوق يمكن ان يكون كبيرا لان مصفوفات BLOSUM قد اشتقت من معلومات قاعدة بيانات اكبر من تلك المستعملة لاشتقاق مصفوفات PAM وبذا يمكن اعتماد قيمها أكثر من قيم PAM.

وقد استتبعت مصفوفات أخرى باستخدام الطريقة نفسها ولكن باعتماد مجموعات بيانات اكبر ومنها مصفوفات Gonnet و Jones-Taylor-Thornton والتي أظهرت اداءً "مكافئاً" لمصفوفة BLOSUM في عمليات الاصطفاف النظامية ولكنها فعالة أكثر في بناء أشجار العلاقات التطورية.

المعنوية الإحصائية لاصطفاف التواليات

عندما تظهر عملية اصطفاف التواليات درجة معينة من التشابه فمن الضروري معرفة ان هذا الاصطفاف ناتجاً عن المصادفة العشوائية ام انه اصطفاف واقعي وله دلالات إحصائية معنوية، وذلك لان الاصطفاف ذات المعنوية الإحصائية الحقيقية يكون قادراً على توفير مصداقية للتمائل الجذري بين التواليات. ولحل هذه المشكلة لابد من إجراء اختبار إحصائي لدرجات الاصطفاف لتوالياين متباعدة ومتساوية في الطول. وبحساب درجات الاصطفاف لعدد كبير من أزواج التواليات المتباعدة يمكن اشتقاق نمط او نموذج توزيعها، ومن هذا التوزيع يمكن إجراء اختبار إحصائي اعتماداً على عدد من الانحرافات المعيارية عن درجة المعدل. وقد أوضحت العديد من دراسات توزيع درجات التشابه ان نمط التوزيع يكون غريباً او خاصاً ويكون مشابهاً لنمط التوزيع الطبيعي ولكن مع انحراف الى أحد الجوانب وذيل في الجانب الآخر كما موضح في الشكل 12.5 :



شكل 12.5 : توزيع نمط التشابه لاصطفاف التواليات

البحث عن التشابه في قواعد البيانات

التطبيق الرئيس لعمليات الاصطفاف المذكورة سابقاً هو البحث عن التواليات المتشابهة لتوالي الاستعلام في قواعد البيانات، وتتضمن العملية عرض توالي الاستعلام لإجراء عمليات اصطفاف زوج من التواليات مع كل توالي مودع في قاعدة البيانات. وعليه فان البحث عن

تشابه توالي الاستعلام مع التواليات المودعة في قواعد البيانات يمثل أحد الركائز المعتمدة في تحديد مواصفات ووظائف التواليات الجديدة. وتطبيق طريقة البرمجة الدائميكية المذكورة سابقا بطيئة ولذا تكون غير عملية في الكثير من الحالات، ولهذا كانت هناك حاجة الى طرق سريعة لمقارنة التواليات.

مواصفات الطرق المستعملة للبحث في قواعد البيانات

تحتاج الطرق المستعملة للبحث في قواعد البيانات الى مواصفات خاصة لغرض اعتمادها ومنها:

- الحساسية Sensitivity
- الانتخابية Selectivity او Specificity
- السرعة Speed

تشير الموصفة الأولى (الحساسية) الى قابلية الطريقة لإيجاد التطابقات الصحيحة بين التواليات قدر الإمكان، وينظر الى التطابق الصحيح على انه نتائج صحيحة موجبة . في حين تشير الانتخابية الى قابلية الطريقة على استبعاد التطابقات غير الصحيحة والتي يمكن ان تنتج من تطابق توالي الاستعلام مع تواليات بعيدة وتعد هذه نتائج موجبة كاذبة .

المعيار الثالث او الصفة الثالثة للطرق المستعملة هو السرعة وتعني الوقت المستغرق للحصول على نتائج البحث في قواعد البيانات وهذا العامل يعتمد على حجم قاعدة البيانات.

وعند اختيار الطريقة لابد من الموازنة بين هذه المؤشرات الثلاث، وتحقيق هذه المتطلبات سوية يكون صعبا، وذلك لان الحساسية والانتخابية او الانتقائية تتناسب عكسيا فزيادة الحساسية يرافقه انخفاض في الانتخابية وعندها تسجل العديد من النتائج الموجبة الكاذبة، ومن الطرف الآخر فان زيادة السرعة في تنفيذ الطريقة يؤثر ويقلل من حساسية وانتقائية الطريقة.

وفي هذا المضمار وكما هو الحال مع مجالات المعلوماتية الحيوية الأخرى هناك نوعين من الخوارزميات:

- النوع الشامل Exhaustive type وتستخدم فيه خوارزميات صارمة لإيجاد الحل الأمثل او المضبوط لمشكلة معينة وذلك باختبار جميع التداخلات الرياضية. وتعد البرمجة الدائميكية مثال على هذه الطريقة.

- النوع الموجه Heuristic type او طريقة الإزاحة (في المفهوم الفيزيائي) او ما يسمى بالطريق المختصر وهو عبارة عن إستراتيجية رياضية لإيجاد حل تجريبي او حل قريب من الحلول المثلى. وبصورة أساسية فان هذا النوع من الخوارزميات يأخذ الطريق المختصر بتقليصه مجال البحث استنادا الى بعض المعايير، وبذا يلاحظ ان إستراتيجية الطريق المختصر لا تعطي ضمان للوصول الى الحل الأمثل او الأكثر دقة.

وأكثر الطرق شيوعا هو استعمال الكلمة Word في حالة برنامج BLAST وما يقابلها K-tuple في حالة برنامج FASTA ، وتتصف الطريقة (كما سيأتي ذكرها لاحقا) بتحديد سلسلة قصيرة Words من تواليات الاستعلام غير المتداخلة ثم تتم مطابقتها مع توالي قواعد البيانات الملائمة وبتطابق هذه الكلمات يهمل بقية التوالي الذي ليس له اي تشابه.

وبالرغم من ان طرق البرمجة الدائميكية هي الأدق مثل خوارزمية Smith-Waterman والتي يعول عليها الا انها بطيئة جدا وتصبح غير مجدية عند البحث في قاعدة بيانات تحوي

على أكثر من 300,000 توالي لذلك تستخدم الطرق الموجهة لأنها الأسرع وتختبر فقط الجزء المحتمل للاصطفاف الذي كان قد اختير بطرق البرمجة الديناميكية النظامية. وتوجد خوارزميتان من النوع الموجه لانجاز البحث في قواعد البيانات وهي BLAST و FASTA، وهذه الطرق ليست مضمونة لإيجاد أفضل اصطفاف أو التماثل الحقيقي ولكنها أسرع بـ 50-100 مرة من طرق البرمجة الديناميكية. والزيادة في سرعة الانجاز تأتي على حساب الحساسية والانتقائية ولكنها في هذه الخوارزميات تكون الكلفة مقبولة لحد ما من قبل العاملين في مجال علوم الحياة.

يستخدم كلا BLAST و FASTA طريقة الكلمة الموجهة Heuristic word method لصف السريع لزوجي التواليات (أي توالي الاستعلام والتوالي الموجود في قاعدة البيانات) وتعمل الطرق على إيجاد بحث مختصر لعدد من الحروف المتطابقة أو القريبة من التطابق بين التواليتين وهذه تمثل الكلمات والتي تشابه النوافذ المستخدمة في طريقة مصفوفة التتقيط السابقة الذكر، والافتراض الأساسي في هذه الحالة هو أن التواليتين المتقاربتين لابد أن يحتويان على الأقل كلمة واحدة متطابقة في المرحلة الأولى ثم يمتد الاصطفاف على طول التواليتين لإيجاد مناطق أو كلمات أخرى مشابهة، وعند إيجاد مناطق متشابهة فإن المناطق المتشابهة هذه يتم ربطها مع بعضها للحصول على الاصطفاف الكامل.

أداة بحث الاصطفاف الموضوعي الأساسي BLAST

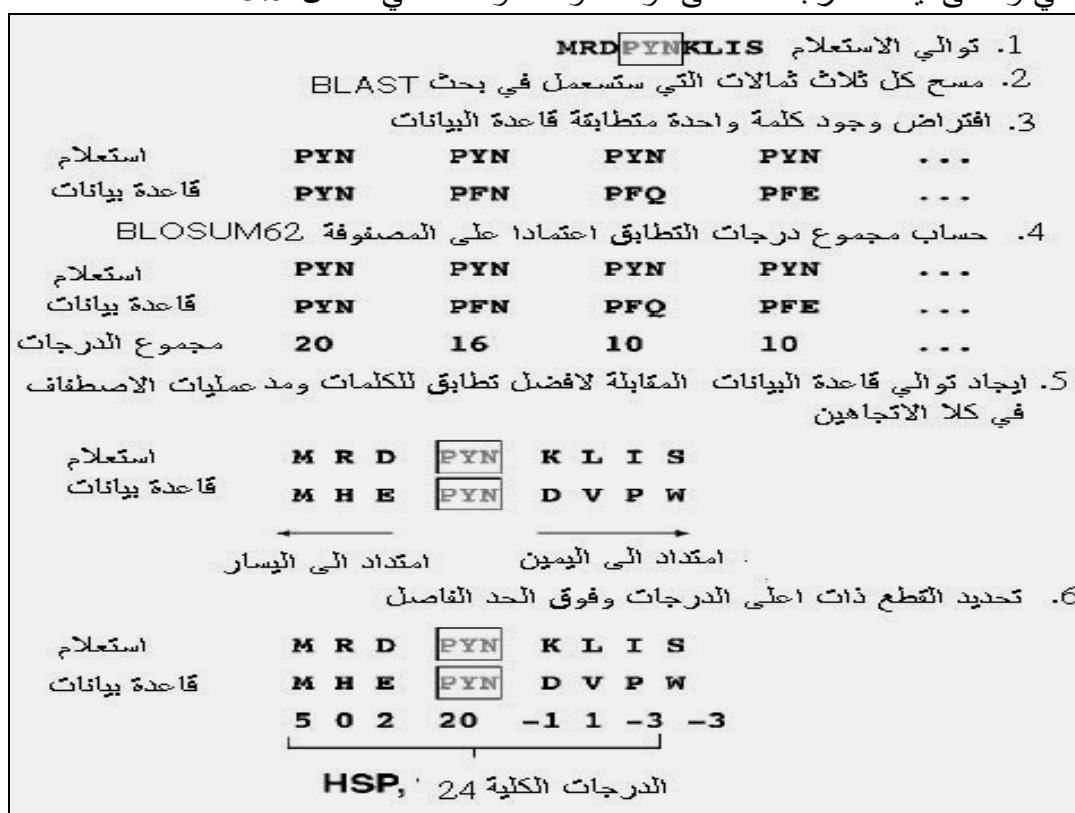
Basic local alignment search tool

طور البرنامج من قبل Stephen Altschul سنة 1990 وهو من النوع الموجه ومنذ ذلك الوقت أصبح واحداً من أكثر البرامج شهرة لتحليل وصف التواليات وكان البديل الأسرع للبرنامج المناظر له FASTA بدون التوضيح بالدقة. يستخدم البرنامج الموجه لاصطفاف توالي الاستعلام مع جميع التواليات في قاعدة البيانات، والهدف هو إيجاد أعلى درجات التطابق للقطع أو الأجزاء بدون فجوات بين التواليات المتقاربة. ووجود مثل هذه الأجزاء أكثر من حد معين يشير إلى التشابه بين زوج التواليات الذي يميز بين التواليات المتقاربة عن التواليات المتباعدة في قواعد البيانات.

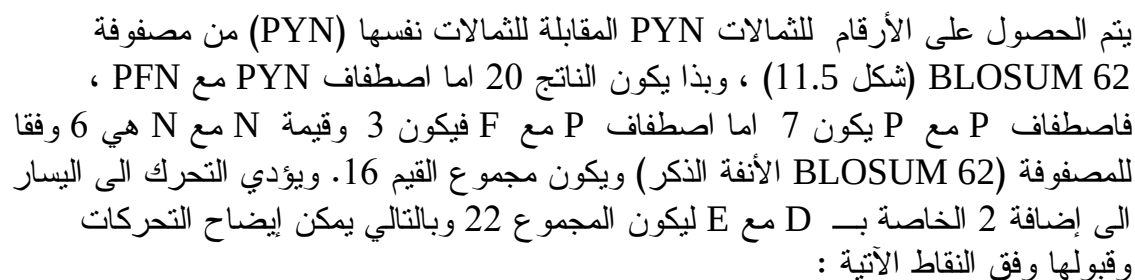
وتتم عملية انجاز الاصطفاف بهذا البرنامج بعدة خطوات هي:

- **الخطوة الأولى** يتم إنشاء قائمة من الكلمات Words من توالي الاستعلام وكل كلمة تتألف بصورة نموذجية من ثلاثة ثمالات لتواليات البروتين وإحدى عشرة ثمالة لتواليات DNA، وطول الكلمة يكون ثابتاً لمعظم تطبيقات البرنامج وتكون مثلى بالنسبة لتوالي الاستعلام وتواليات قواعد البيانات، ولا يغير الطول إلا في حالات خاصة مثلاً عند البحث عن المكررات أو تواليات الاستعلام القصيرة جداً والقائمة تضم كل الكلمات المحتملة في توالي الاستعلام، وتدعى هذه الخطوة "البذار".
- **الخطوة الثانية** يتم البحث عن الكلمات في التوالي الموجودة في قاعدة البيانات وهذه الخطوة تحدد تواليات قاعدة البيانات الحاوية على الكلمات.
- **الخطوة الثالثة** تبحث عن وجود الكلمات المتطابقة وتعطي درجات بواسطة مصفوفة الاستبدال وتؤخذ بنظر الاعتبار إذا كان عدد الكلمات المتطابقة أعلى من الحد الفاصل.

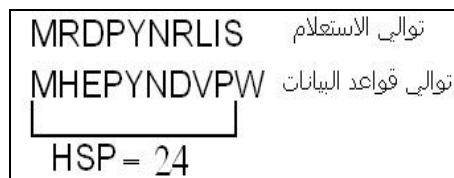
- **الخطوة الرابعة** تشمل عمليات مد اصطفاف زوج التواليات بالاتجاهين وفي أثناءها يتم تسجيل الدرجات باستعمال مصفوفة الاستبدال نفسها.
- وتستمر عملية الامتداد حتى تصل درجة عمليات الاصطفاف الى اقل من الحد الفاصل نتيجة لعدم التطابق ويكون الحد الفاصل للبروتينات 22 و 20 للـ DNA. وتدعى نتائج الاصطفاف المتجاور لزوج من القطع بدون فجوات بزواج القطع ذات الدرجات العالية High-scoring segment pair (HSP). وفي الإصدار الأصلي للـ BLAST كانت تعد HSP التقرير النهائي وتدعى ايضا الدرجة العظمى، والخطوات موضحة في الشكل 13.5 :



شكل 13.5 : البحث عن التطابق باستعمال BLAST



- 392



وقد طور برنامج BLAST لأداء عمليات الاصطفاف بوجود الفجوات ويتم اختبار أعلى درجة للقطع وتمد في الاتجاهين باستخدام البرمجة الديناميكية الحاوية على الفجوات. وتستمر عملية الامتداد ما دامت درجة الاصطفاف أعلى من درجة الحد الفاصل وبعدها تنتهي عملية الاصطفاف ولكن قد تكون مؤقتة لتعود فيما إذا ارتفعت قيمة الاصطفاف عن قيمة الحد الفاصل. ومثل هذه العمليات تكون لازمة قبل إعطاء التقرير النهائي لعملية الاصطفاف.

أنواع برامج BLAST

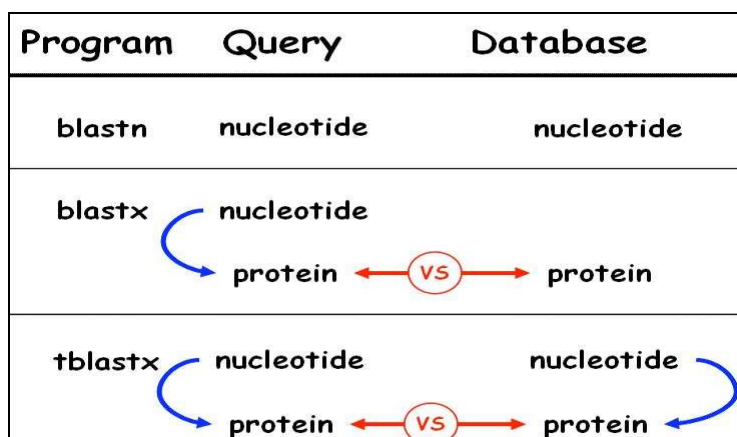
يتضمن برنامج BLAST مجموعة من البرامج المتخصصة الموجودة وربما تظهر إصدارات أخرى لأداء مهام أخرى ومنها:

BLASTN يستعمل لصف تواليات الاستعلام للنوكليوتيدات مقابل تواليات النوكليوتيدات الموجودة في قواعد البيانات المختلفة والتي تحدد أثناء تنفيذ البرنامج الذي يوفر عدة إمكانيات في هذا المجال.

BLASTP يستعمل لصف تواليات الاستعلام البروتينية مقابل تواليات البروتينات في قواعد البيانات المختلفة الخاصة بالبروتينات.

BLASTX يستعمل لصف تواليات الاستعلام النوكليوتيدية بعد ترجمتها إلى بروتينات باستعمال أطر القراءة الستة المحتملة مقابل تواليات البروتينات في قواعد البيانات المختلفة الخاصة بالبروتينات.

TBLASTN يستعمل في هذا البرنامج تواليات الاستعلام النوكليوتيدية التي ترجمت لكل احتمالات أطر القراءة الستة مقابل قواعد البيانات النوكليوتيدية الحاوية على التواليات المترجمة لأطر القراءة الستة ، يمكن اختصارها بالشكل 15.5 :



شكل 15.5 : أنواع برنامج BLAST

bl2seq برنامج يستعمل لإيجاد التشابه بين تواليان اثنتان فقط أحدهما هو توالي الاستعلام وآخر يتم اختياره من قاعدة البيانات لأغراض خاصة، كما في حالة البحث عن بروتين لكائن ممرض لبروتينات أخرى منتخبة من بروتينات المضيف لغرض إيجاد مدى التشابه، وعند عدم وجود التشابه يكون البروتين أو توالي الاستعلام هدفا دوائيا مهما فضلا عن استخدامه لأغراض أخرى. وفي هذا الموضع يذكر PSI-LAST المار الذكر (الفصل الرابع) ضمن سياق إيجاد النسق .

المعنوية الإحصائية لبرنامج BLAST

يكون تقرير برنامج BLAST على شكل قائمة من تطابقات التواليات المزدوجة من توالي الاستعلام وتوالي قاعدة البيانات مرتبة وفق القيم الإحصائية المعنوية كما في الشكل 16.5 :

Sequences producing significant alignments:			Score (Bits)	E Value	
gi 45024927 gb AY545216.1	Homo sapiens cytochrome P4502D6 (C...		2099	0.0	G
gi 181303 gb M33388.1 HUMCYP2D6	Human cytochrome P450 IID6 (CYP2		2099	0.0	G
gi 40806230 ref NG_003180.1	Homo sapiens cytochrome P450, fa...		2091	0.0	
gi 30524949 emb BX247885.1	Human DNA sequence from clone RP...		2091	0.0	
gi 181305 gb M33189.1 HUMCYP2DG	Human debrisoquine 4-hydroxyl...		2091	0.0	G
gi 17065905 emb AL021878.2 HS257I20	Human DNA sequence from c...		1873	0.0	E
gi 181320 gb M33387.1 HUMCYP8P	Human debrisoquine 4-hydroxyla...		1497	0.0	E
gi 20128060 ref NG_000853.1	Homo sapiens cytochrome P450, fa...		1465	0.0	G
gi 30336 emb X58467.1 HSCYP2D7A	Human CYP2D7AP pseudogene for cy		1465	0.0	E
gi 30337 emb X58468.1 HSCYP2D7B	Human CYP2D7BP pseudogene for cy		1449	0.0	G
gi 20128463 ref NG_000854.1	Homo sapiens cytochrome P450, fa...		1084	0.0	G
gi 30452 emb X07618.1 HSDB1A	Human mRNA for cytochrome P450 db1		904	0.0	U
gi 55732727 emb CR860959.1	Pongo pygmaeus mRNA; cDNA DKF2p47...		317	4e-83	G

شكل 16.5 : تقرير نتائج BLAST النصية

وتساعد الدرجات الإحصائية للتمييز بين التواليات ذات العلاقة التطورية والأخرى المتباعدة، وبصورة عامة يعرض التقرير التطابقات التي هي فوق الحد الفاصل. ولذلك كانت الحاجة الى مؤشر إحصائي يأخذ بنظر الاعتبار العدد الكلي لعمليات الاصطفاف وهنا العدد يتناسب طرديا مع حجم قاعدة البيانات وفي برنامج BLAST يكون الدليل الإحصائي ممثلا بقيمة التوقع Expectation value (E-value) وهي تمثل القيمة الاحتمالية للاصطفافات الناتجة من البحث في قاعدة البيانات الحاصلة بسبب الصدفة العشوائية، وقيمة E لها علاقة بقيمة P المستخدمة لتخمين عملية اصطفاف منفرد من زوج من التواليات ويقوم برنامج BLAST بمقارنة توالي الاستعلام مع كل التواليات في قاعدة البيانات وتحدد القيمة بالمعادلة الآتية :

$$E = m \times n \times p$$

اذ تمثل m مجموع أعداد الثمالات في قاعدة البيانات، n تمثل عدد الثمالات في توالي الاستعلام ، p تمثل احتمالية اصطفاف HSP نتيجة المصادفة العشوائية. فعلى سبيل المثال اذا كانت طول توالي الاستعلام 100 ثمالة والقاعدة المستعملة للمسح تحوي على 10^{12} ثمالة وقيمة p للمنطقة بدون فجوات ذات الدرجة HSP هي 10^{-20} فان قيمة E تكون حاصل ضرب القيم الثلاث:

$$E = 10^{12} \times 100 \times 10^{-20}$$

$$= 10^{-6}$$

وهذا يدل على ان احتمالية حدوث التطابق بين توالي الاستعلام وتواليات قاعدة البيانات عن طريق الصدفة العشوائية هو بمقدار $10^{-6} \left(\frac{1}{1000000}\right)$. وبذا فان قيم E الواطئة تعني ان التطابق العشوائي يكون اقل وبالتالي يكون التطابق أكثر معنوية . والتفسير التجريبي لقيمة E كما يأتي:

اذا كانت قيمة E اقل من 10^{-50} ($E < 10^{-50}$) فهذا يعني ان التطابق هو نتيجة للعلاقة الجذرية . واذا كانت قيم E بين 0.01 و 10^{-50} فيدل ذلك على وجود التماثل الجذري، اما اذا كانت قيم E بين 0.01 و 10 فان التطابق يكون غير معنويا" وان كان يمكن ان يعنى وجود علاقة تماثل ولكنها ضعيفة ومؤقتة وهي بحاجة الى دلائل أكثر لتأكيد هذه العلاقة ، في حين اذا كانت قيم E أكبر من 10 ($E > 10$) فهذا يدل على عدم وجود علاقة او ان هناك علاقة بعيدة جدا تقع خارج نطاق الطريقة المستعملة.

ونظرا لتأثر قيم E نسبيا بحجم قاعدة البيانات اذ تزداد القيم بزيادة حجم القاعدة المستعملة للمسح وبما ان العلاقة التطورية بين التوالين تكون ثابتة والتي تتناقص بشكل كبير عند استعمال قواعد بيانات كبيرة وهذا يعني فقدان التماثل او نقصانه بشكل كبير ولذلك كان لابد من بديل لقيم E، ومن هنا جاء استعمال درجة الشك Bit score إضافة الى قيم E كمؤشر إحصائي في تقرير برنامج BLAST. وقيمة درجة الشك (S') تقيس تشابه التواليات بغض النظر عن طول توالي الاستعلام وحجم قاعدة البيانات وتغيير القيمة استنادا الى درجة اصطفاف زوج من التواليات وتحدد بالصيغة الآتية:

$$S' = (\lambda \times S - \ln K) / n2$$

وتمثل λ ثابت توزيع كمبيل Gumbel distribution constant
S تمثل درجة الاصطفاف الخام

K يمثل الثابت الحاصل بمصفوفة الدرجة المستعملة

وهذا يشير الى ان درجة الشك العالية تعني التطابق العالي بين التوالين وذو معنوية عالية. وبذلك فان درجة الشك توفر دليل إحصائي ثابت للبحث في قواعد البيانات المختلفة في الحجم وكذلك البحث في قاعدة البيانات نفسها التي ربما تزداد في الحجم بمرور الزمن (انظر الشكل 16.5).

التعامل مع مناطق التكرار

يمكن ان تحتوي تواليات DNA على المكررات الكثيرة وهي شائعة في جينومات الخلايا حقيقية النواة، وكذا الحال مع تواليات البروتينات ، ويمكن ان تحوي على مكررات من ثمانية واحدة ومثل هذه المناطق تكون بتعقيد اقل Low complexity regions. وبطبيعة الحال فان هذه المناطق توجد في التواليات المودعة في قواعد البيانات وتشكل حوالي 15% من مجموع تواليات البروتينات المودعة في قواعد البيانات العامة . ووجود هذه المناطق في تواليات الاستعلام يؤدي الى ظهور تطابق زائف وغير صحيح وتؤدي الى ظهور درجات اصطفاف عالية زائفة للتواليات غير المترابطة ولتجنب هذه الحالة فلا بد من تصفية تواليات الاستعلام والتواليات في قواعد البيانات قبل إجراء عملية الاصطفاف ولذلك تجرى عملية الحجب عليها.

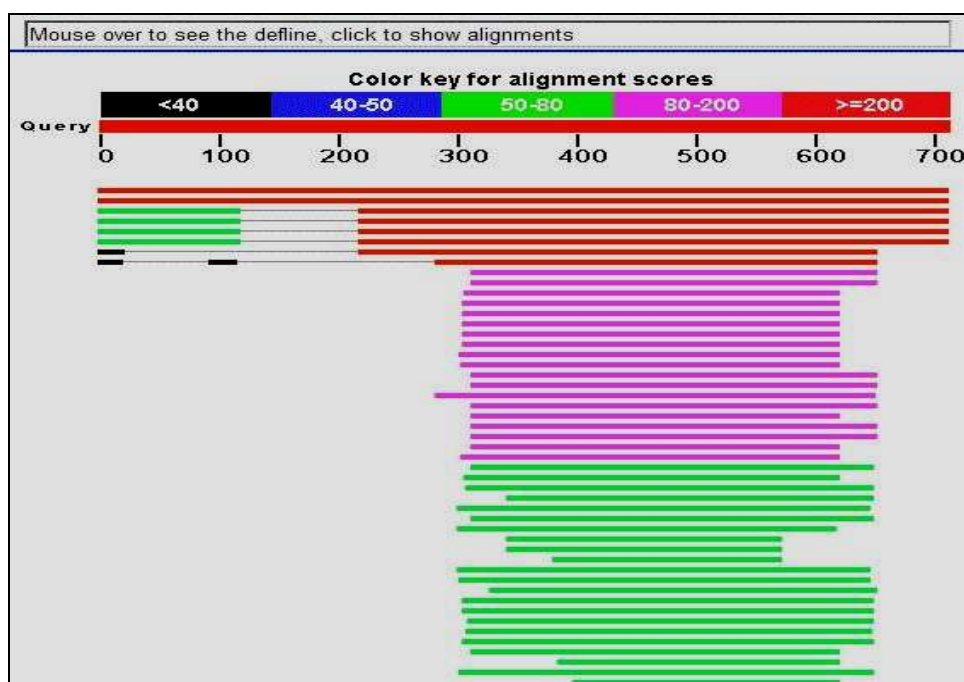
وتكون عملية الحجب على نوعين وهي Hard masking وذلك باستبدال الثمالات المتكررة بـ N للنوكليوتيدات و X لثمالات الحوامض الامينية وهذه الرموز الغامضة يتم إهمالها بالنسبة لبرنامج BLAST ولا تظهر في نتائج الاصطفاف وبالتالي تمنع ظهور النتائج الموجبة الكاذبة، ولكن لهذه الطريقة بعض المعوقات منها انخفاض درجات التطابق للتماثل الجذري نتيجة لقصر التواليات المعرضة لعملية الاصطفاف .

الطريقة الثانية هي Soft masking وفيه تحول التواليات المكررة المسببة للمشاكل الى حروف صغيرة والتي تهمل عند بناء قاموس الكلمات في برنامج BLAST (والتي تمثل الخطوة الاولى في عمل البرنامج كما ذكر سابقا) ولكنها هنا تبقى عاملة في عمليات الامتداد (الخطوة الرابعة في عمل البرنامج المذكورة سابقا) وكذلك تعمل في أمثلة عملية الاصطفاف.

ولذلك كان من الضروري تصفية التواليات الحاوية على المكررات قبل البدء بصفها والبحث عن مشابهاتها في قواعد المعلومات ويوجد عدد من برامج الحاسوب التي تقوم بهذه المهمة، ومن الجهة الثانية يفضل استعمال قواعد البيانات المنقحة والمشذبة مثل (nr) او ما تسمى بعض الأحيان Refseq database.

تقرير برنامج BLAST

ان مخرجات تقرير BLAST بعد عمليات الصف تكون مختلفة منها التصويرية كما في الشكل (17.5) ويظهر فيها توالي الاستعلام الى الأعلى ثم تظهر تواليات قواعد البيانات المتطابقة معه، فضلا عن احتواء التقرير على مدرج لوني لإيضاح درجات التطابق.



شكل 17.5 : تقرير برنامج BLAST الصوري

التدرج اللوني يسمح بالتشخيص السريع لمدى التطابق لتوالي الاستعلام مع تواليات قاعدة البيانات، وتتأثر الألوان درجة التطابق، الأحمر يكون الأكثر صلة واللونين الأخضر والأزرق معتدلة الصلة والأسود يشير إلى عدم التطابق والصلة. ويمثل طول الخط الأفقي وحدة اصطفاف التوالي نسبة إلى توالي الاستعلام. ويرتبط كل خط مباشرة مع مثيله من اصطفاف زوج التواليات بالصيغة النصية التي تكون مرتبة وفقاً لقيم E ويتكون الجزء النصي من التقرير قائمة بالمتطابقات، ثم تقرير حول الاصطفاف لكل عملية تطابق كما موضح في الشكل 18.5 :

Hit Accession Number B26183 of PIR database	
Hit Sequence Description : pol polyprotein – radiation murine leukemia virus	
Length of Hit Sequence = 1196	
P-value = 3.400000e-06	
HSP Alignments:	
Score: 133	Frame = 2 Identities = 36% Positives = 68%
2351 LPQGWKGSFAIFQSSMTKILEPFRKQNPDIYQYMDLYVGSLEIGQHRTKIEELR	2524
LPQG K SP++F+ ++ + L FR Q+PD+++ QY+DDL +++ E++ ++ L+	
308 LPQGFKNSTLTFDEALHRLADFRIQHPDLILLQYVDDLLLAATSELDCCQGGTRALLK	365
Score: 79	Frame = 2 Identities = 31% Positives = 73%
2054 IGPENPYNTPVFAIKKKDSTKWRKLVDFRELNRRRTQDF	2167
+ ++P NTP++++KK ++ R + ++RE+N+R +D+	
208 VPCQSPWNTPLLPVKKPGTNDYRPVQGLREVNKRVEDI	245
Score: 78	Frame = 2 Identities = 32% Positives = 54%
3917 KTIHTDNGSNFTSATVKAACWWAGIKQEFQIPYNPQSQGVVSMNKKELKKIIGQV	4081
+ + TDNG+ F S ++ GI + +Y PQS G VE MN+ +K + ++	

شكل 18.5 : نوع من تقارير BLAST حول الاصطفافات

ويتكون التقرير المعطى على ترويسة تحوي رقم الجين ، فضلاً عن وصف لنتائج الجين، وعدد الثمالات، أما المقطع الثاني فيكون خاص بالمعلومات الإحصائية لنتائج البحث مثل قيمة E وقيمة درجة الشك Bit ونسبة التطابق والفجوات. ثم وصف لعملية الاصطفاف إذ يكون توالي الاستعلام إلى الأعلى وتوالي قواعد البيانات تحتها، وبينهما تدون الثمالات المتطابقة المتناظرة في حين تعلم الثمالات المتشابهة ولكن غير المتماثلة بعلامة + .

برنامج FASTA

اسم البرنامج مشتق من FAST-All أي انه يستخدم لعمليات مقارنة تواليات البروتينات وكذلك تواليات النيوكليوتيدات السريعة. وهو في الواقع أول البرامج المستعملة للبحث عن التشابه بين تواليات الاستعلام وقواعد البيانات. ويستعمل البرنامج إستراتيجية أو دالة الكشف Hash function وهي طريقة لاختصار البيانات كما هو مستعمل في تسمية الشهور مثل (01 لشهر

كانون الثاني و 10 لشهر تشرين الأول) او كما في اختصار الأشهر التي تكون أيامها 31 وهي سبعة والأشهر التي أيامها 30 وهي أربعة أشهر بالربط بين عناصر مجموعة الأشهر مع مجموعة الأيام ضمن علاقات الرياضيات الحديثة. وتعتمد الدالة على جدول الكشف Look up table الذي يوضح بسهولة عملية التوازن بين عناصر المجموعات.

يستعمل برنامج FASTA الإستراتيجية المذكورة آنفا للبحث عن التشابه بين تواليات الاستعلام مع قواعد البيانات لطول معين هو K. ويتم عمل البرنامج بعدة خطوات:

الخطوة الأولى: تحديد الكلمات الصغيرة او الكتوب Ktup في التوالين المزمع صفهما باستعمال إستراتيجية الكشف ووضعها في جدول الكشف. فعند صف التوالين :

التوالي الأول A M P S D G L

التوالي الثاني G P S D N A T

يكون جدول الكشف لهما كآلاتي :

جدول 2.5 : الكشف عن تشابه الحوامض الامينية وفق برنامج FASTA

الحامض الاميني	موقع الحامض الاميني		التوازن
	التوالي الأول	التوالي الثاني	
A	1	6	-5
D	5	4	1
G	6	1	5
L	7	-	-
M	2	-	-
N	-	5	-
P	3	2	1
S	4	3	1
T	-	7	-

وتختلف مواقع الحوامض الامينية في التواليات وتحسب لها قيم التوازن بالآتي
موقع الحامض الاميني الالنين (A) في التوالي الأول - موقع الحامض الاميني في التوالي الثاني

فهي بالنسبة للالنين $A = 1 - 6 = -5$

والكتوبات التي تمتلك مقدار متساوي من قيم التوازن القيمة (1) في الجدول أعلاه يتم ربطها مع بعضها لتكون مناطق التواليات المتجاورة وهي تشابه بذلك الخطوط القطرية في المصفوفات النقطية . والكتوب Ktup يماثل الكلمات في برنامج BLAST. والكتوب اما تتكون من حرف واحد او اثنين من الحروف . والقيمة 2 تستعمل للبروتينات اما بالنسبة للـ DNA فتستعمل القيم من 1- 6 وأكثرها شيوعا والموصى بها هي 4 أو 6 وعندما تكون تواليات الاستعلام البروتينية او النيوكلوتيديّة قصيرة فلا مناص من استعمال الكتوب (حرف واحد). ويعد حجم الكتوب من المؤشرات المهمة في تحديد حساسية البرنامج والسيطرة عليها، فزيادة حجم هذه الكلمة يؤدي الى نقصان عدد التطابقات الأساسية.

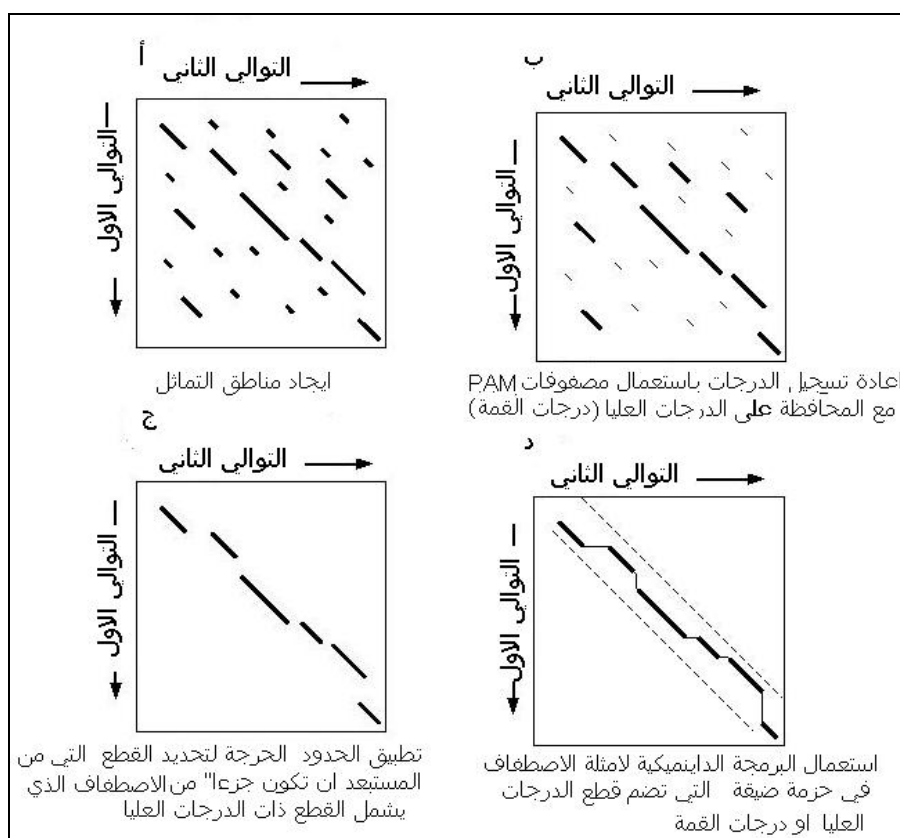
وتسجيل التطابق في الخطوة الأولى يستعمل مصفوفات PAM250 أو BLOSUM50. ويحفظ البرنامج عشرة مناطق ذات أفضل تطابق لاستعمالها في الخطوة اللاحقة.

الخطوة الثانية: يقوم البرنامج بإعادة مسح المناطق العشرة الناتجة من الخطوة الأولى باستعمال مصفوفة PAM250 وتسجل درجات التطابق وهذه تسمى الدرجات الأولية (Init1).

الخطوة الثالثة: عند وجود العديد من المناطق الأولية التي لها درجات أعلى من الحد الفاصل Cutoff value يعاد التأكد منها لغرض ربطها مع بعض بوجود الفجوات. وتحسب درجات التشابه من جمع درجات المناطق الأولية مطروحا منها جزء الفجوة ودرجات التشابه الأولي Initial similarity score (initn) تستعمل في ترتيب التواليات وفق التشابه.

وتكون عملية ربط المناطق المتشابهة في اصطفاف واحد يكون معتمدا على عملية البرمجة الديناميكية. ولا يتم الربط الا اذا كانت درجة التشابه اكبر من قيمة الحد الفاصل والتي تكون قيمتها أعلى بواحد من الانحراف المعياري لمعدل الدرجات المتوقعة من التواليات المتباعدة في المكتبة التي يكونها البرنامج أثناء العمل. وقد قدرت القيمة اللازمة لربط التواليات لتوالي استعلام مكون من 200 ثمانية وكتوب بقيمة 2 لتكون 28.

الخطوة الرابعة: تتم عملية الصف باستعمال البرمجة الديناميكية المتمثلة باستعمال خوارزمية Smith-Waterman للحصول على أفضل اصطفاف. والخطوات موضحة في الشكل 19.5:



شكل 19.5 : خطوات عمل برنامج FASTA

وحزمة FASTA (FASTA package) كما هو الحال مع BLAST تضم عدة برامج منها FASTAN للبحث عن تشابه التواليات مقابل قواعد النيوكلووتيدات وبرنامج EASTAP للبحث عن تشابه التواليات مقابل قواعد بيانات البروتينات وتوجد مواقع مزودة للحزمة للبحث عن تغايرات النيوكلووتيد المفرد SNP وآخر خاص بمكنون البروتين والعمليات الخاصة بالخياطة البديلة وآخر خاص بالجينومات وكثيرة غيرها.

المعنوية الإحصائية للـ FASTA

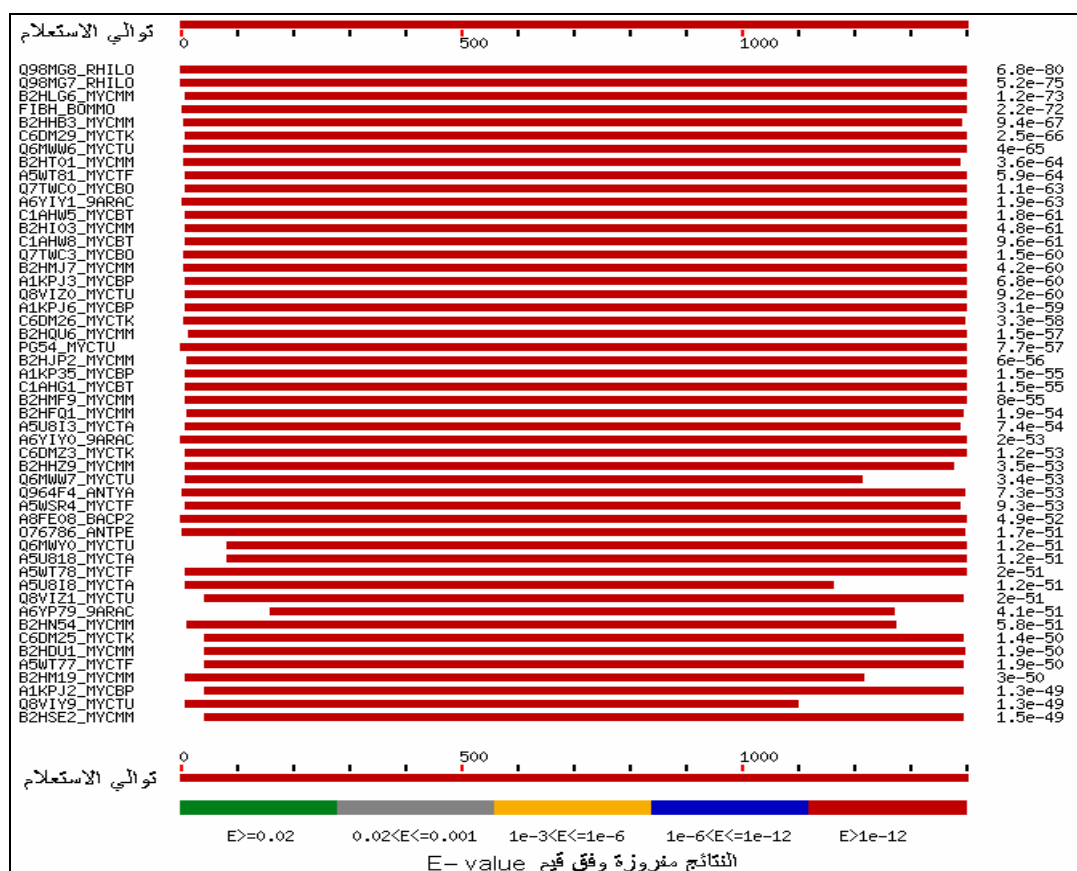
تحدد الأهمية الإحصائية لنتائج FASTA بعدد من القيم منها قيمة E ودرجة الشك وهي مشابهة لما في برنامج BLAST، وهناك قيمة أخرى تستخدم في هذا البرنامج وهي قيمة Z وهي قيمة معيارية وتوضح الأهمية الإحصائية لمدى اصطفااف توالي الاستعلام مع تواليات قواعد البيانات البعيدة عنها، والقيم المرتفعة للـ Z تعني معنوية أعلى لتطابق التواليتين، فإذا كانت درجة Z أكبر من 15 دل ذلك على أن هناك تطابق معنوي كبير مع علاقة تماثل جذري، أما القيم بالمدى 5-15 دل ذلك على احتمالية وجود تماثل جذري عالي والقيم أقل من 5 فتشير إلى علاقات ضعيفة.

تقارير برنامج FASTA

نتائج البرنامج يمكن أن تكون على شكل تقرير مصور يعتمد الألوان في تفسير وتحديد النتائج كما موضح في الشكل 20.5 .

ويظهر قيم E ، كما أن هناك دليل لوني لقيم E التي على ضوءها يحدد مدى التطابق وفرز التواليات أو ترتيبها.

الجزء الآخر من التقرير يكون نصي يظهر فيه تعريف بالتوالي وهويته وبعض القيم الإحصائية ، وتظهر اصطفاافات توالي الاستعلام وتوالي قاعدة البيانات، ويظهر التطابق بنقطتين عمودية في حين التشابه يشار إليه بنقطة، أما عدم التطابق فيترك فارغا بشكل مشابه نوعا ما لما ذكر أعلاه .



شكل 20.5 : النتائج المصورة لبرنامج FASTA

اصطفاف التواليات المتعددة (MSA) Multiple Sequence Alignment

في العديد من المهام التي تؤديها المعلوماتية الحيوية ومنها اصطفاف زوج من التواليات يكون غير كافياً لعدد من الأغراض، لذلك يتم اللجوء الى اصطفاف عدد من التواليات MSA وتعني صف ثلاث او أكثر من التواليات الحيوية مثل البروتينات او DNA او RNA وفي العادة تكون هذه التواليات ذات أصول مشتركة اي بينها علاقة تطورية سواء كانت قوية او متباعدة، لذلك تكون هذه المهمة ذات أهمية كبيرة في علم البايولوجي الجزيئي الحاسوبي لانها توفر معلومات بايولوجية أكثر من صف زوج من التواليات. ومن صف التواليات المتعددة يمكن التعرف على مدى التماثل وكذلك التشابه والتماثل الجذري. وزيادة درجات التشابه تؤدي في النهاية الى استنتاج التماثل الجذري وان التواليات تعود الى أصول مشتركة .

وعليه كان لاصطفاف عدد من التواليات أهمية في إيجاد العلاقات وأشجار التطور وذلك باستهداف صف تواليات الدومينات التي تقوم بالفعاليات وهذه تكون أهم من صف التواليات نفسها وذلك لان اصطفاف المناطق الثابتة (مثل الدومينات) التي تكون في الغالب مهمة من النواحي الوظيفية يمكن إيجادها بطريقة الاصطفاف المتعددة ، اذ تمكن الطريقة من التعرف على مواقع الفعاليات المهمة مثل مواقع الارتباط ، والمواقع الفعالة للإنزيمات وغيرها من المواقع النشطة في أداء وظائف معينة.

ويلعب هذا النوع من الاصطفاة دورا مهما في إيجاد التراكيب الثانوية والثلاثية للبروتينات وبذلك فهي تساعد في إيجاد العلاقة بين تركيب ووظيفة البروتينات. وتستعمل في تصميم البواديء، فضلا عن ان بعض برامجها يمكن ان تظهر الفجوات في التواليات وكذلك مواقع حدوث الطفرات او طفرات الحذف والإقحام (Indels).

آليات عمل الاصطفاة المتعددة ومعوقاتها

يعد اصطفاة التواليات المتعددة امتدادا طبيعيا لاصطفاة زوج من التواليات وتبحث العملية عن إيجاد أفضل التطابقات لعدد من التواليات وذلك بالبحث في قواعد البيانات . ومن الناحية النظرية يمكن استخدام البرمجة الديناميكية لصف عدد من التواليات الا ان هذا يحتاج الى وقت طويل كما انه يحتاج الى إيجاد مصفوفات بعدد التواليات المستعملة، فمثلا في صف زوج من التوالي يكفي إنشاء مصفوفة ثنائية الأبعاد (كما مر ذكره سابقا) ومع ثلاثة تواليات تكون الحاجة الى مصفوفة ثلاثية الأبعاد وتزداد الأبعاد مع زيادة عدد التواليات المستعملة لذلك فان البرمجة الديناميكية غير ممكنة فضلا عن زيادة الوقت اللازم لاداءها (كما ذكر آنفا) ومجال ذاكرة الذي تحتاجه، ففي حالة الاصطفاة الشامل لزوج من التواليات يكون الوقت ممثلا بالعلاقة $O(L^2)$ اذ تمثل L طول التوالي و O تمثل تعقيد الحاسوب Computational complexity بالنسبة للوقت وما تحتاجه من الذاكرة للخرن ، وهذا له علاقة بنوعية الحاسوب أي جيل الحاسوب وسعة الذاكرة .

ولكن في حالة اصطفاة عدد كبير من التواليات N فان الوقت يصبح ممثلا بالعلاقة $O(L^N)$ ففي الحالة الأولى اي عند صف زوج من التواليات الحيوية مثل البروتينات او الجينات التي يكون فيها الطول اقل من 10,000 ($L < 10,000$) يكون الوقت اللازم معقولا ومقبولا ولكن عند استعمال عدد كبير يكون عامل الوقت غير عملي فضلا عن الحاجة الى حيز كبير من ذاكرة الحاسوب . لذلك استعملت خوارزميات الطرق الموجهة بشكل أساسي او بالتناوب مع طرق أمثلة الاصطفاة مثل طريقة البرمجة الديناميكية.

واغلب طرق او خوارزميات الاصطفاة المتعدد والتي تعمل في مجال الطرق الموجهة تعتمد أساسا على اصطفاة لزوج من التواليات ومنه تنفرع الاصطفاة المتوالية لباقي التواليات وغيرها من التحويلات الجانبية ففي معظم البرامج والتي أكثر تمثيلا لعمليات الاصطفاة المتعدد مجموعة برامج Clustal، يتم صف زوج التواليات الذي يؤدي الى إنتاج عدد من الاصطفاة الممثلة بالعلاقة الآتية:

$$N(N-1)/2$$

ثم تستعمل درجات هذه الاصطفاة لبناء شجرة ثنائية للعلاقة التطورية بينها ثم يقوم برنامج الاصطفاة ببناء الاصطفاة للتواليات المتعددة بترتيب يعتمد على شجرة التطور هذه التي تسمى في بعض الأحيان بشجرة الدليل Guide tree ففي البداية يتم صف التواليات المتباعدة من الناحية التطورية وتليها التواليات الأخرى والذي يؤدي الى إيجاد نسق الاصطفاة

$$N/2$$

ثم يتم صف النسق الناتج مع بعضه البعض مؤديا الى إنتاج نسق اصطفاة $N/4$ وهكذا الى ان يتم اصطفاة جميع التواليات المستعمل عنها (N). ان عملية إيجاد الدرجات وبناء الأشجار التطورية الثنائية هو $O(NL^2)$ (التي تمثل Computational complexity للوقت والحيز

لدخول عامل طول التوالي فيها) ولبناء الشجرة تستعمل الصيغة $O(L^2 \log N)$ ثم تستعمل الخوارزمية لإيجاد وقت متعدد الحدود Polynomial time الذي يمكن ان يستعمل لاصطفاف مئات من التواليات في وقت مقبول.

وقد تعاني هذه البرامج من بعض الجوانب السلبية نظرا لحدوث خطأ في عمليات الاصطفاف المبكرة ولذلك يمكن ان تغلق العملية، وعليه فان هذه الخوارزميات يمكن ان لا تعطي ضمان للحصول على أفضل اصطفاف. وقد طورت العديد من الخوارزميات لاصطفاف التواليات المتعددة ومعظمها يعتمد على الاصطفاف الموضعي لان استعمال الاصطفاف الشامل لعدد كبير من التواليات يكون غير عملي، كما ان بعض الخوارزميات عمدت الى معالجة بعض المشاكل منها ما هو متعلق بوجود الفجوات او Indels التي عولجت في إصدارات البرامج المختلفة، كما ان بعضها يعطي إشارات للتنوع الحيوي كما في إشارات بعض نتائج البرامج مثل (-). وهذه البرامج تعاني أيضا من المشاكل الدائمة التي تستعمل طريقة الاستقراء الرجعي Bottom-up او ما يسمى بتسلق التل (Hill-climbing).

دالة الدرجة Scoring function

ان اصطفاف التوالي المتعددة يعمل على ترتيب التوالي بطريقة بحيث تعظم عدد الثمالات المتطابقة من كل توالي استنادا الى دالة الدرجة. تعتمد دالة الدرجة لاصطفاف التواليات المتعددة على فكرة مجموع الأزواج Sum of pairs (SP) في احتساب درجات دالة الدرجة SP، حيث يحتسب كل عمود درجاته بواسطة تجميع الدرجات لجميع تطابقات زوج التواليات الممكنة، وتحديد درجات الفجوات وعدم التطابقات. تمثل الدرجة للاصطفاف الكلي مجموع جميع درجات الأعمدة (جدول 3.5) ان الهدف من استخدام خوارزميات اصطفاف التواليات المتعددة هو لتحقيق أعظم درجات لدالة الدرجة SP.

جدول 3.5 : احتساب دالة الدرجة SP

seq. 1	G	K	N	(G,T)=-2	(K,R)=2	(N,N)=6
seq. 2	T	R	N	(T,S)=1	(R,H)=0	(N,E)=0
seq. 3	S	H	E	(G,S)=0	(K,H)=-1	(N,E)=0
Total				-1	1	6
المجموع الكلي	حاصل الجمع $(-1)+(1)+(6)=6$					

لو فرض ثلاث تواليات يرام إجراء اصطفاف توالياتها المتعددة فتحتسب مجموع الدرجات الخاصة بتمائل كل مواقع أزواج التواليات. وان هذه الدرجات تعتمد على مصفوفة BLOSUM62. وان مجموع الدرجات للاصطفاف يساوي 6. والتي تعني ان الاصطفاف $2^6=64$ مرة من احتمالية حدوث تماثل جذري بين التواليات مقابل حدوثه نتيجة المصادفة.

الطرق المستعملة للاصطفافات المتعددة

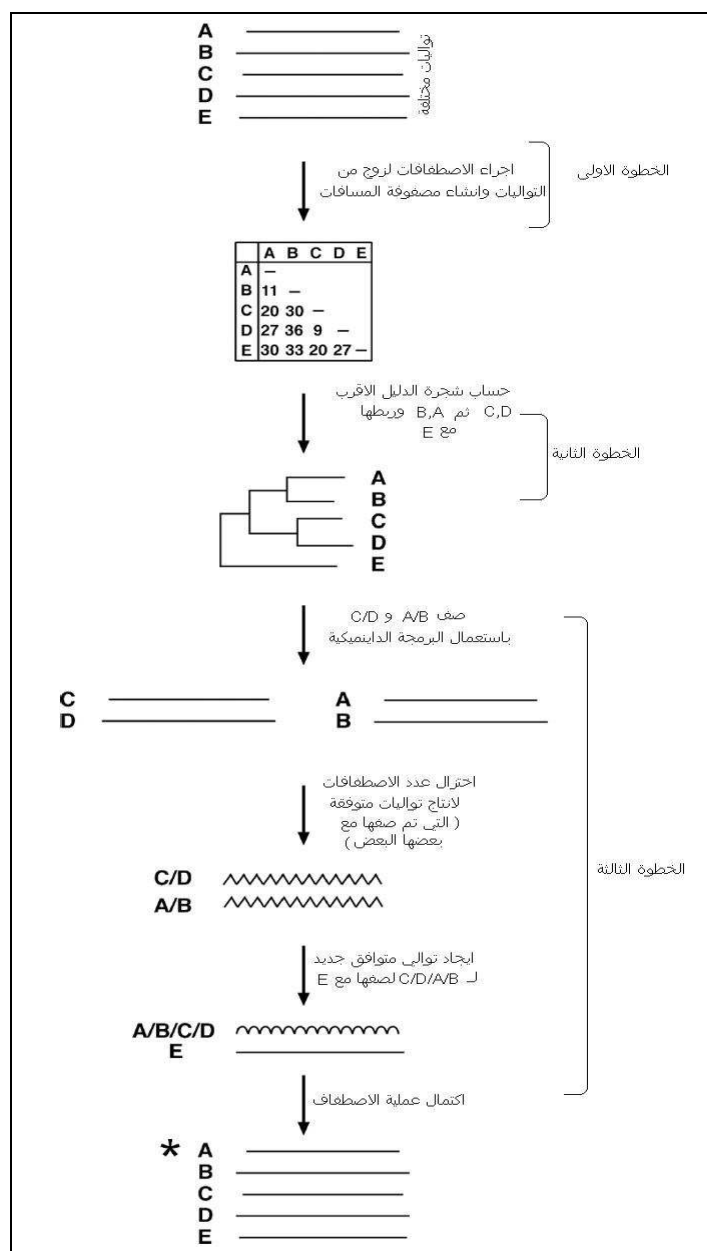
تكون الطرق بصورة عامة اما متزامنة اي يتم صف كل التواليات في وقت واحد وتكون هذه أفضل من النقاط الإشارات البايولوجية المهمة في عوائل التواليات المتباعدة والتي تكون محط التركيز في حالة اصطفافات زوج من التواليات او تكون الطرق مجزأة وهي الأعم. ونظرا

لعدم إمكانية استعمال البرمجة الدائيميكية في اصطفاف التواليات المتعددة كما ذكر آنفاً، تم اللجوء الى استعمال الطرق الموجهة ومنها.

طريقة الاصطفاف المتدرج Progressive alignment method

وتسمى ايضا طريقة الشجرة او الطريقة الهرمية وتعتمد كباقي الطرق على عملية اصطفاف زوج من التواليات لتبني عليه باقي الاصطفافات وتتم بعدة خطوات كما موضح في الشكل

21.5



شكل 21.5 طريقة الاصطفاف المتدرج

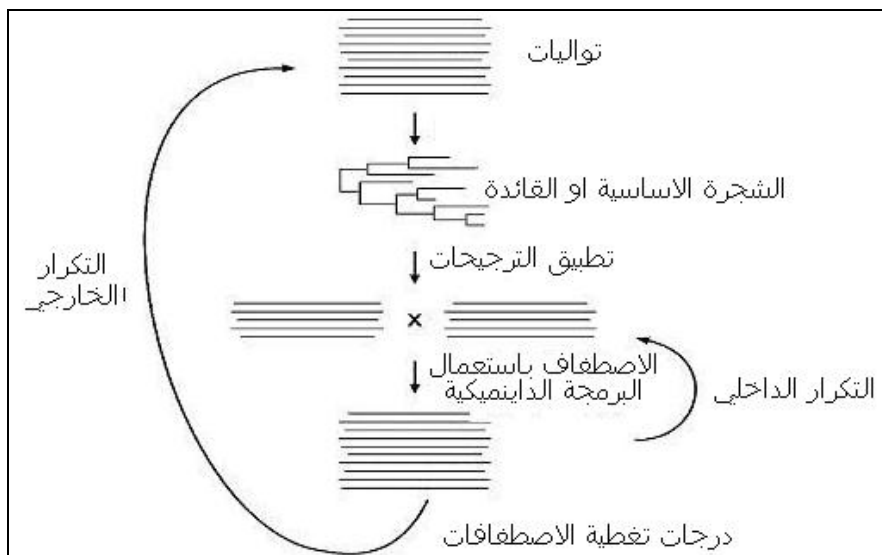
- كما تظهر في برامج Clustal (راجع شكل 24.5 أ)
- **الخطوة الأولى :** يتم صف زوج من التواليات الملائمة بطريقة الاصطفاف الشامل باستعمال خوارزمية Needleman – Wunsch وتسجيل درجات التشابه او النسبة المئوية للتشابه اعتمادا على مصفوفات استبدال محددة ، وكل من القيم المسجلة تشير الى المسافة التطورية بين التواليات، وتحول القيم المسجلة الى قيم مسافات لإنشاء مصفوفة المسافة بين كل التواليات المشمولة بالدراسة. لذلك تكون الخطوة الأولى مهمة جدا وتستعمل فيها التواليات ذات العلاقة الأكبر وذلك لأنها حساسة لأي اختلال في الدقة في عمليات اصطفاف الزوج الأول من التواليات.
- **الخطوة الثانية :** بناء شجرة الدليل ويتم بتحليل بسيط لعلاقات التطور استنادا الى مصفوفة المسافات التي حددت في الخطوة الأولى لأزواج من التواليات باستخدام طريقة ربط الجوار Neighbor-joining method او استعمال طريقة المعدل الرياضي غير المرجح Unweighted pair Group Method using Arithmetic Average (UPGMA) والشجرة تمثل علاقة التطور لجميع التواليات وهناك حاجة للتأكد من حقيقة هذه العلاقات نظرا لكون الشجرة هي تمثيل تقريبي اذ ليس هناك صيغة معينة وثابتة لبناء أشجار العلاقات التطورية . ولكن يمكن استخدام شجرة الدليل هذه لصف التواليات.
- **الخطوة الثالثة :** يتم صف التواليات الأكثر تقاربا اعتمادا على شجرة الدليل باستخدام خوارزمية البرمجة الديناميكية ويتم التعامل مع ناتج هذا الاصطفاف كتوالي واحد في خطوة الاصطفاف اللاحقة اذ يتم صفها مع التوالي الأقرب اعتمادا على شجرة الدليل باستخدام البرمجة الديناميكية. والطريقة لا تكون مثالية للاصطفاف الشامل، كما ان الخطأ في الاصطفاف الأول يتضخم بشكل كبير في النهاية. وبما ان الطريقة تستعمل النهج الموجه لذلك ليس هناك ضمان كبير في انها تعطي الاصطفاف الأمثل وتبقى مصداقيتها البايولوجية مشكوك فيها. كما انها في الأساس طريقة اصطفاف شامل لذلك فهي غير ملائمة لاصطفاف التواليات المختلفة الطول، ونتيجة لاستخدام جزاءات الفجوات المتصلة في الطريقة فان الفجوات الطويلة غير مسموح فيها ونظرا لاعتمادها على صف زوج من التواليات فان اي خطأ في هذه الخطوات لا يمكن تصحيحه وبالتالي يتضخم بالخطوات اللاحقة وهذا الخطأ يبدو أكثر وضوحا عند التعامل مع التواليات المتباعدة اي بمعنى آخر ان الطريقة بعد الاصطفافات الأولية التي تدخل نتائجها في عملية الاصطفاف سوف لا تؤخذ بنظر الاعتبار في العمليات اللاحقة وهذا يزيد من كفاءة الطريقة ولكن على حساب الدقة وهناك طريقة مقاربة لها هي طرق الاصطفاف شبه المتدرج Semi progressive methods تعتمد الى ترك الطريقة الموجهة غير المنضبطة Lossy heuristic لتحسين الأداء وتستعمل مع تواليات خاصة. والطريقة يمكن ان تستوعب صف 100-1000 توالي ومن أفضل برامجها Clustal ومشتقاته وإصداراته المختلفة، وكذلك مجموعة برامج T-Coffee وهي برامج أبطأ من المجموعة الأولى الا انها أكثر دقة خاصة بالنسبة للتواليات المتباعدة لانها تعتمد على إجراء الاصطفافات الموضوعية والشاملة لزوج من التواليات وكذلك لجميع الأزواج المحتملة .

طريقة الاصطفاف التكراري Iterative alignment methods

طريقة سريعة اي تعتمد النهج التوجيهي ويمكن ان تستعمل في صف اي عدد من التواليات وأدخلت في العديد من البرامج واستعملت لمعالجة الأخطاء والضعف في طريقة الاصطفاف المتدرج وسميت تكرارية لأنها تعمل بشكل مشابه لطرق الصف المتدرج وتعاد عمليات الاصطفاف للتواليات الأولية مع عمليات إضافة التواليات الجديدة، ونتائجها مقبولة في اصطفاف التواليات المتوسطة البعد عن بعضها وعند استعمالها للعوائل او المجاميع المتباعدة فهي تعتمد الى استعمال الاصطفاف الموضعي وليس الشامل.

تعتمد طريقة التكرار على فكرة ان الحل الأمثل يمكن ان يوجد بواسطة تكرار الحلول المثلى الجزئية وتبدأ الطريقة بإجراء اصطفاف رديء النوعية والذي يتم تحسينه وتطويره بواسطة إعادة الاصطفاف الى حين الوصول الى الحد الأمثل وان كانت الطريقة توصل الى ما دون الحالة المثلى وكما ذكر أنفا أنها تعتمد النهج التوجيهي لذلك لا ضمان في انها تجد الاصطفاف الأمثل. وذلك لأنه في العادة يوجد أكثر من اصطفاف أمثل لزوج التواليات وهذا يجب ان يؤخذ بنظر الاعتبار نظرا لحدوث الطفرات في مناطق مختلفة من الجزيئات الحيوية التي يصعب تعديلها خاصة عند استعمال صف زوج التواليات او الاصطفاف الثنائي لتبني عليه عمليات الاصطفاف المتعدد وعليه فان الأخيرة تكون حساسة لتغيير المؤشرات في الاصطفاف الثانوي خاصة عند صف التواليات المتباعدة او العوائل الثانوية من التواليات. وتستخدم طريقة التكرار المتداخل للاصطفافات المتعددة التي تتجز بمجموعتي التكرار الخارجي والتكرار الداخلي.

في التكرار الخارجي يتولد اصطفاف عشوائي أولي يستخدم لإيجاد تكرار شجرة بطريقة المعدل الرياضي غير المرجح UPGMA لبناء شجرة العلاقة التطورية ، اما في مرحلة التكرار الداخلي والذي يقسم الى أكثر من دورة او مرحلة وفي المرحلة الأولى تقسم التواليات الى مجموعات عشوائية بعد ذلك تثبت مواقع الاصطفاف لكل مجموعة ، وتعامل كل مجموعة على انها توالي مفرد لتجري لها عملية اصطفاف باستخدام طريقة أكثر انضباطا وهي البرمجة الديناميكية الشاملة . وتكرر العملية الى حين الوصول الى درجة مجموع الأزواج الثابتة وعندها يستخدم ناتج الاصطفاف لبناء شجرة جديدة باستخدام طريقة المعدل الرياضي غير المرجح ، ثم تطبق قيم ترجيحية جديدة لتحديد درجات الاصطفاف المثلى ، ويستعمل الاصطفاف الأمثل الجديد لإعادة الاصطفاف مرة أخرى في التكرار الداخلي ، وتكرر العملية لعدة مرات الى حين الوصول الى مرحلة عدم حصول اي تحسن لدرجات الاصطفاف الكلية. وهذه الطريقة تستعمل في البرامج المعتمدة على الطريقة مثل PRN. وأساسيات الطريقة موضحة في الشكل 22.5 :

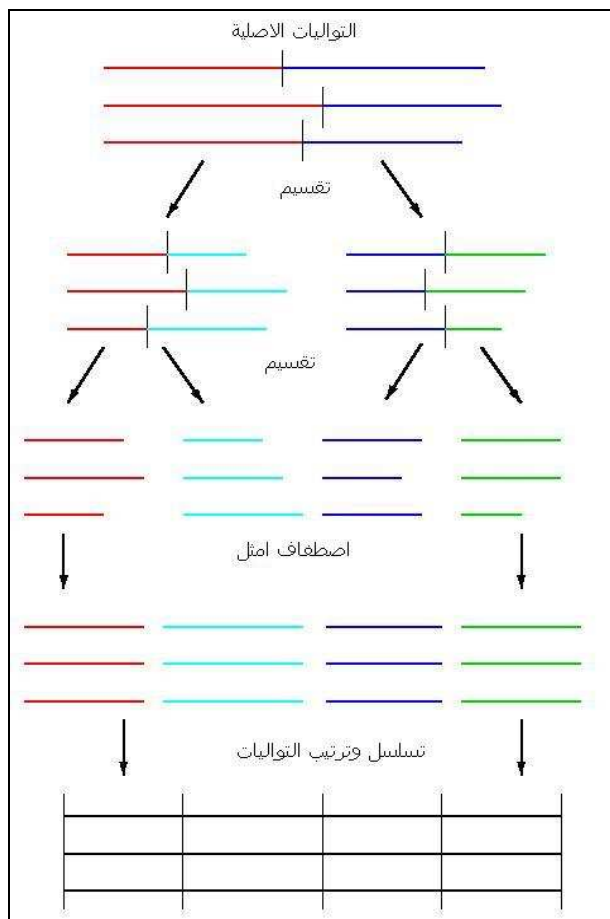


شكل 22.5 : طريقة الاصطفاف التكراري

وهناك بعض البرامج التي تستعمل الطريقتين أي الاصطفاف المتدرج و الاصطفاف التكراري لتحسين الأداء مثل CHAOS/DIALIN و -log Multiple sequence alignment by expectation (MUSCLE) .

طريقة القطع والتسلسل (DCA) Divided and conquer alignment

طريقة سريعة من مجموعة النهج الموجه تستعمل لصف تواليات لجزيئات حيوية مختلفة مثل البروتينات وعوائل RNA وتواليات أخرى. وأكثر البرامج المعتمدة على هذه الطريقة تستخدم مصفوفات استبدال مختلفة مثل PAM و BLOSUM و Gonnet وغيرها وبدرجات مختلفة. تعتمد الطريقة الى قطع التواليات في مناطق محددة غالبا ما تكون في المركز والتي يفترض ان تكون بطول مقبول مثلا 40 ثمانية ($L=40$) لغرض صفها بشكل امثل وذلك لان التواليات الطويلة ($L \geq 100$) تؤدي الى إطالة وقت تنفيذ البرنامج وكذلك زيادة المساحة المحجوزة في ذاكرة الحاسوب كما ان التواليات القصيرة ($L < 5$) فإنها تؤدي الى إيجاد فراغات في نهاية الاصطفافات اي انها تنتج اصطفافات رديئة النوعية. وفي حالة التواليات الطويلة يتم قطعها مرة ثانية لحين الوصول الى الطول المقبول المتمثل بالقيمة Z (Stop length) وتمثل الحد الفاصل الذي تتوقف عنده عملية القطع ، ومواقع القطع يتم تحديدها بمعادلات رياضية لان القطع غير الصحيح يؤدي الى نتائج غير صحيحة او غير ملائمة لعمليات الاصطفاف النهائية. والخطوة التي تلي عملية التقطيع يتم فيها عملية الاصطفاف بطرق تحاول الوصول الى الاصطفاف الأمثل ، ثم بعد ذلك تجمع ضمن خطوة التسلسل Concatenate لتعطي الاصطفاف النهائي كما موضح في الشكل 23.5 :



شكل 23.5 : طريقة القطع والتسلسل

والنتائج المعطاة بهذه الطريقة جيدة ولكن ليست المثلى خاصة عند صف عدد كبير من التواليات الطويلة. وفي هذه الطريقة يتأثر الوقت اللازم بطول التوالي وكذلك مدى تشابه التواليات الاستعلام، فضلا عن حيز الذاكرة المستعملة والجدول التالي (4.5) يوضح نتائج إحدى التطبيقات لتواليات مختلفة:

جدول (4.5): علاقة طول التوالي بالوقت وحجم الذاكرة عند استعمال طريقة DCA لأنواع من البروتينات.

البروتين	طول التوالي	قيمة Z	الوقت (ثانية)	حجم الذاكرة (MB)
قطع قصيرة				
1	94-76	64	58.7	29
2	106-89	64	152.1	32
3	109-95	128	11.5	8
4	117-108	64	28.2	10
5	104-98	64	5.8	8
6	110-98	64	6.6	9

قطع متوسطة				
30	97.7	128	237-201	7
?	> 9000	64	263-224	8
10	20	128	213-203	9
215	2342.4	128	259-237	10
10	11.0	128	216-206	11
9	19.1	128	254-242	12
قطع طويلة				
411	1920	128	474-304	13
16	98	32	572-435	14
629	2595	256	483-421	15
229	2046	128	482-452	16
24	31	256	447-424	17
684	4737	512	541-526	18

وتختلف البروتينات المذكورة في أصول اشتقاقها ودرجة التشابه بينها. ولذلك يلاحظ ان بعض البرامج تحدد في الواجهة أقصى ما يمكن تحمله من استعمال الذاكرة اذ ان بعض برامج هذه المجموعة يمكن ان تشغل 500 MB وتمتتع بعض مزودات الخدمة بالقيام بالعملية وتنبه الى الحد الذي يمكن ان يتوقف عنده البرنامج ، او تطلب في حالة كون عملية الاصطفاف تستغرق وقت طويل وطلب البريد الالكتروني لإرسال النتائج بعد الحصول عليها. وهناك برامج وطرق أخرى يمكن ان تستعمل في اصطفاف التواليات المتعددة منها طريقة الاصطفاف على أساس القطاعات Block-based alignment وكذلك استعمال نماذج ماركوف المخفية وهي نماذج احتمالية يمكن ان تستعمل بعد تحسين سرعتها وتستعمل بصورة خاصة مع التواليات الحاوية على مناطق متداخلة، وكذلك التواليات المتباعدة وتكون قليلة الحساسية للضوضاء الناتجة عن الاستبدالات الثابتة وشبه الثابتة.

اصطفاف التراكيب Structural alignment

هذا التوجه يكون خاص بالبروتينات وبعض الأحيان بتواليات RNA ، وفيه تستعمل المعلومات المتوفرة حول التركيب الثانوي والثلاثي للجزيئات الحيوية للمساعدة في عمليات الاصطفاف، وتستعمل لاصطفاف اثنين او أكثر من التواليات. ونظرا لاعتمادها على ما متوفر من المعلومات حول التركيب لذلك تستعمل للجزيئات المعروفة التركيب المعتمدة على دراسات الأشعة السينية والرنين النووي المغناطيسي .

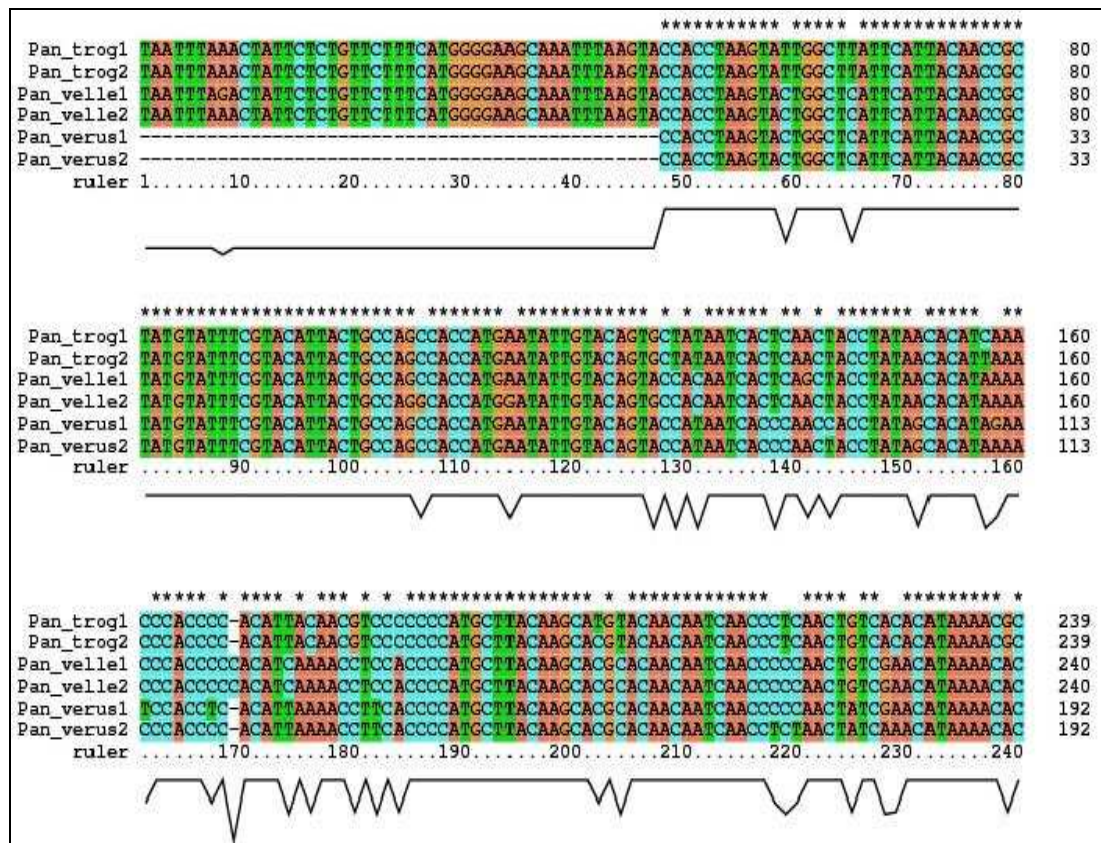
ولكون تركيب البروتين وجزيئات RNA هو أكثر ثباتا من حيث التغيرات مقارنة بالتواليات الخام (التركيب الأولي) لذلك تكون نتائج هذه الاصطفافات أكثر موثوقية خاصة بين التواليات البعيدة جدا عن بعضها والتي انفجرت بشكل كبير والتي تكون مقارنة توالياتها الأولية غير موثوقا بها عند كشف التشابه.

ويستعمل اصطفاف التركيب كطريقة قياسية في تقييم الاصطفافات المعتمدة على التماثل الجذري في حالة التنبؤ عن تركيب البروتينات نظرا لوضوح اصطفاف المناطق لتواليات

البروتينات التي تكون متشابهة التركيب وعدم الاعتماد على معلومات التواليات الأولية بشكل حصري. وتوجد العديد من البرامج التي تظهر هذه الاصطفافات.

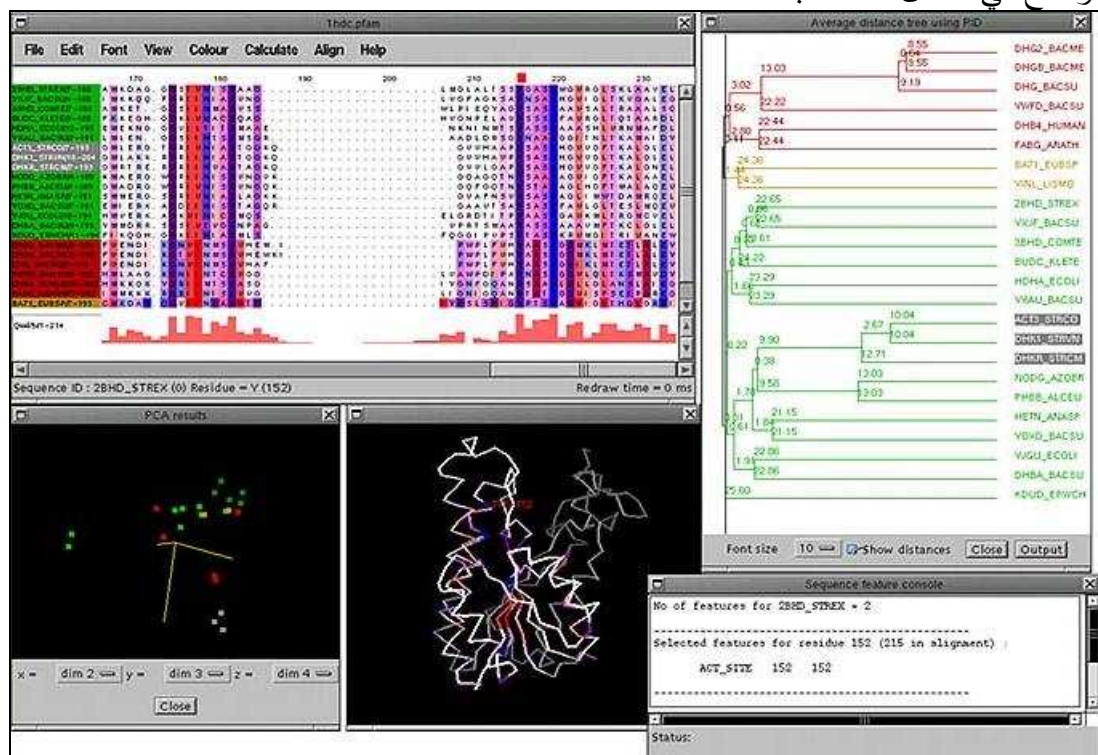
طرق إظهار اصطفافات التواليات المتعددة

تكون عمليات إظهار Visualization النتائج مهمة خاصة بالنسبة لاصطفاف التواليات المتعددة التي تتبع نهج الطرق الموجهة حيث يتوقع وجود أخطاء فيها لذلك تحوي العديد من البرامج على مظهر للصور Multiple sequence alignment viewer والذي يكون بمثابة تصديق لعمليات الاصطفاف ، وبذا يمكن ملاحظة الهفوات التي قد تحويها المخرجات النصية. وتختلف مخرجات البرامج في طريقة عرض اصطفاف التواليات المتعددة وأكثرها تعتمد الألوان (في حالة البروتينات) اذ تكون الألوان خاصة بكل مجموعة من الحوامض الامينية اعتمادا على الصفات الفيزيوكيميائية. اما القواعد النروجينية فتكون لها ألوان ثابتة في معظم البرامج ، كما ان كل برنامج له بعض العلامات الخاصة وطرق تمثيل التوافق والشكل التالي(24.5) يوضح اصطفاف مجموعة من التواليات النيوكلوتيدية باستعمال برنامج Clustal X وهو من أشهر البرامج المستعملة في هذا المجال

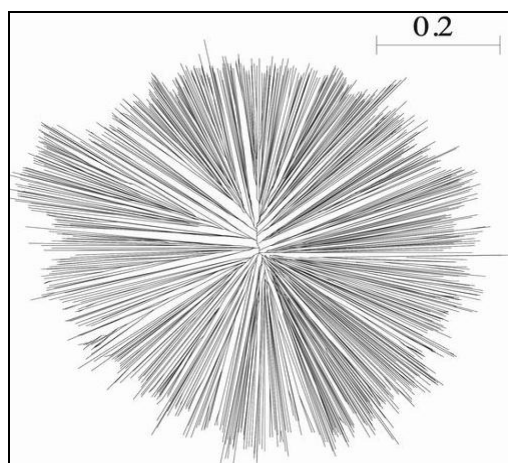


شكل 24.5 (أ) : اصطفاف مجموعة من التواليات النيوكلوتيدية باستعمال برنامج ClustalX

وتظهر حالات التطابق بشكل خط أفقي مستقيم تحت التواليات اما في حالة البروتينات فان بعض البرامج تظهر أكثر من صورة مع عملية التطابق مثل شجرة التطور والمسافات الخاصة بالتواليات فضلا عن إظهار معلومات إضافية وكذلك إظهار التراكيب الثانوية والثلاثية كما موضح في الشكل 24.5 ب :



شكل 24.5 (ب) : إظهار مع معلومات الإضافية مع عملية الاصطفاف متمثلة بشجرة التطور والمسافات الخاصة والتراكيب الثانوية والثلاثية وإضافة الى ما ذكر أعلاه يمكن ان تظهر الاصطفافات بشكل مركزي كما موضح في الشكل 25.5 :



شكل 25.5 : مخرجات برامج الاصطفاف بالطريقة المركزية

معيار التصديق لاصطفافات التواليات المتعددة

خوارزميات الاصطفاف والبرامج وغيرها من الشؤون يمكن ان تعبير او يصادق عليها باستعمال معيار مرجعي Benchmark reference ، فهناك بعض القياسات الثابتة التي تستعمل كمرجع للمقارنة وهذه المعايير قد تكون قياسات اما ما يسمى محكات Touchstones او مؤشرات وهي اي مرجع تقاس اليه الأشياء الأخرى ويستعمل للمقارنة والتقييم. وفي مجال عمليات الاصطفاف والبرامج في المعلوماتية الحيوية فان المعايير المرجعية هي قواعد بيانات خاصة مثل BALiBASE الخاصة باصطفافات التواليات المتعددة وتحتوي على بيانات للاصطفافات التركيبية المستقاة من العديد من المراجع وتعد من المقاييس التي تقارن بها عمليات اصطفاف التواليات.

العلاقات التطورية Phylogenetic relationships

يعد تحديد العلاقات التطورية بين الأحياء أحد المهام الرئيسية في المعلوماتية الحيوية ويعتمد تحليل التواليات للجزيئات الحيوية الأساس في معرفة عملية التطور ورسم أشجار التطور. ويوحى التشابه او الابتعاد بين التواليات الحيوية ذات العلاقة الذي يحل في إطار الأشجار التطورية الى هذه العلاقات، والى المدى الذي مرت به التواليات من تغيير وتطور فمثلا التوالي الآتي:

acctga → agcta

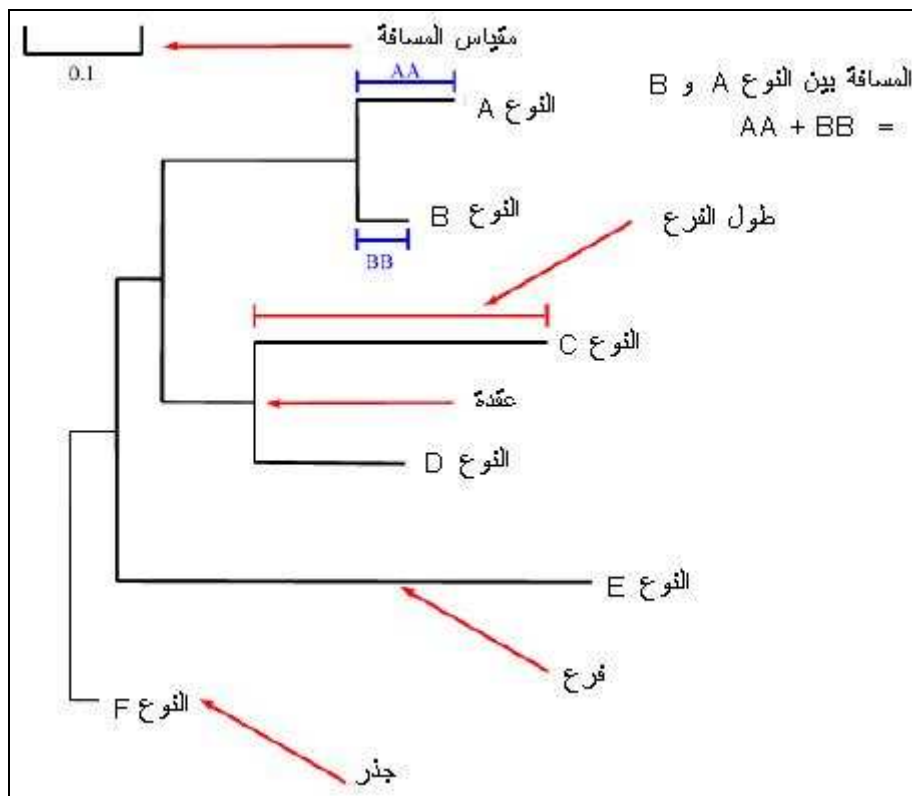
يشير الى حصول حالة استبدال وبعدها حالة حذف لحين الوصول الى الحالة الناتجة

agcta	agctga	acctga
حذف	استبدال	-

ففي حالة الاستبدال ، تم استبدال g مكان c وفي حالة الحذف تم حذف c .

فعليه فان مسافة التغيير Edit distance مساوية 2

كما ان التشابه العالي للتواليات يشير الى ان التغيير عن الأصل المشترك او الجد المشترك هو جديد الحدوث ، اما في حالة التشابه الواطئ فانه يشير الى انفراج وتباعد التواليات من زمن بعيد وهذا يعكس فرضية الساعة الجزيئية Molecular clock والتي تشير الى حدوث معدل ثابت من الطفرات وبالتالي التغيرات التطورية والتي تستخدم لاستنتاج الوقت المنقضي منذ الانفراج الأول للجينات. لذلك يمكن عد إيجاد العلاقات التطورية الجزيئية أحد مظاهر المعلوماتية الحيوية المستعملة في التنبؤ بمدى حالة التطور الحاصلة للأحياء الموجودة اليوم. والأشجار التطورية Phylogenetic trees هي رسم او صورة ترتب فيه الأنواع على الأفرع التي تربط اعتمادا على العلاقة وانحدارهم التطوري في أبسط أشكالها وتوضيح بعض المعلومات عن الشجرة الممثلة في الشكل 26.5 :

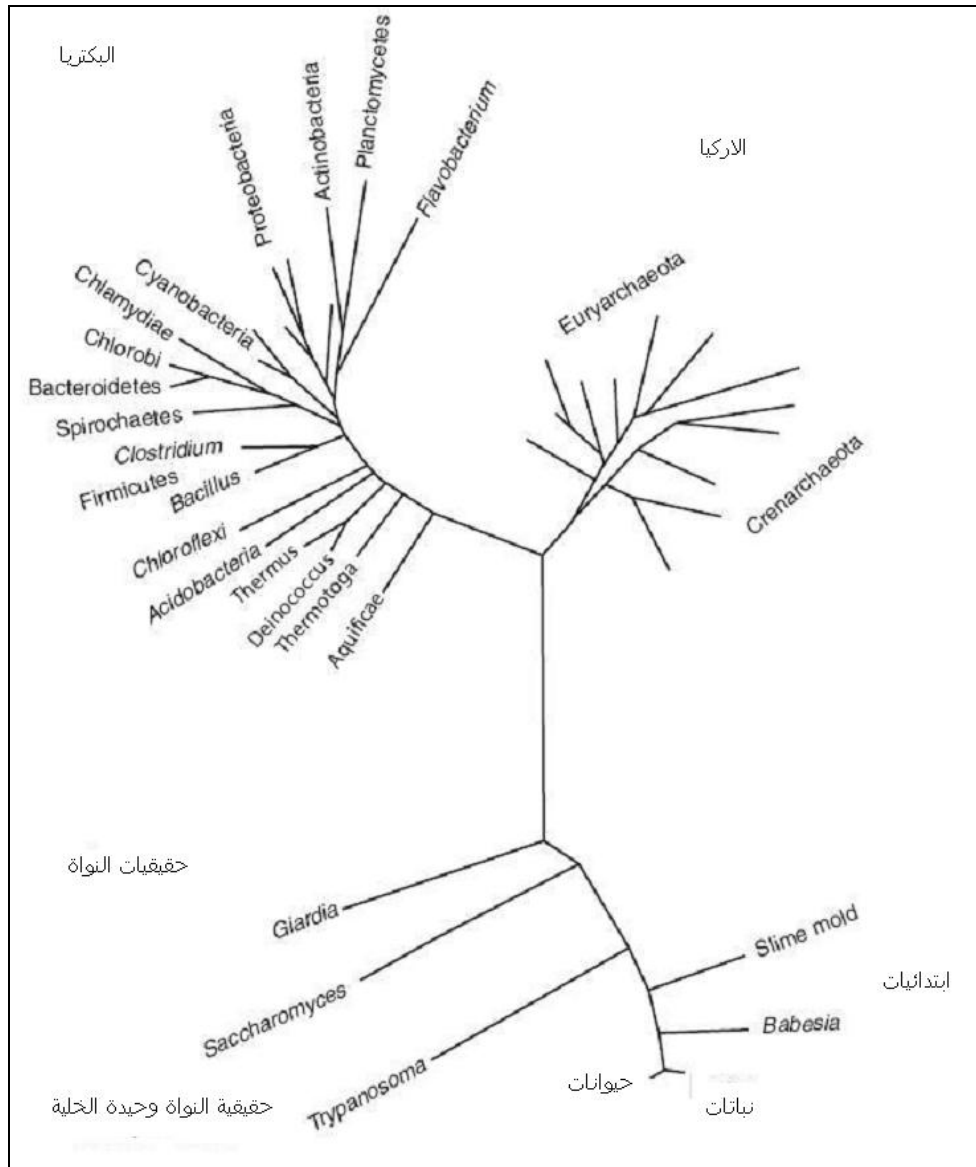


شكل 26.5 : توزيع الأنواع على الشجرة التطورية

وتستعمل لدراسة العلاقات التطورية ولاستنتاج أو تحديد العلاقات وتحديد تأريخ التطور. وبعض الأحيان تسمى العلاقة التطورية بـ Cladistics.

التطور وسرعة نمو الأحياء المختلفة

نظراً لكون عملية التطور تمثل التغير في الجينومات وما يتبعها من جزيئات حيوية فهي تكون رهناً بعدد من الظروف مثل معدل وقت الجيل للكائن وكذلك معدل عمليات الايض وتأثير الظروف البيئية. وبما ان كل من البكتريا والاراكيا سريعة النمو والانقسام فكانت هناك فرصة كبيرة لعمليات التطور فيها، وهذا ينعكس بالتنوع الكبير لهذه الأحياء والذي يفوق تلك الخاصة بالأحياء حقيقية النواة، وكان هذا مدعماً من دراسة التنوع الحيوي الذي لا يحكم عليه من دراسة مظاهر الكائنات وانما من الفحص الدقيق لموادها الوراثية والشكل التالي (27.5) يوضح ان مجموعة البكتريا هي المتغلبة من ناحية عدد الأجناس مقارنة بالحيوانات والنباتات.



شكل 27.5 شجرة التطور للأحياء وتغلب البكتريا على باقي الأحياء

وفضلا عن هذا تشير الدراسات التطورية الى ان البكتريا والاركيا موجودة قبل 3 بليون سنة في حين تمثل الحيوانات والنباتات القادمين الجدد الى الأرض منذ حوالي نصف بليون سنة. وإضافة الى سرعة النمو لكل من البكتريا والاركيا فان هذه الأحياء تستضيف عناصر وراثية خارجية تؤثر في أنماطها المظهرية، وذلك لان حدوث الطفرات التلقائية يحتاج وقت طويل لتثبت في الأحفاد، في حين ان نقل المعلومات أفقيا بين الأحياء يحتاج الى ثواني وهذا الشائع في هذه المجاميع، ولكن من جهة ثانية فان إمكانية الحفاظ على DNA في الخلايا المستقبلية واستقراره يحتاج الى وقت طويل للكشف عنه، وفي هذه الحالة يتم الاحتفاظ فقط بالمواد الوراثية ذات الفائدة للمضيف قبل ان تظهر على المجتمع الصفة المنقولة. وما يجعل من

استقرار المواد الوراثية القادمة هو الضغط الانتخابي الذي تتعرض له الخلايا والذي يجعل من انتشار الصفة وبالتالي تؤدي الى حرف المجتمع (وهي الوحدة الممكن التعامل معها في حالة البكتريا والاركية لصغر حجمها) . فمثلا عند تعرض المجتمع البكتري للمضادات الحيوية فان هذا يعمل بمثابة حاصد شرس يدمر المجتمع بكامله، ولكن هناك بعض الخلايا المثابرة تبقى حية وأحفاد هذه الخلايا هي التي تستطيع التكاثر وإنتاج مجتمع جديد مقاوم. والخلايا المقاومة التي أنتجت المجتمع الجديد يمكن ان تكون في إحدى الحالات الآتية:

- خلايا مقاومة موجودة في المجتمع ولكنها قليلة لا يمكن الكشف عنها الا بتسليط ضغط انتخابي.
- خلايا حصلت فيها طفرات بالصدفة.
- خلايا اكتسبت المقاومة من أحياء مقاومة موجودة معها في البيئة نفسها.

ومثل هذا التطور يكون ضئيلا" والبعض يعده تطبعاً وليس تطورا" ولكن في الحقيقة لا توجد اختلافات كبيرة او أساسية بين الاثنين وذلك لان العديد من خطوات التطبع تتجمع في عمليات التطور الكبيرة.

وبالرغم من ان التطور في الأحياء حقيقة النواة يتبع الأساسيات نفسها ولكن يكون من الصعب ملاحظتها كما هو الحال مع البكتريا ، لان وقت الجيل في حقيقيات النواة يكون طويلا وتتكاثر بشكل جنسي أساسا، وبذا فالأنماط المظهرية يمكن ان تغطي على عمليات التطور خاصة على المستوى الجزيئي. والبكتريا وان كانت نموذجا لملاحظة عمليات التطور الا ان الفيروسات تكون أسرع منها.

ومما ذكر أعلاه وما فهم من موضوع التطور الذي يعني التغيير المستمر فالجينومات البكتيرية قادرة على التغيير مع الوقت وبصورة أسرع من الأحياء الأخرى وبذلك تكون جينوماتها عرضة للكبر حيث تكبر بإضافة العاثيات وتواليات الإقحام IS والبلازميدات وغيرها من العناصر الوراثية المتحركة، ولكن في الوقت نفسه فان التطور يمكن ان يؤدي الى انكماش الجينوم كما يحصل في اختزال الجينومات للأحياء المتعايشة والمتطفلة، في حين توجد أحياء بجينومات ثابتة ولا تتغير بشكل كبير ولكن تحوي على تنوع وراثي عالي، ومنها يستخلص ان التنوع والتطور هي عمليات مستمرة تترك علاماتها او بصماتها على الجينومات البكتيرية في كل الأوقات .

التطور الجزيئي

يعني التطور في مجال علوم الحياة هو تطور الأشكال البايولوجية من أشكال سابقة لها اي من أصولها ربما المشتركة نتيجة لعمليات التحوير والانتخاب الطبيعي للأفضل منها. وتعرف العلاقات التطورية على انها دراسة تاريخ تطور الكائنات الحية باستعمال أشجار التطور التصويرية لتمثل أصول الكائنات. ومثل هذه الدراسات غالبا ما تتم باستعمال سجلات الحيوانات المتحجرة الحاوية على معلومات تخص الأشكال المظهرية لأصول الكائنات، ولكن هذه السجلات محدودة الفائدة لعدد من الأسباب منها انها غير متاحة لكل الكائنات وتأثرها بالمناطق الجغرافية وغيرها من الأسباب مما يجعل الدراسات في هذا التوجه غير مجدية فمثلا معظم مجموعة الأحياء المجهرية ليس لها سجلات حفريّة او سجلات لمتحجراتها مما يؤدي الى استحالة دراستها في هذه الحالة. وعليه كان الاعتماد على المعلومات الجزيئية التي فرضت

نفسها بكونها أكثر ملائمة من صفات الكائن الأخرى لإيجاد تحليل الغصون Clades وتحليل الأشجار التطورية مثل DNA اي إيجاد شجرة الجينات Gene tree والبروتينات التي توفر منظور تطوري مفيد للكائنات الحية الموجودة اليوم. والمواد الوراثية تتراكم فيها الطفرات والتغيرات عبر الزمن مما يجعلها ملائمة ان تكون متحجرات جزيئية Molecular fossils يمكن ان تحلل وتدرس للأحياء ذات العلاقة وبذا يمكن التوصل الى التأريخ التطوري للجينات ومن ورائه الكائنات .

يكون استخدام البيانات الجزيئية أكثر وضوحا من سجلات المتحجرات للأحياء وأسهل في الحصول عليها ولا يوجد انحياز في النماذج المستعملة ومتاحة للباحثين مما وفر الفرصة للباحثين لإعادة كتابة وتركيب التأريخ التطوري. وبذا نشأ علم التطور الجزيئي Molecular phylogenetics الذي يمكن تعريفه على انه دراسة العلاقات التطورية للجينات والجزيئات الحيوية العملاقة بتحليل الطفرات الحاصلة في مواقع مختلفة في توالياتها ومن ثم إيجاد الفرضيات حول العلاقات التطورية بينها بالاعتماد على مدى التشابه لهذه الجزيئات، وبالتالي يمكن استنتاج العلاقة التطورية بين الأحياء. ومن المعروف ان القوة الدافعة لعمليات التطور هي الانتخاب الطبيعي والتي تفترض حقيقة ان الأشكال والأحياء غير الملائمة يجب ان تتلاشى او تستأصل نتيجة لتغير الظروف البيئية وبالتالي يبقى الأفضل، ويكون الأساس وراء هذه هو حدوث الطفرات التلقائية التي تؤدي الى التنوع الحيوي في المجتمعات. وبذا يكون التنوع الوراثي المصدر الرئيس للمواد الوراثية التي يعمل عليها الانتخاب الطبيعي. لذلك يطرح السؤال الآتي من أين يأتي التنوع والاختلاف الذي يكون خاضعا للانتخاب الطبيعي؟ ويكون الجواب بداية ان متطلبات التطور هو وجود أحياء مختلفة التي يتم انتخاب المجتمعات الثانوية منها ، ويكون التنوع الوراثي في مجتمع متجانس نتيجة :

- وجود أخطاء أثناء تضاعف DNA

- وجود الفيروسات

- وجود تواليات الإقحام (IS)

- وجود الجينات القافزة Transposons

والثالث الأخيرة تقحم او تحذف من الكروموسومات تحت ضوابط معينة والبعض الآخر بدون ضوابط . فضلا عن وجود آليات أخرى، وهذا يعني وجود عمليات تحول بمستوى معين ولكن بطيئة (في بعض الأحياء) على مدى زمني طويل تؤدي الى التنوع الوراثي، وهذا يعني ان عامل الزمن اللازم لتراكم خطوات التطور التي يكون تأثير كل خطوة فيها بسيط جدا ولكن مجموع التأثير سيؤدي الى اختلاف واضح.

فرضيات المنهج الدراسي للعلاقات التطورية الجزيئية

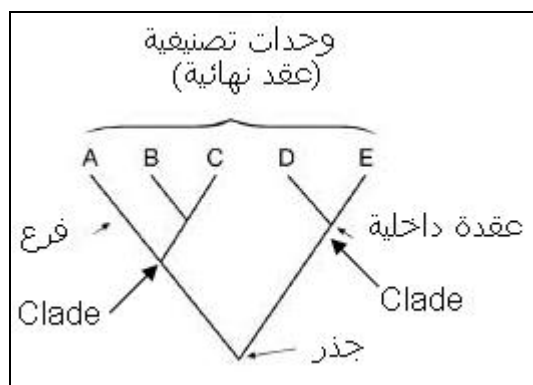
لأجل استعمال المعلومات الجزيئية لغرض إعادة النظر في تركيب تأريخ التطور لابد من وجود بعض الفرضيات لجعل هذا التوجه يسير في نطاق منهجي ومن هذه الفرضيات أولا ان التواليات الجزيئية المستخدمة في إيجاد العلاقات التطورية لابد ان تكون متماثلة جذريا في أصولها ثم انفجرت بمرور الزمن ، الافتراض الثاني ان عملية الانفراج كانت ثنائية التشعب . الفرضية الثالثة ان كل موقع في التواليات يتطور بشكل مستقل عن الآخر. وأشجار العلاقات

التطورية التي تبنى على أساس اصطفايف التواليات يمكن ان تفشل فيما اذا كانت البيانات كبيرة ومعقدة.

مصطلحات ومفاهيم في العلاقات التطورية الجزئية

ان المسافة بين تواليتين يمكن ان تحسب لتعطي نموذجا للتطور وهذه يمكن ان تجري على عدة مستويات :

- الواطئ ويكون على مستوى جين منفرد ويكون البحث هنا عن تماثل تواليات الجين.
 - العالي ويكون على مستوى الجينوم الكلي وفيها تستعمل كل الجينات وتتدخل في الحسابات عدد مرات الإدلاف اللازمة التي أدت الى وجود الجينات الحالية.
- وتحتاج مثل هذه الدراسات الى مفردات خاصة بها ، ومنها ما يستعمل في رسم أشجار التطور، ويوضح الشكل التالي (28.5) أحد الأشجار التطورية البسيطة ثنائية التفرع:



شكل 28.5 : شجرة تطور ثنائية التفرع

فالخطوط تدعى بالأفرع والتي تمثل نهاياتها مثل A, B وغيرها الأجناس او الأنواع من الأحياء الموجودة اليوم اي تمثل الوحدات التصنيفية Taxa (مفردتها Taxon) وتمثل اي مجموعة مسماة من الأحياء وليس بالضرورة ان تكون غصن . اما نقطة التقاء الفرعين فهي العقدة التي تمثل أول الأصول المشتركة للأجناس والأنواع المشتركة فيها. اما التقاء الفروع الرئيسية فانه يمثل عقدة الجذر التي تمثل الأصل العام المشترك للمجاميع الموجودة ضمن الشجرة التطورية. وفي حالة الوحدات التصنيفية المذكورة في الشكل مثل E و D او B و C والتي تشترك في أصل او جد مشترك فيطلق عليها الغصن وهي مجموعة تطورية متحدة غير متفرعة Monophyletic group والتي قد تكون مجموعة من الأحياء او الجينات التي تشترك في أصل او جد مشترك و Clade تعني باللغة اللاتينية الغصن Twig او الفرع Branch وهنا فان هذه المجموعة لا تشترك مع مجموعة تصنيفية أخرى وتسمى بالمجاميع التصنيفية الشقيقة . اما المسار الذي يربط الوحدة التصنيفية الى العقدة الجذرية او الخط التطوري المشترك يعرف بالخط Lineage والذي يكون في الغالب مرادفا لفرع الشجرة الذي يقود الى المجموعة التطورية المتحدة.

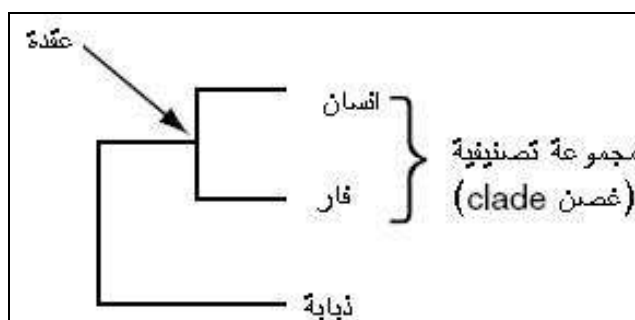
اما في حالة اشتراك الوحدات التصنيفية بأكثر من أصل مشترك (قبل الوصول الى قاعدة الجذر) كما في حالة الوحدات (C,B) او (E,D) فان هذه المجموعة يطلق عليها مجموعة التطور النظيرة Paraphyletic.

دراسة الغصون Cladistics

مصطلح يطلق بعض الأحيان على العلاقات التطورية لان كلمة Clade (المذكورة أعلاه) تعني في المفهوم العام غصن وفي مفهوم التطور هي تمثل مجموعة من الأحياء المنحدرة من جد واحد. وبعض الأحيان يمكن ان تستعمل اسماً لطريقة معينة تستعمل افتراض معين للعلاقات التطورية. والخيمة التي تغطي موضوع دراسة الغصون هي مجموعة من الأحياء او فروع تشترك في تأريخ تطوري مشترك وتكون ذات علاقة فيما بينها أكثر من علاقتها مع مجاميع أخرى اي انها مجموعة تمتاز بصفة مشتركة معينة مميزة لها لا توجد في المجموعات الأخرى . وهذه الصفة المشتركة يمكن ان تكون اي شيء يمكن ملاحظته ووصفه.

والتحليل يشمل مقارنة مفردات متعددة في وقت واحد مثل مقارنة الأنماط المظهرية او القواعد النتروجينية او تواليات حوامض امينية. ولكن هناك عدة افتراضات لهذا التحليل منها:

- ان المجموعة من الأحياء تكون ذات علاقة بأصل مشترك.
 - يكون هناك تفرع ثنائي لتكوين الأغصان Cladogenesis وان كان هذا الافتراض متعثراً" بعض الأحيان.
 - الافتراض الآخر هو ان التغيير في مفردات الخط التطوري Lineage تحدث بمرور الوقت وتكون الأساس في تحليل وإيجاد Cladistics.
- وتمثل العلاقة في التحليل بالشكل المبسط أدناه (29.5):



شكل 29.5 : الخط التطوري Lineage

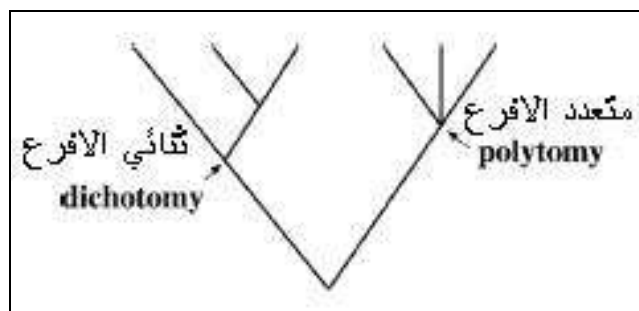
ويتضح من الشكل ان الفيران أقرب الى حشرة الذباب أكثر من قرب الإنسان الى هذه الحشرة. ومن المصطلحات الأخرى المهمة في موضوع الأشجار التطورية هو ما يخص اي من الجزيئات العملاقة التي تتطور، فهناك تطور جين وهناك تطور بروتين. ولذلك كان هناك اختلاف بين تطور الجين وتطور الكائن الحي.

ولدراسة تطور الجين تستغل مقارنة التماثل الجذري للتواليات التي لها أصول مشتركة ولكن قد لا تكون لها فعالية مشتركة، وتعد التواليات متماثلة جذريا اذا اشتركت بدرجة تشابه أكثر من

الحد الفاصل مستورثة من الجد المشترك وتكون لها تراكيب مشتركة ولو ان التركيب في الجد قد يكون من الصعوبة تحديده لانه تعرض للتحويلات .
وحالة التماثل الجذري قد تكون قائمة التناظر Orthologs او متناظرة التشابه Paralogs او دخيلة التشابه Xenologs . وحالة التناظر تمثل حالة تماثل جذري نتج عند نشوء الأنواع او تحديد الأنواع Speciation وتمثل الجينات المشتقة من أصل او جد مشترك الذي انفرج او انفصل عن الأحياء التي كانت له علاقة بها، ومثل هذه المتشابهات تكون لها وظائف مشتركة. حالة التناظر المشابهة لحالة تماثل جذري نتجت من تضاعف الجين وتمثل الجينات المشتقة من أصل جين مشترك تضاعف في الكائن نفسه ثم بعد ذلك انفرجت وهذه تميل ان تكون لها وظائف مختلفة.

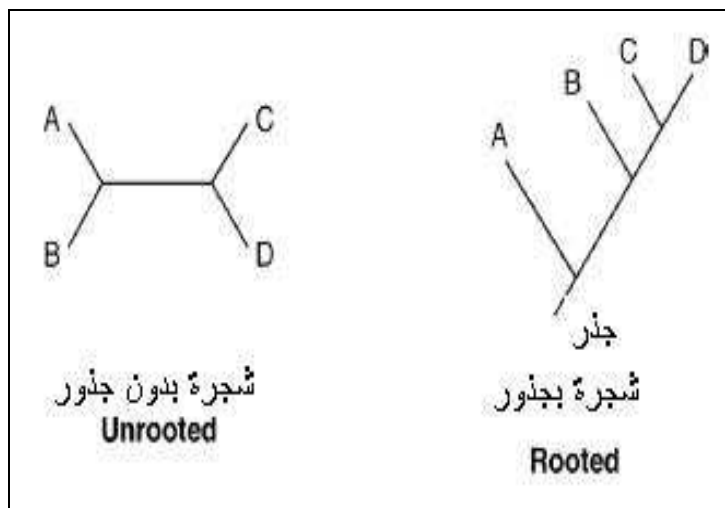
اما حالة التناظر الدخيل فهي ناتجة عن الانتقال الأفقي للجينات بين اثنين من الأحياء. وتحديد الجين الذي انتقل حديثا الى المضيف يكون صعبا ولكن المحتوى (%G+C) تكون مختلفة بشكل كبير جدا عن الجينات العادية للمضيف مما يشير الى الأصل الخارجي للجين ولكن تبقى مسألة طريقة انتقاله أفقيا مشوبة بالشك. والجينات الغريبة في الغالب تكون لها وظائف مشتركة الا في بعض الحالات حيث تحصل تغيرات يحدثها الجين في البيئة الكروموسومية التي ينغرس فيها مما يؤثر في الوظيفة التي يؤديها.

ومن جهة ثانية فان الأشجار التطورية يمكن ان تتخذ أشكالا مختلفة والذي يطلق عليه طبوغرافية الشجرة اي نمط توزيع الأفرع فيها. فهناك أشجار ثنائية التفرع Dichotomy اي ان كل نقطة تتفرع الى فرعين متناظرة في الغالب ، اما الأنواع الأخرى فيمكن ان يتفرع من النقطة الواحدة أكثر من فرع اي ان العقدة تكون متعددة الأفرع ويطلق عليها الشجرة متعددة الأفرع Polytoimy كما موضح في الشكل 30.5 :



شكل 30.5 : أشجار ثنائية متعددة الأفرع

ويوجد نمط آخر لأشجار التطور فالبعض منها تشترك في نقطة الأصل وتسمى الشجرة المجذرة Rooted tree، والنوع الآخر بدون جذر Unrooted tree اي تكون غير قطبية كما موضح في الشكل 31.5 :



شكل 31.5 : الأشجار المجذرة وغير المجذرة

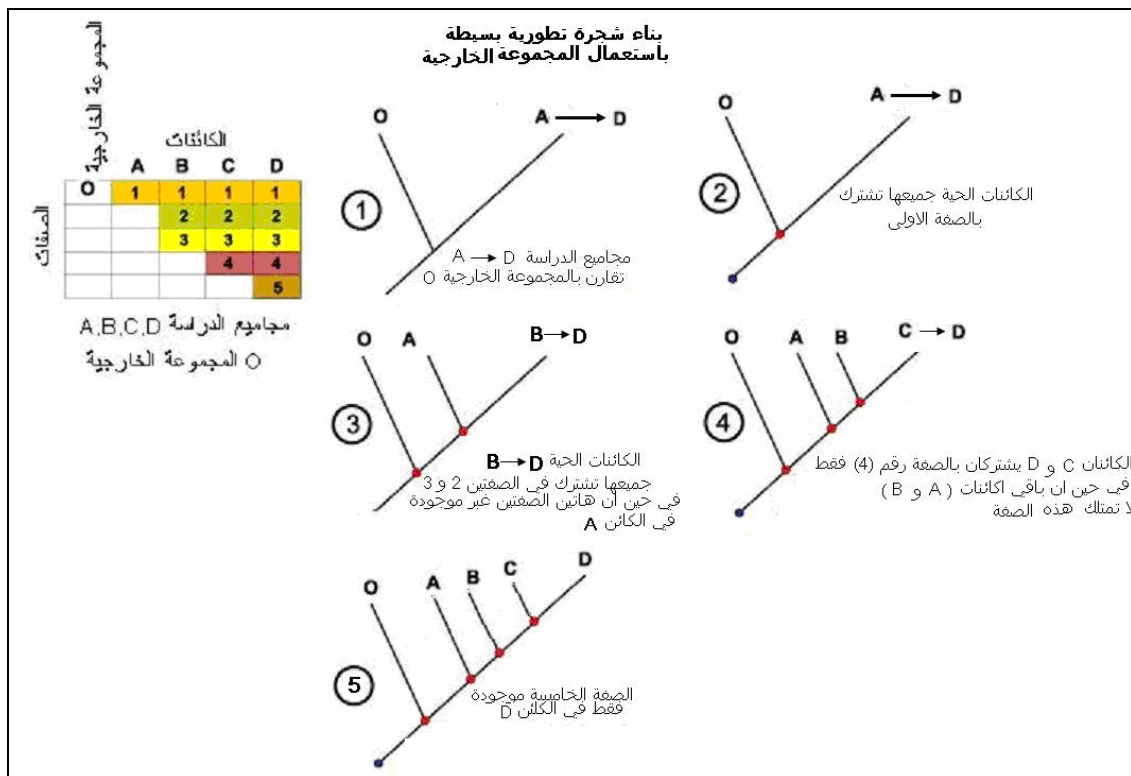
وبطبيعة الحال فإن الشجرة غير المجذرة أو بدون جذر لا تعطي معلومات عن الأصل المشترك وانما تستعمل لتوضيح العلاقة بين الوحدات التصنيفية مع بعضها، ولذلك فهي لا تفيد في دراسة العلاقات التطورية بين الوحدات التصنيفية . في حين توفر الأشجار المجذرة توضيحاً لاتجاه مسار التطور، وتشير إلى أن التواليات تحت الدراسة يكون لها أصل مشترك أعلى الذي تنطلق منه مسارات توضيح التطور الحاصل والذي يقود من عقدة إلى أخرى ، ولذلك فهي توفر معلومات أكثر من الأشجار الخالية من الجذور، ولغرض تحويل الأشجار الخالية من الجذور إلى أشجار حاوية على الجذور لابد أولاً من معرفة أين تقع نقطة الجذر.

وبالنسبة للأشجار ذات الجذور فإن جذرها الأصلي في الوقت الحاضر هو منقرض وهناك طريقتين لمعرفة ذلك ، الأولى استعمال طريقة المجموعة الخارجية Outgroup وتمثل السيطرة التجريبية والتي هي توالي يكون له تماثل أو تشابه جذري للتوالي تحت الدراسة ولكنها انفصلت عنها في أوقات مبكرة من عملية التطور ، وينتخب التوالي الخارجي من مجموعة تصنيفية بعيدة إلى حد ما عن الكائن الذي تواليه قيد الدراسة مثل استعمال تواليات من الطيور لغرض تحديد العلاقات التطورية لتواليات من اللبائن وبذا تشير إلى أن تواليات الطيور انفصلت عن اللبائن منذ زمن بعيد (كما ذكر في مستهل الموضوع) نظراً لقلّة التماثل الجذري. وعلى العموم فإن اختيار تواليات المجموعة الخارجية يجب أن لا يكون بعيداً جداً عن تواليات المجموعة الداخلية In group (وهي في هذه الحالة اللبائن) لأن ذلك سيؤدي إلى أخطاء في بناء شجرة التطور. أما في حالة استعمال توالي من مجموعة قريبة جداً فإن ذلك يلغي فكرة المجموعة الخارجية . فمثلاً استعمال البكتيريا كمجموعة خارجية يمكن أن يؤدي إلى اضطراب العملية نظراً لأن معظمها تحوي على جينات أو تواليات انتقلت إليها بالنقل الأفقي عن طريق الاقتران أو التثبيغ أو التحول المعروفة .

الطريقة الثانية تستعمل عند غياب توالي مجموعة خارجية ملائمة وتتم باستعمال طريقة التجذير الوسطي Midpoint rooting approach وفيها تحدد نقطة الوسط بين المجاميع المتباعدة بقياس طول الأفرع واعتماد نقطة الوسط كجذر للشجرة ويكون ذلك باعتماد فرضية

الساعة الجزئية المارة الذكر والتي تفترض ان التواليات تتطور بمعدل ثابت. بحيث ان تراكم الطفرات فيها يتناسب طرديا مع زمن التطور ومنه يمكن ان تمثل أطوال الأفرع كأساس للقياس انفراج التواليات عن بعضها ولو ان هذا التوجه لا يمثل الحقيقة الواقعة نظرا لان التغيرات وكذلك الطفرات لا يمكن ان تنتج من منظور التغيير الذاتي او التلقائي لان هناك عمليات أخرى يمكن ان تؤثر في نوعية ومعدل تغير التواليات.

والطريقة الأخرى التي لا تعتمد على المجموعة الخارجية لإيجاد الجذور تشمل تحليل الجينات المتضاعفة او متناظرة التشابه Paralogs فاذا كانت كلها متناظرة في معظم او كل الكائنات الداخلة في الدراسة فعندها يمكن وضع الجذر في المكان الصحيح. وعمليا ولبناء شجرة بأبسط صورها لمجاميع مختلفة مثل A و B و C و D ولغرض تحديد جذرها تقارن بمجموعة خارجية (O) كما موضح في الشكل 32.5 :



شكل 32.5 : بناء شجرة تطورية بسيطة باستعمال المجموعة الخارجية

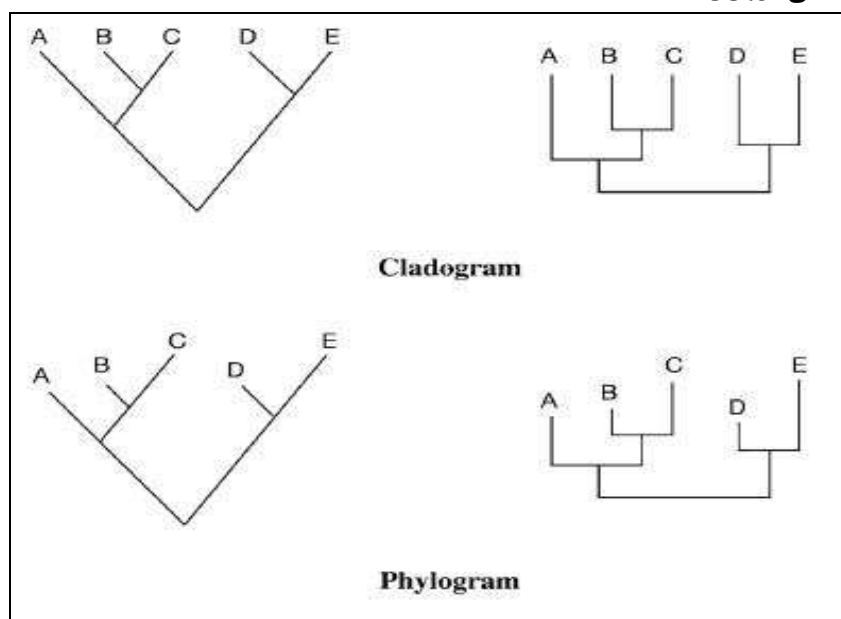
تقارن المجاميع الأربعة مع المجموعة الخارجية وتسجل نقاط التشابه كما موضح في جدول الشكل أعلاه وترسم كما في الخطوة (1). وتستعمل المجموعة الخارجية كنقطة مرجعية للشجرة لتمييز المجاميع القديمة من المجاميع الجديدة، فمثلا A تمثل مجموعة قديمة في الشجرة في حين D تمثل مجموعة جديدة في الشجرة . في الخطوة الثانية (2) ترسم عقدة لتمثل الجد الأصلي للمجموعة الخارجية ومجاميع الدراسة اي تكون أصلا او جدا مشتركا لكل المجاميع الواقعة

فوقها. الخطوة الثالثة (3) ترسم عقدة للكائن A ... وهكذا وصولا الى الخطوة الخامسة والأخيرة.

والمسافة لفرع المجموعة الخارجية يمثل الوقت النسبي للأصل والمسافة بين العقد تمثل الأزمان عندما تغير الجد الأصلي في حالة التطور اعتمادا على الصفات.

التمثيل الصوري لأشجار التطور

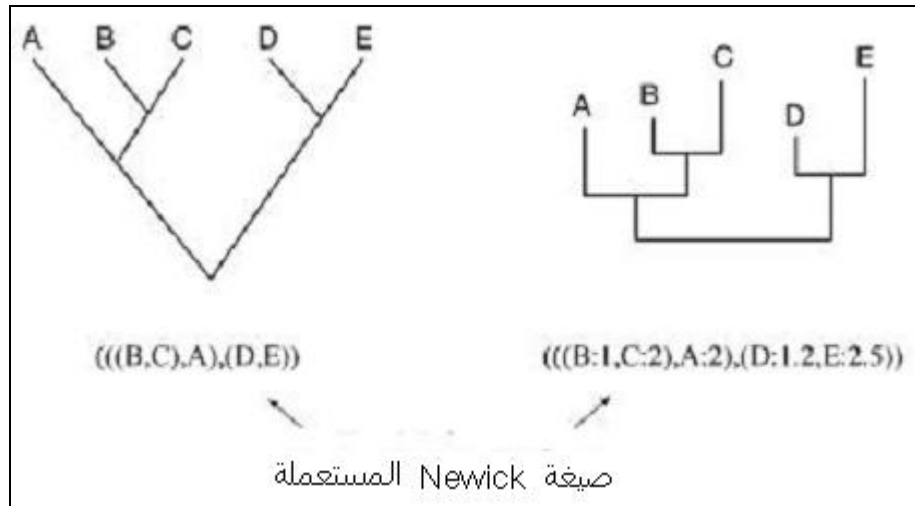
ان طبوغرافية مخططات أشجار التطور تحدد العلاقات بين الوحدات التصنيفية. ويمكن ان ترسم الأشجار بعدة طرق مثل الشكل المتساوي الأطراف والآخر غير متساوي الأطراف كما موضح في الشكل 33.5 :



شكل 33.5 : التمثيل الصوري لأشجار التطور

وفي مثل هذه الأشكال يمكن تدوير الصورة دون التأثير في العلاقات بين وحدات التصنيف. ويلاحظ من الجزء الثاني من الرسم (الشكل 33.5) الذي يمثل الشكل التطوري Phylogram ، ان أطوال الفروع التي تكون لها معاني خاصة فهي تشير الى مدى الانفراج في التطور ولمثل هذه الأشجار مقاييس خاصة بها، وذلك لان هذا النوع من الأشجار يوضح العلاقة التطورية وكذلك توفر المعلومات حول زمن الانفراج النسبي للأفرع. اما في الجزء الأعلى Cladogram فان طول الأفرع لا تمثل العلاقات التطورية وانما توضح فقط ترتيب وحدات التصنيف.

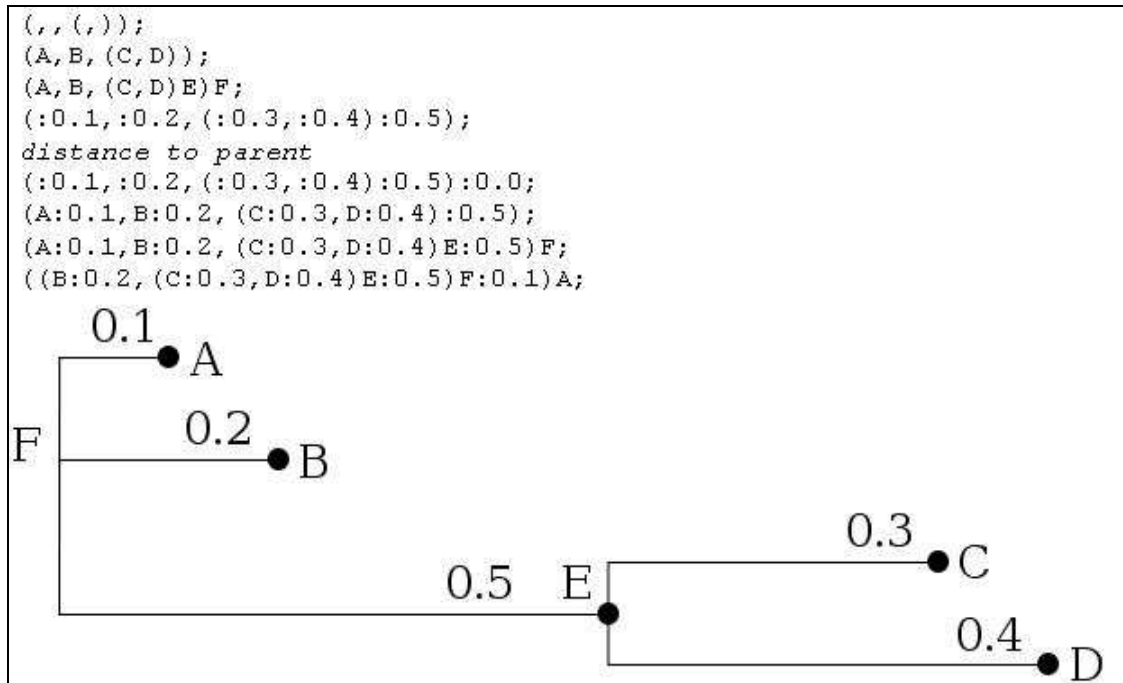
ولغرض رسم الأشجار بالحاسوب هناك عدة صيغ لكتابة المعلومات التي يستعملها الحاسوب لرسم الشجرة كما موضح في الشكل 34.5 :



شكل 34.5 : كتابة الصيغ عند رسم الأشجار التطورية

وتكون الصيغة بشكل أقواس متجمعة، حيث يمثل كل قوس العقدة التي تحويها من الوحدات التصنيفية التي يفصل بينها بفاصلة. ويمكن ان تدخل المعلومات مع طول الأذرع ويمكن التحكم بالشجرة الناتجة بصيغ Newick كما موضح في الشكل 35.5 :

5



شكل 35.5 : صيغ Newick لإدخال البيانات لبناء الأشجار التطورية

وملاحظة عامة فان أعداد الأشجار يزداد بشكل تزايدى مع ازدياد الوحدات التصنيفية ليصل الى أرقام فلكية. ولكن المحددات الحسابية (التي سيأتي ذكرها لاحقا) تسمح باستغلال عدد صغير من الأشجار الممكنة والرقم يعتمد في الأساس على عدد الوحدات التصنيفية فضلا عن اعتماده على عوامل أخرى منها:

- المؤشرات المستعملة في الطريقة وسرعتها.
- Hardware للحاسوب المستعمل.
- Software للحاسوب او البرنامج المستعمل، حيث ان بعض الخوارزميات سريعة مقارنة بالأخرى ، كما ان بعض برامج الحاسوب تقوم بعمليات متعددة في الوقت نفسه، وبعض الحواسيب تحدد عدد ونوع الأشجار التي يمكن ان تخزن في الذاكرة.
- نوعية الخوارزميات التي تختلف في الأشجار التي تبنيها، فبعضها يكون شاملا ويمكن ان يعطي كل أنواع الأشجار الممكنة وأخرى تتعامل مع القليل من المؤشرات المستعملة وبطبيعة الحال فان عمليات البحث تتأثر بالطبيعة التركيبية للبيانات وحجمها.

العلاقة التطورية للجينات مقابل العلاقة التطورية للأنواع

ان العلاقة بين تطور الجينات وتطور الأنواع لا تكون متوازنة. وذلك لان تطور الأنواع وما يظهر عليها من أنماط مظهرية يكون حصيلة لتطورات وتغييرات عدد من الجينات وليس جينا واحدا. وعليه فان دراسة تطور الجين الذي يمكن ان يستنتج من دراسة تواليه او توالي البروتينات التي ينتجها (في حالة الجينات المشفرة) تعكس فقط تطور الجين نفسه او البروتين الذي يشفر له.

ولذلك كانت هناك عدة نقاط لعدم الالتقاء بين تطور الجين والنوع الذي يوجد فيه، ومن هذه النقاط ان الصفات المظهرية التي تظهر على النوع قد يكون بعضها ناتجا عن إدخال جينات غريبة الى النوع كما في حالة النقل الأفقي للجينات. وكذلك التحكم بفعالية الجين وتنظيمه فقد يبقى الجين ساكنا تحت ظروف معينة وربما يتحول الى جين كاذب فضلا عن داييميكية الحالة الوراثية التي تغير من الجينات بواسطة إدلاف الجينات وغيرها من عمليات ترتيب المواد الوراثية. ومن هنا فرض مفهوم ان تطور التواليات لا يعني حدوث تطور في مسار تطور الكائنات بشكل مترام.

ومن هنا كانت نقطة التفرع الداخلي في شجرة الأنواع تمثل إنتاج نوع معين مما يشير الى حدوث تغييرات عديدة في شجرة الجينات في حين تمثل العقدة الداخلية لشجرة الجين حدوث بعض الأحداث الخاصة بالجين مثل تضاعف الجين. وبذلك فان الحدثين لا يكونان متوافقين . ومن هنا يستنتج انه لبناء شجرة تطور النوع لابد من استعمال أشجار تطورية لعوائل جينات مختلفة لتعطي النمط النهائي لتطور النوع.

فهم العلاقات التطورية باستعمال الدراسات الجزيئية

وفر التطور في الكيمياء الحيوية وسائل جديدة لبناء الأشجار التطورية، فتحديد تواليات DNA و RNA والبروتينات هي من التقنيات الجزيئية المستعملة في تشخيص وتوصيف الكائنات وكذلك فهم العلاقات التطورية بينها. وهناك مفهوم الساعة الجزيئية التي تشير الى ان معدل الطفرات في تواليات الحوامض النووية والبروتينات يحدث بمعدل ثابت وتكون تقريبا متناسبة

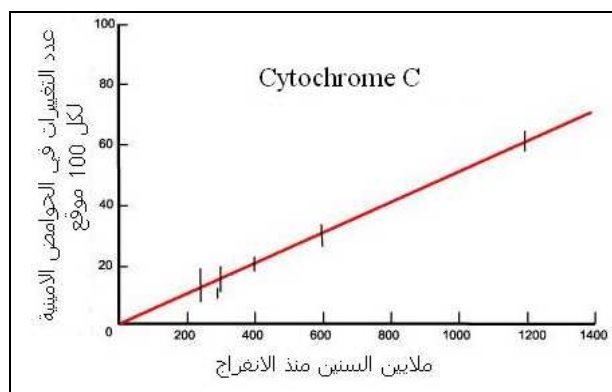
مع زمن التطور وهذه الفرضية تستخدم في عدد من خوارزميات وبرامج إعداد الأشجار التطورية ولكنها ليست صحيحة بشكل مطلق.

وقد اعتمد في مجال تحديد الزمن المنقضي وتأثيره في حدوث التغيرات في التواليات الجزيئية الحيوية على بعض البروتينات الثابتة أي أن معدل حدوث الطفرات فيها يكون منخفضاً جداً وعلى سبيل المثال استعمال بروتين سايتوكروم Cytochrome C الذي تم تحليل أكثر من 100 بروتين من أحياء حقيقية النواة ، ودراسة البروتين يمكن أن تدعم فرضية الساعة الجزيئية، والمخطط التالي (شكل 36.5) يوضح توالي البروتين في الإنسان والجرذ والخميرة:

Human Cyt C:	MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIW	61
Alignment:	MGDVEKGKKIP+ KC+QCHTVEKGGKHKTGPNLHGLFGRKTGQA G+SYT ANKNKGI W	
Rat Cyt C:	MGDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAAGFSYTDANKNKGITW	61
Human Cyt C:	GEDTLMEYLENPKKYIPGTMIFVGIIKKKEERADLIAYLKATNE	105
Alignment:	GEDTLMEYLENPKKYIPGTMIF GIKKK ERADLIAYLKATNE	
Rat Cyt C:	GEDTLMEYLENPKKYIPGTMIFAGIIKKKGERADLIAYLKATNE	105
Below: Alignment of human and yeast cytochrome c amino acid sequences using BLAST. Sequence similarity between the two proteins is 64%.		
Human Cyt C:	GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIW	61
Alignment:	G +KG +F +C QCHTVEKGG HK GPNLHG+FGR +GQA GYSYT AN K ++W	
Yeast Cyt C:	GSAKKGATLFPKTRCLQCHTVEKGGPHKVGPNLHGIFGRHSGQAEGYSYTDANIKKNVLWD	66
Human Cyt C:	EDTLMEYLENPKKYIPGTMIFVGIIKKKEERADLIAYLKKA	102
Alignment:	E+ + EYL NPKKYIPGTM F G+KK+++R DLI YLKKA	
Yeast Cyt C:	ENNMSEYLTNPKKYIPGTMAPGLKKEKDRNDLITYLKKA	107

شكل 36.5 : تغيرات Cytochrome في الإنسان والفار والخميرة

ووجد أن التشابه بين الإنسان والجرذ 91% في حين أن التشابه بين الإنسان والخميرة 64% وهذا يشير إلى أن المسار التطوري بين الإنسان والخميرة قد حدث منذ زمن بعيد وقبل الانفراج بين الإنسان والجرذ. ولذلك استعمل البروتين في حساب وقت الانفراج كما موضح في الشكل 37.5 :



شكل 37.5 : تطور بروتين Cytochrome C

- وتحديد زمن الأحداث باستعمال الساعة الجزيئية أثار بعض الشكوك ولعدة أسباب منها:
- الاختلاف في معدلات التطور يمكن ان يخفي معدل وموقع الانفراج وخاصة في حالة البروتينات.
 - الاختلاف في وقت الجيل للكائن وكذلك الاختلاف في معدلات عمليات الايض يمكن ان يؤثر في معدل حدوث الطفرات.
 - البيئة التي يعيش فيها الكائن والظروف التي يتعرض لها يمكن ان تولد لديه أنظمة أصلاح تؤثر في معدل الطفرات وكذلك إعادة الطفرات الحاصلة الى أصلها مما يؤثر في المشاهدات المسجلة والمعتمدة في بناء أشجار التطور.
- ومثل هذه العوامل يمكن ان تجعل الساعة الجزيئية اقل أهمية في عمليات التنبؤ. وكلمة يمكن ان تقال قبل الخوض في أنواع الأشجار وطرق تصميمها وبناءها هي انه ليس من المعقول ان كل تحاليل العلاقات التطورية يمكن ان تؤدي كل المهام المنوطة بها بشكل دقيق. وبالرغم من ان هناك عددا من الخوارزميات والبرامج قد تم إيجادها وتطويرها فان مدى ملائمتها ومصداقيتها تعتمد على تركيب البيانات وحجمها . وان جدارة اي طريقة او هفواتها تكون عادة عرضة للنقد اللاذع التي تناقش في المجالات والمراجع المهمة بالدراسات التطورية. ويكون الأساس في القبول المستخدم الذي يجب ان يكون على دراية واسعة بالطرق كي يختار ما يلاءم بياناته.

الجوانب الرياضية لبناء أشجار التطور

ان الهدف الرئيس من دراسة علاقات التطور الجزيئي هو لتصحيح إعادة بناء تاريخ التطور اعتمادا على تشعب وانفراج التواليات بين الكائنات والذي يعني إيجاد مخطط الأشجار الصحيح مع أطوال أفرعها الصحيحة وهذا قد يكون في غاية الصعوبة وذلك لان الأعداد المحتملة من المخططات يكون بأعداد كبيرة حتى مع أعداد الوحدات التصنيفية المعتدلة. ويحدد عدد الأشجار ذات الجذور N_R لوحدة تصنيفية عددها n بالمعادلة الآتية:

$$N_R = (2n - 3)! / 2^{n-2} (n - 2)!$$

وفي هذه الصيغة فان $(2n-3)!$ تمثل مفكوك التعبير الرياضي، والمفكوك هو حاصل الضرب الموجب للأعداد من 1 الى $(2n-3)$ ، فعلى سبيل المثال مفكوك 5! هو

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

وفي حالة الأشجار بدون جذور فيكون عدد مخططات N_u مستنتجا من الصيغة الآتية:

$$N_u = (2n - 5)! / 2^{n-3} (n - 3)!$$

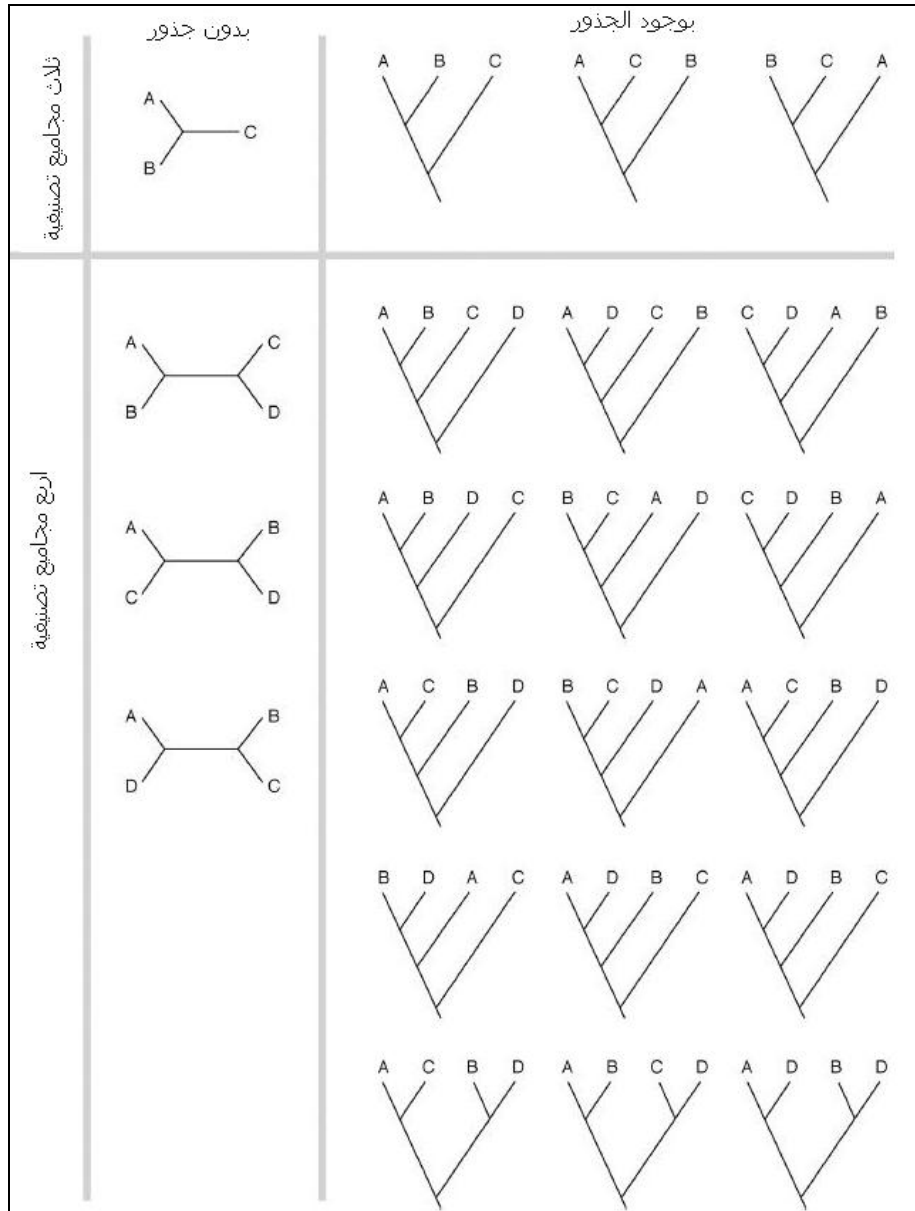
وعليه فان عدد المخططات لمجموعة من الوحدات التصنيفية $(n=3)$ هناك شجرة واحدة ممكنة من الأشجار بدون جذور، وعند تطبيق المعادلة الخاصة بالأشجار ذات الجذور فان عدد الأشجار (N_R) يكون 3 كما في الحسابات الآتية:

$$N_R = (2 \times 3 - 3)! / 2^{3-2} (3 - 2)! \\ = 3$$

في حين عند تطبيق المعادلة الخاصة بالأشجار بدون جذور تكون النتيجة 1

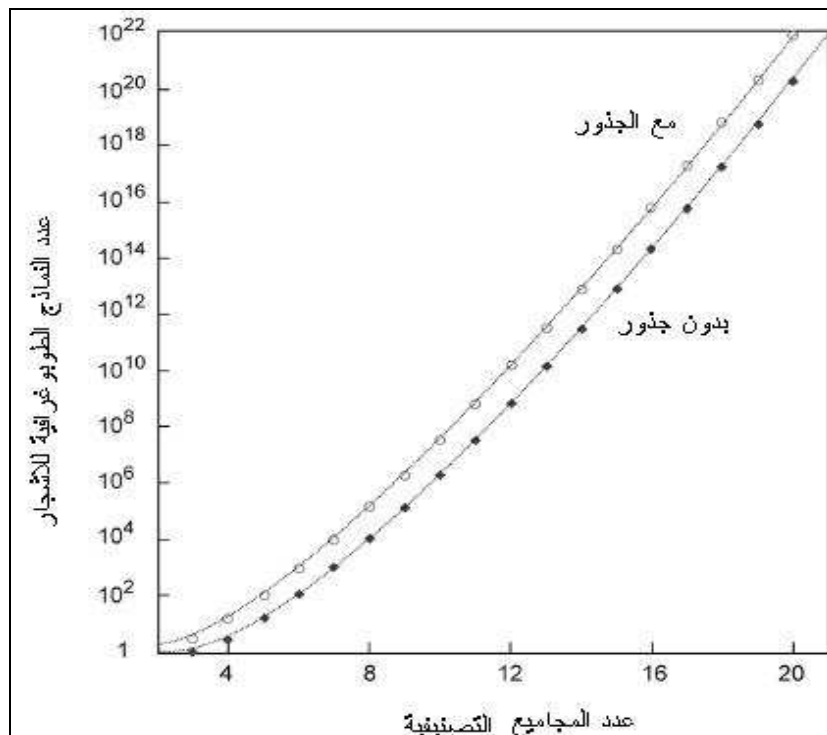
$$N_u = (2 \times 3 - 5)! / 2^{3-3} (3 - 3)! = 1$$

اما في حالة زيادة عدد الوحدات التصنيفية الى أربعة ($n=4$) فيكون عدد الأشجار بدون جذور 3 وفي حالة الأشجار المجذرة يكون 15 من تطبيق المعادلات أعلاه كما موضح في الشكل 38.5 :



شكل 38.5 : تمثيل لأعداد الأشجار التطورية لمجاميع مصنفة مختلفة في العدد بوجود او غياب الجذور

وعليه عندما تكون هناك ستة وحدات تصنيفية يصل العدد الى 105 بالنسبة للأشجار بدون جذور و 945 من الأشجار المجذرة وهكذا اي هناك زيادة أسية في عدد الأشجار التي يمكن الحصول عليها عند زيادة أعداد الوحدات التصنيفية كما موضح في الشكل 39.5 :



شكل 39.5 : الزيادة الأسية لعدد الأشجار بازدياد عدد المجموعات التصنيفية

بناء أشجار التطور

تبنى أشجار التطور الجزيئية بعدة خطوات :

اختيار الواسمات الجزيئية وهذه قد تكون تواليات للنوكليوتيدات او تواليات بروتينية ويكون الاختيار مهما اذ انه يؤثر في الحصول على الشجرة الصحيحة. والاختيار هنا يعتمد على مواصفات التواليات وكذلك الغرض من الدراسة. فعند دراسة أحياء قريبة من بعضها جدا يمكن استعمال توالي النوكليوتيدات الخاصة بجينومات المايكروبيات لانها أسرع تطورا من البروتينات (نظرا للأسباب المذكورة حول الشفرات وعلاقتها مع الحوامض الأمينية التي ذكرت في بداية الفصل).

اما عند دراسة الأحياء المتباعدة فالاختيار يحتاج الى استعمال تواليات بطيئة التغير وقد تكون هذه تواليات نيوكليوتيدية مثل تواليات RNA الرايبوزومي وخاصة 16S rRNA او استعمال تواليات بروتينية، وفي حالة كون الدراسة تهدف الى دراسة أحياء متباعدة جدا كما في مقارنة البكتريا مع الخلايا حقيقية النواة فيفضل استعمال تواليات البروتينات الثابتة وترك أنواع البروتينات الأخرى والتواليات النيوكليوتيدية.

وفي العديد من الأحيان يفضل استعمال التواليات البروتينية لأنها أكثر ثبوتاً من التواليات النيوكليوتيدية نظراً لوجود حقيقة اختلاف الشفرات الوراثية Codon degeneracy (يراجع الفصل الثالث) وبالتالي فالتغيير في الشفرة الوراثية وخاصة في القاعدة النتروجينية في الموقع الثالث من الشفرة ليس من الضروري أن يؤدي إلى تغيير الحامض الأميني الناتج منها . وهذا يعني أن التواليات النيوكليوتيدية هي مجال كبير للتغيير وهذا لا يلاءم القواعد المنصوص عليها لبناء أشجار تطورية ، في حين أن البروتينات لا تعاني من هذه المشكلة بشكل كبير حتى في التواليات المتباعدة . فضلاً عن ما ذكر أعلاه فإن التفاضل في استخدام الشفرات الوراثية للحامض الأميني الواحد يشكل مجالاً آخر للاختلاف.

ومن جهة أخرى يلاحظ أن التواليات البروتينية تكون أكثر حساسية للاصطفاف من DNA وذلك لأن الأولى لها 20 مفردة وهي الحوامض الأمينية للاصطفاف في حين هناك 4 مفردات للـ DNA وهي القواعد النتروجينية. وقد وجد بالتجربة أن إمكانية تطابق تواليين من DNA متباعدة جداً يمكن أن يصل إلى 50% في حين تنخفض النسبة في البروتينات إلى 10%. ولذلك كان من غير الممكن استعمال التواليات النيوكليوتيدية للاصطفاف في حالة الأحياء ذات العلاقة متوسطة القرب، فضلاً عن أن اصطفاف تواليات DNA المشفرة للبروتينات حتى مع إدراج الفجوات لتضخيم درجات الاصطفاف يكون غير مجدياً لحدوث أخطاء في قراءة الشفرات الوراثية وبالتالي تكون عمليات الاصطفاف لا معنى لها من النواحي البايولوجية. وعلى الرغم من تفضيل استعمال التواليات البروتينية إلا أن التواليات النيوكليوتيدية تبقى تزود بالعديد من المعلومات في بعض الحالات مثل في حالة التواليات القريبة من بعضها جداً إذ أنها تصف الاستبدالات المترادفة وغير المترادفة التي تكون بمثابة شواهد توضح أحداث الانتخاب الإيجابي أو السلبي.

ولغرض فهم أحداث الانتخاب السلبي أو الإيجابي لابد من التمييز بين الاستبدالات المترادفة والأخرى غير المترادفة . ففي حالة الاستبدالات المترادفة يحدث تغيير في التوالي المشفر دون أن يؤثر ذلك في الحامض الأميني في البروتين الناتج وذلك لوجود عدد كبير من الحوامض الأمينية التي يشفر لها بأكثر من شفرة وتصل ذروتها في حالة الأرجنين والليوسين والسيرين الذي يشفر لها بست من الشفرات الوراثية فضلاً عن وجود أربع شفرات تشفر للثريونين وكذلك البرولين والكلايسين والالنين والبقية تتراوح بين 2-3 شفرات لكل حامض أميني، أما الاستبدالات غير المترادفة فهي تمثل التغيير في الشفرة الوراثية الذي يؤدي إلى تغيير في الحامض الأميني الذي تشفر له وبالتالي تغيير توالي البروتين.

ولذلك فإن مقارنة نوعي الاستبدالات المذكورة يساعد في فهم عملية التطور الحاصلة للتواليات. فعند حصول الاستبدالات غير مترادفة على سبيل المثال بمعدلات أعلى من الاستبدالات المترادفة فإن ذلك سيؤدي إلى تغيير البروتينات أي أنها تعاني من معدلات عالية من التطهير التي يمكن أن تساهم في بزوغ وظائف جديدة للبروتين ومثل هذه الحالة يشار إليها على أنها تطور إيجابي أو تطبيعي Adaptive evolution . أما في حالة حدوث العكس وزيادة الاستبدالات المترادفة بمعدلات عالية فإنه يؤدي إلى تغييرات معتدلة على مستوى الحوامض الأمينية وبالتالي البروتينات التي توجد فيها ويكون هذا هو الانتخاب السلبي أو الانتخاب التنقي Purifying selection وهذا ما يحصل في البروتينات المهمة جداً والضرورية لحياة الخلية

وتشارك في المسارات الأيضية المهمة فهي يجب ان لا تكون تحت سقف التغيرات ولذلك يلاحظ انه مثل هذه البروتينات تكاد تكون ثابتة ومتشابهة في معظم ممالك الأحياء.

الاصطفاف في بناء الأشجار التطورية

تعد الخطوة الثانية والمهمة في بناء وإيجاد الأشجار التطورية وربما تعد هذه الخطوة هي الأهم من الباقي لأنها تؤسس لمواقع المناظرة في التطور. وبذا فالاصطفاف الصحيح هو الذي ينتج أشجار تطورية صحيحة لان المواقع المصطفة يفترض فيها ان تكون ممثلة للعلاقات النسبية ، في حين ان الاصطفاف غير الصحيح يؤدي الى حدوث أخطاء أساسية في تركيبية الشجرة وبالتالي ربما الوصول الى علاقات خاطئة . ولذلك كان لابد من العناية بعملية الاصطفاف والوصول الى الصحيح منها قدر المستطاع .

وللوصول الى اصطفافات صحيحة يجب اختيار برامج صارمة وملائمة فمثلا عند صف تواليات بروتينية يفضل استعمال برنامج T-Coffee، كما يستحب استعمال أكثر من برنامج وتمحيص نتائجها لتشخيص الاصطفافات الأكثر عقلانية كما انه يستحب ان تعرض النتائج وتقرن مع المعايير المرجعية في قواعد بياناتها وبالتالي يمكن التخلص من الأخطاء المستورثة بعمليات الاصطفاف الآلية . وغالبا ما يكون التنقيح اليدوي هاما لحسم التأكد من نوعية الاصطفاف والتي تحتاج الى التأكيد على تطابق ثملات العوامل المرافقة والثمالات ذات الصفات الفيزيوكيميائية المتشابهة فضلا عن إمكانية استعمال العناصر التي لها تراكيب ثانوية متوفرة وغيرها من النواحي يمكن ان توصل الى اصطفاف صحيح.

كما انه من الضروري التقرير في استخدام الاصطفاف الشامل او استعمال جزءاً منه لاستبعاد المناطق غير الواضحة المعالم قبل إجراء تحليل العلاقات التطورية والأجزاء التي يمكن ان تحذف تعتمد على الباحثين فالبعض يفضل استبعاد مناطق الإقحام والحذف (Indels) واستخدام المواقع المشتركة بين التواليات فقط من مجموعة البيانات، ولكن عملية الحذف لبعض الأجزاء تكون غير محبذة لأنها ستؤدي الى فقدان بعض إشارات العلاقات التطورية ، فمثلا بعض الفجوات تعود الى مواقع مهمة للحذف والإقحام التي تكون بمثابة الطمغة لتحديد المجاميع الثانوية في التواليات لذلك كان من الضروري الإبقاء عليها عند بناء أشجار العلاقات التطورية. وفضلا عن التعديل اليدوي هناك بعض برامج الحاسوب القادرة على تعديل وتمحيص وتحسين عمليات الاصطفاف وذلك بزالة المناطق المصطفة بشكل ضعيف وجعل الاصطفافات أكثر ملائمة للاستعمال في تحليل العلاقات التطورية.

وفي اصطفافات البروتينات تستعمل مصفوفات PAM و BLOSUM بكافة درجاتها. اما اصطفافات RNA وخاصة rRNA هي التي أدت الى فصل المملكة الحيوانية الى ثلاث أقسام ثانوية هي حقيقية النواة و اركيا والبكتريا قبل حوالي 1.5 بليون سنة. وتكون عمليات اصطفاف DNA سببا في بعض المشاكل، اذ ليس كل الكائنات تتكون موادها الوراثية من الجينات فهناك DNA السقط وكذلك وجود الانترونات التي يجب ان تميز عن الاكسونات في الجين الواحد، لذلك كان تحديد تواليات الجينومات ذو تأثير بسيط في تحديد العلاقات التطورية، فضلا عن ان عدد قليل من الأحياء قد تم تحديد توالياتها الجينومية فضلا عن ان هذه الأحياء قد لا توجد علاقة تطورية قوية بينها.

الاستبدالات المتعددة

ان ايسط مقياس لتحديد التباعد بين تواليين هو حساب عدد الاستبدالات في الاصطفااف. وتعرف قيمة الاستبدال بالمسافة الظاهرة بين تواليين وقد لا يمثل حوادث التطور الحقيقية التي قد حصلت فعلا، فمثلا عندما يلاحظ ان A استبدلت بـ C فان هذه الحالة قد تكون مرت عدة مراحل قبل الوصول الى الحالة الظاهرة يعني:

$$A \rightarrow C$$

قد تكون مباشرة ولكن في الوقت نفسه قد تكون غير مباشرة

$$A \rightarrow T \rightarrow G \rightarrow C$$

كما ذكر في مستهل الموضوع كما ان الملاحظات المشاهدة مثلا $G \rightarrow C \rightarrow G$ اي إمكانية حدوث طفرات راجعة أدت الى ملاحظة G هذا فضلا عن إمكانية حدوث طفرات متوازية في التواليين. ومثل هذه الاستبدالات المتعددة والانفراج يمكن ان تؤدي الى تغطية وحجب المسافات التطورية بين التواليين (يعني ما يشاهد اليوم قد لا يكون هو نفسه في الماضي) وهذا ما يعرف بالنشوء المتجانس للتغيرات والطفرات وإذا لم يصحح فانه يؤدي الى الحصول على أشجار غير صحيحة، ولذلك كان هناك ضرورة لاستعمال طرق او نماذج إحصائية لاستنتاج المسافات التطورية الصحيحة بين التواليين .

[والنشوء المتجانس Homoplasy يطلق عليه مصطلحات أخرى منها النشوء المتوازي Parallel evolution او النشوء العكسي Evolution reversal او النشوء المنفراج Convergent evolution]

والاختبارات الإحصائية اللازمة تحتاج الى بعض المقومات منها :

اختيار نماذج الاستبدال:

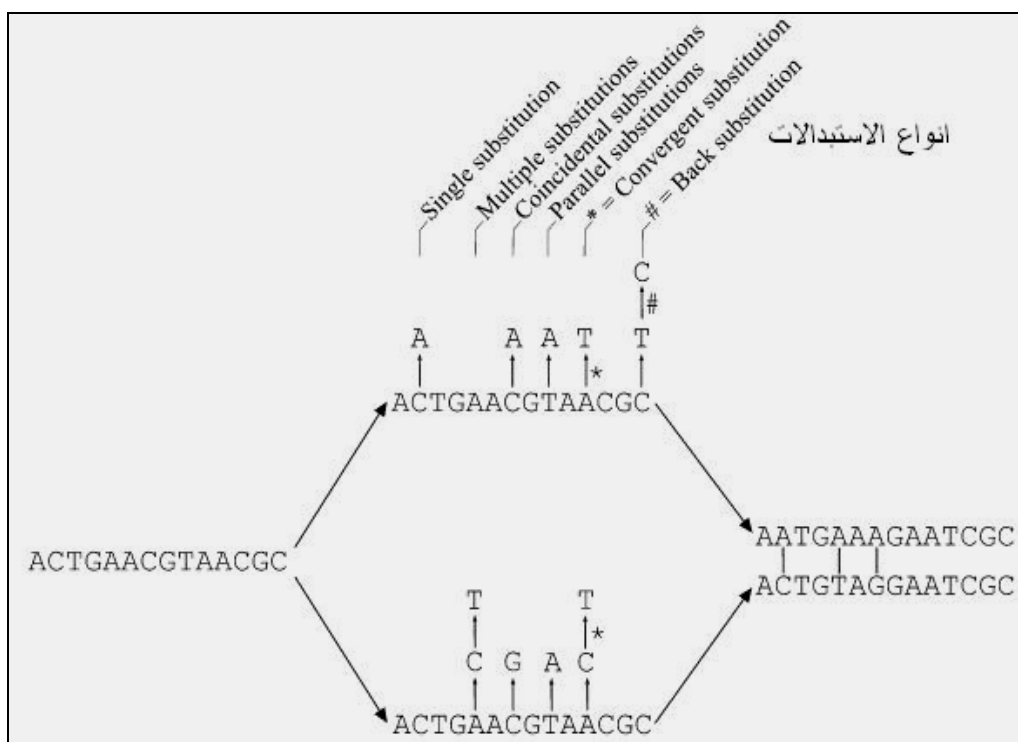
تدعى النماذج الإحصائية المستعملة لتصحيح النشوء المتجانس بنماذج الاستبدال او نماذج التطور Evolutionary models وبالنسبة لدراسة العلاقة التطورية باستعمال DNA هناك بعض النماذج المتوفرة وتختلف فيما بينها في الكيفية التي تتعامل بها مع الاستبدالات المتعددة لكل نيوكلوئيد والمعوقات في استعمالها يكون عند التعامل مع الاستبدالات الكثيرة جدا في موقع معين من التوالي والذي يكون موجودا في التواليات المتباعدة جدا وهذا يعني ان الانفراج التطوري يكون خارج نطاق النموذج الإحصائي الذي يكون غير قادرا" على تصحيحه وبالتالي لا يمكن تحديد المسافة التطورية، وهذا يعني ان أشجار العلاقات التطورية ممكنة للتواليات التي بينها تشابه الى درجة معقولة، اما التواليات المتباعدة جدا فتستعمل لها نماذج إحصائية خاصة بها.

مصفوفات المسافات Distance matrices

تستعمل مصفوفات المسافات في عدد من طرق بناء الأشجار التطورية ، وأصل تطبيقها يكون البيانات الظاهرية Phonetic data باستعمال مصفوفات المسافات الثنائية. وهذه المسافات يعاد تعديلها لإنتاج الأشجار Phylogram التي يكون فيها طول الأذرع يدل على معلومات معينة. ومصفوفات المسافات يمكن ان تأتي من أرقام من مصادر مختلفة مثل:

- اصطفاافات زوج من التواليات
- دراسات مناعية

- دراسات المظهر الخارجي
 - المسافات الوراثية
 - طول القطع الناتجة من تأثير إنزيمات القطع وغيرها من الإنزيمات.
- ويمكن تهيئة البيانات الى مفردات تستعمل في العلاقات التطورية. ومن ضمن المصفوفات المستعملة هي المسافات الوراثية Genetic distances بين تواليين والتي تمثل عدد الطفرات او الأحداث (التغيرات) التطورية بين الأنواع منذ انفراجهم عن بعض. وبسط طريقة لحساب المسافة الوراثية هو حساب عدد الاختلافات بين اثنين من التواليات. ولكن بعض الأحيان تكون هذه الطريقة غير مجدية لعكس التاريخ التطوري للتواليات عند تحديد تواليات الأحياء الموجودة في الوقت الحاضر كما موضح في الشكل 40.5 :

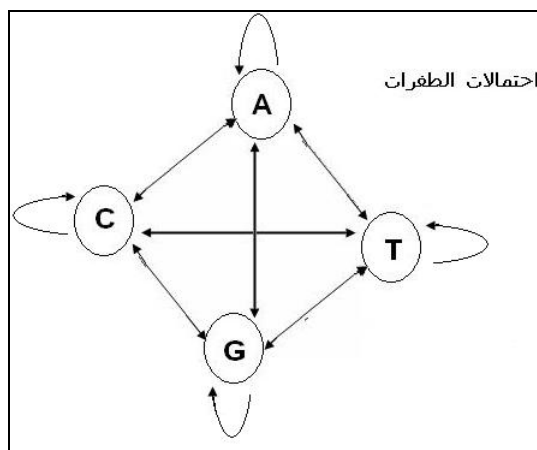


شكل 40.5 : تحديد تواليات الأحياء القديمة والأخرى الموجودة في الوقت الحاضر

ومن النماذج الإحصائية التي تتعامل مع هذه المشكلة Jukes-Cantor model . ولذلك تحتاج عمليات الرسم الى اصطفاقات تواليات متعددة كمدخلات لعمليات الحساب والرسم وتحديد المسافات. وتعرف المسافات على انها الجزء ألكسري لعدم التطابق عند مواقع الاصطفاقات، والفجوات اما تهمل او تحسب على انها حالة عدم تطابق.

نموذج Jukes-Cantor

أحد النماذج الإحصائية البسيطة المستعملة في مجال بناء الأشجار التطورية ويستعمل لعمليات استبدال النيوكلووتيدات. والذي يفترض فيه ان كل النيوكلووتيدات يتم استبدالها باحتمالية متساوية كما موضح في الشكل 41.5 :



شكل 41.5 : احتمالات الطفرات في النيوكلووتيدات

والصيغة التالية تمثل عمليات استنتاج المسافات التطورية التي تشمل التغيرات غير المرئية (الخفية)

$$d_{AB} = - (3/4) \ln [1 - (4/3) P_{AB}]$$

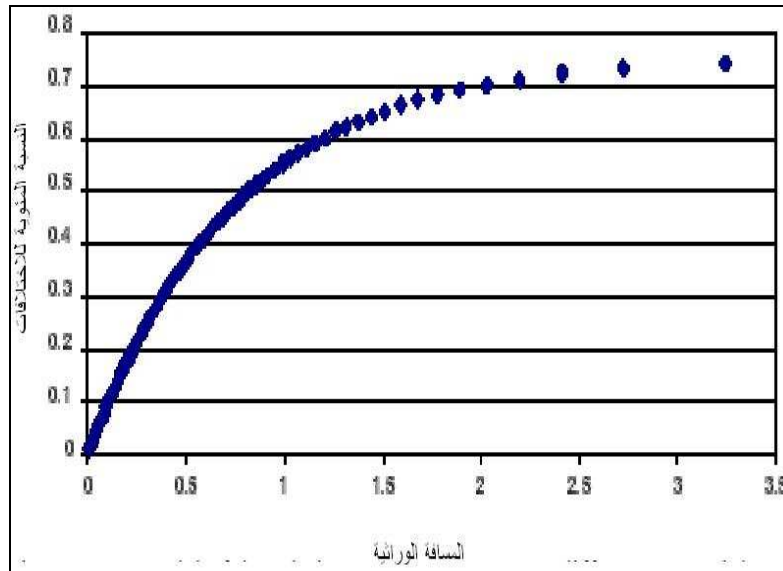
d_{AB} هي المسافة التطورية أو الوراثة بين التواليتين A و B
 P_{AB} هي مسافة التوالي المشاهدة مقاسة من نسبة الاستبدالات مقسومة على الطول الكلي للاصطفا، اي تمثل النسبة المئوية للاختلافات فعلى سبيل المثال صف تواليتين A و B كل منهما بطول 20 نيوكلووتيد ووجد ان 6 منها مختلفة، فان التواليتين يختلفان بنسبة 30% $\left(\frac{6}{20}\right)$
 ولها مسافة مشاهدة او ملاحظة 0.30 ولتصحيح الاستبدال المتعدد باستخدام نموذج Jukes-Cantor فان تصحيح مسافة التطور يكون استنادا الى المعادلة أعلاه:

$$\begin{aligned} d_{AB} &= -3/4 \ln [1 - (4/3) \times 0.30] \\ &= -3/4 \ln [1 - 0.4] \\ &= -3/4 \ln [0.6] \\ &= -3/4 [-0.5108] \\ &= 0.38 \end{aligned}$$

ولذلك فاذا كانت نسبة التشابه قليلة بحدود 25% فهذا يعني P_{AB} ستكون 0.75 وعند تطبيق المعادلة سيؤدي هذا الى ان تكون المسافة d_{AB} كبيرة جدا

$$\begin{aligned} d_{AB} &= -3/4 \ln [1 - 4/3 \times 0.75] \\ &= -3/4 [1 - 1] \\ &= \infty \end{aligned}$$

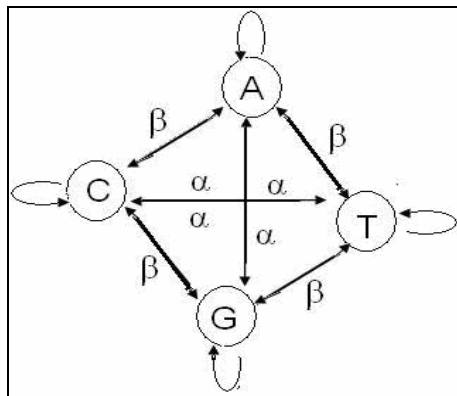
وهذا يعني ان النموذج يمكن ان يتعامل بحدود معينة مع التواليات المتقاربة وان المسافة التطورية تزداد كلما ازدادت المسافة المشاهدة، وتصل الى حد الإشباع عند 75% وعندها لا يمكن حساب المسافة الوراثية كما موضح في الشكل 42.5 :



شكل 42.5 : حد الإشباع للمسافات التطورية عند حساب المسافة الوراثية

نموذج Kimura

نموذج آخر لتصحيح المسافة التطورية ويدعى نموذج كيمورا – ثنائي المؤشرات Kimura two-parameters model ، وهو نموذج أكثر تعقيدا وفيه تكون معدلات استبدال القواعد النتروجينية من صنف واحد Transition يختلف عن معدل استبدال القواعد من صنفين مختلفة Transversion وهو الحال الأكثر واقعية ووفقا لهذا النموذج فان حدوث الحالة الأولى يحدث أكثر من حالة Transversion كما موضح في الشكل 43.5 :



شكل 43.5 : استبدالات أصناف القواعد النتروجينية

وهذا يمكن ان يعطي مسافة تطورية واقعية أكثر ويستعمل هذا النموذج الإحصائي الصيغة التالية في الحسابات:

$$d_{AB} = - (1/2) \ln (1 - 2 P_{ti} - P_{tv}) - (1/4) \ln (1 - 2 P_{tv})$$

d_{AB} مسافة التطور بين التوالين A و B

P_{ti} التكرار المشاهد لعمليات تغيير القواعد النروجينية من الصنف نفسه

P_{tv} التكرار المشاهد لعمليات تغيير القواعد النروجينية من أصناف مختلفة

وبتطبيق نموذج كيمورا للمقارنة بين التوالين A و B المختلفان بنسبة 30% (كما في المثال أعلاه) كانت نسبة الاختلاف للاستبدال الصنف الواحد هي 20% وللحالة الأخرى 10% وان المسافة التطورية تكون

$$d_{AB} = - \frac{1}{2} \ln (1 - 2 \times 0.20 - 0.10) - \frac{1}{4} \ln (1 - 2 \times 0.1) = 0.4$$

والشكل التالي (44.5) يوضح النماذج الإحصائية للنوعين المذكورة أعلاه.

A					A				
T	α				T	β			
G	α	α			G	α	β		
C	α	α	α		C	β	α	β	
	A	T	G	C		A	T	G	C
موديل Jukes-Cantor					موديل كيمورا				

شكل 44.5 : النماذج الإحصائية المستعملة لحساب استبدالات القواعد النروجينية

وفضلا عن النماذج الإحصائية المذكورة أعلاه هناك العديد من النماذج التي يمكن استعمالها. اما بالنسبة لتواليات البروتينات فيمكن استعمال النماذج الإحصائية المذكورة أعلاه لتصحيح المسافات التطورية. وبالنسبة لنموذج كيمورا تستخدم الصيغة الآتية:

$$d = - \ln (1 - p - 0.2p^2)$$

p هي المسافة المشاهدة بين اثنان من التواليات

وكل النماذج الإحصائية تعطي نتائج متشابهة عند الانفراج القليل، ولكن عند الانفراجات الكبيرة ووجود نسبة مئوية عالية من الاختلافات لابد من اختيار نماذج استبدال أخرى ملائمة.

تباينية الاستبدال

في النماذج الإحصائية المدرجة آنفا يفترض ان الاستبدال يحصل في المواقع بمعدل ثابت ولكن هذا لا يمثل الواقع، فمثلا في تواليات DNA فان الاستبدال يكون مختلفا في الشفرات الوراثة، اذ تكون القاعدة الثالثة من الشفرة الوراثة عرضة للتغيير بمعدلات أسرع من الشفرتين الأولى والثانية وهذه يتم تلافيها في الأنظمة الحيوية بظاهرة المراوغة بحيث يكون الناتج في اغلب الأحيان هو الحامض الاميني نفسه بالرغم من تغير الشفرة الثالثة وهذا ما يلاحظ عند تعدد

شفرات الحامض الاميني الواحد اذ تكون القاعدة الثالثة هي المتغيرة اما في حالة البروتينات فان تغير بعض الحوامض الامينية يكون نادرا وذلك يعود الى اعتبارات وظيفية.

والتغاير في المعدلات عند التطور يسمى تباين المعدل الموضعي Among-site rate heterogeneity والذي يمكن ان يؤدي الى ظهور نتائج كاذبة في الأشجار التطورية. وهذا التوزيع للمواقع المختلفة يتبع نموذج توزيع كما γ وهو دالة احتمالية عامة لها منحنيات توزيع بأشكال مختلفة اعتمادا على قيم او مؤشرات خاصة بالتوزيع ، ولذلك فان المسافات التطورية يجب ان تأخذ بنظر الاعتبار هذا التباين الموضعي وتصحح على ضوءه . وعامل تصحيح توزيع كما هو القيمة α وتطبق بالنسبة لنموذج Jukes-Cantor يكون وفق الصيغة الآتية:

$$d_{AB} = (3/4) \alpha [1 - 4/3 P_{AB}]^{-1/\alpha} - 1$$

حيث ان α هي عامل التصحيح للتوزيع المتباين γ اما في حالة استخدام نموذج كيمورا الإحصائي فان عملية التصحيح تكون وفق الصيغة الآتية:

$$d_{AB} = (\alpha/2) [1 - 2P_{ti} - P_{tv}]^{-1/\alpha} - (1/2) (1 - 2 P_{tv})^{-1/\alpha} - 1/2]$$

وقد استعملت قيمة التصحيح α الخاصة بالتوزيع γ في العديد من برامج بناء أشجار التطور.

طرق وبرامج بناء الأشجار التطورية

يوجد في الوقت الحاضر مجموعتان من الطرق لبناء وإيجاد أشجار العلاقات التطورية ولكل طريقة مزاياها الجيدة ومحددات استعمالها وذلك فان تمثيل التاريخ الطبيعي للحياة في نظام تصنيفي واحد يكون بعيد المنال ولكن هناك طرق تحاول تقريب هذا المنال ومنها:

المجموعة الأولى: طريقة المسافات تعتمد على المسافات التي تمثل عدم التطابق بين التواليات المصطفة لاشتقاق الشجرة. وهذه الطرق تفترض ان كل التواليات (قيد الدراسة) هي متماثلة جذريا ، وهذه الطرق يمكن ان تعطي أشجار حقيقية فيما اذا كانت أحداث الانفراج الوراثي مسجلة بشكل دقيق (في حالة استعمال تواليات الحوامض النووية) ويضطرب الانفراج عند حد أعلى عندما يكون التوالي قد وصل الى حد الإشباع من الطفرات. وتعتمد الطرق على بناء مصفوفات للمسافات وتضع التواليات المتقاربة جدا تحت العقدة الداخلية نفسها ، وطول الأذرع يشير الى المسافات المشاهدة بين التواليات. والمسافات التي تستعملها هي الناتجة من طرق الصف المتدرج Progressive alignment والاصطفاف المتكرر لعدد من التواليات (المارة الذكر).

وتنتج الطرق أشجار مجذرة وأخرى بدون جذور اعتمادا على الخوارزمية المستعملة في الحسابات. وفي بعض الأحيان يتم تجنب استعمال الطرق المعتمدة على المسافات لعدم قابليتها على استخدام المعلومات بكفاءة للمناطق عالية التغاير والتي يمكن ان تتسبب في ظهور أشجار ثانوية. كما ان بعض العلاقات يمكن ان تفقد من الشجرة أثناء عمليات اختزال المسافات للمفردات، فضلا عن ان الطرق تستعمل بيانات غير مباشرة (يعني المسافات ومصفوفاتها) مما يؤثر في توزيع المفردات على الشجرة. كما ان المعقد منها يمكن ان يؤدي الى انحراف وانحياز في المسافات . وفي كل رسم ينتج من هذه الطرق فان طول الأفرع لا يمكن في بعض الأحيان تقديره بدقة لان بعض التغيرات في التواليات لا يمكن ان تكتشف نظرا لفشل النموذج المستعمل او الخطأ في تصميم التجربة او انقراض النموذج الأصلي وتسمى هذه بتأثير كثافة

العقدة Node density effect والتي لا يمكن التخلص منها حتى بعد إجراء تعديل للمسافات في البيانات الوراثية.

المجموعة الثانية: تعتمد على مفردات خاصة متمثلة بالتواليات الجزيئية الخاصة بكل وحدة تصنيفية. والفرضية الأساسية فيها ان مواقع معينة تكون متشابهة بين التواليات المتعددة وبذا يمكن تحديد الأصل المشترك للتواليات من تتبع مجاميع المعلومات الموجودة فيها، الفرضية الأخرى لهذه الطرق هي ان كل مفردة او ثمالة تتطور بشكل مستقل لذلك تعامل على انها وحدة تطويرية مستقلة.

والبيانات المستقاة سواء من إنشاء مصفوفات المسافات او اعتماد مفردات او صفات لتحديد العلاقات التطورية التشريرية والجزيئية تكون ملائمة جدا في حالة توافق بين البيانات الجزيئية والشواهد التشريرية المأخوذة من الأحياء الموجودة في الوقت الحاضر او من سجل المتحجرات ، ولذلك كان تنوع في إيجاد العلاقات التطورية فهي تستقي المعلومات باستعمال كل أنواع البيانات مثلما ذكر منها البيانات الجزيئية وكذلك البيانات الايضية وبيانات المظاهر الخارجية والبيانات الجغرافية والجيولوجية.

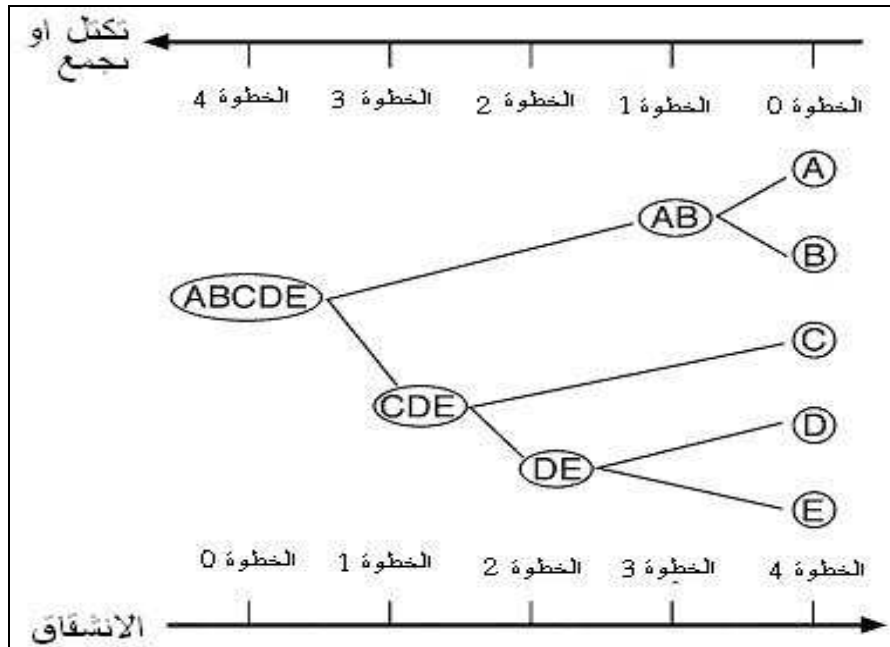
الطرق المعتمدة على المسافات

كما ذكر سابقا ان هذه الطرق تعتمد على إيجاد مصفوفات للمسافات. وتقسم الخوارزميات اللازمة (أي المعتمدة على المسافات) الى خوارزميات العقدة Clustering algorithms وتكون معتمدة على مصفوفة المسافات وتبدأ حساباتها من التواليات الأكثر تماثلا والأخرى المعتمدة على إيجاد امثل الأشجار لانتخابها Optimality algorithms والأخيرة تعتمد الى مقارنة عدد من مخططات الأشجار ثم اختيار الأفضل التي تكون أكثر ملائمة من حيث تطابق المسافات المحسوبة مع المسافات التطورية الحقيقية .

طرق العقدة Clustering methods

تستعمل الطرق الأشكال المتفرعة Dendrograms لإظهار العلاقة بين الأحياء (العناقيد)، فالأحياء القريبة من بعضها تربط بخط أفقي مكونة مجموعة او عنقود وقد يكون كائن جديد وفي الحالة الأخيرة فان المعلومات يتم البحث عنها من أقرب زوج ، وهي تعتمد على حساب المسافات كما ذكر أعلاه والتي يمكن ان تكون اي قياس متري. وطريقة العناقيد طريقة عامة تستعمل لتحليل البيانات الإحصائية في مجالات أخرى غير المعلوماتية الحيوية.

وتوجد أنواع من طرق العقدة فهي قد تكون هرمية Hierarchical وفي هذه الحالة يتم البحث او إيجاد عناقيد معينة باتباع العناقيد التي قبلها. وقد تكون متكتلة او متجمعة Agglomerative وتبدأ بكل عنصر كعنقود منفصل ثم تندمج لتكون عناقيد اكبر. او تكون منفصلة Divisive وتبدأ من القمة الى الأسفل اي تبدأ من المجموع الكلي ثم تنفصل الى عناقيد صغيرة. وتوجد العديد من الخوارزميات والبرامج التي تستعمل طرق العقدة. وتشمل طرق عديدة سيأتي ذكرها تباعا والشكل التالي (45.5) يوضح مسار كل نوع من أنواع العقدة .



شكل 45.5 : طرق العنقدة

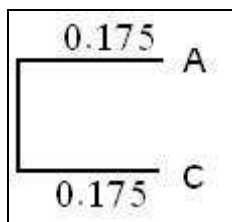
طريقة المعدل الرياضي غير المرجح (UPGMA) Arithmetic Average

وهي أبسط طرق العنقدة الهرمية إذ تبنى فيها الشجرة بطريقة العنقدة المتتالية واستعمال مصفوفات مسافات معينة واستعملت الطريقة بداية مع دراسات الترحيل الكهربائي للبروتينات ولكنها تستعمل في الوقت الحاضر لإنشاء شجرة الدليل لاستعمالها في خوارزميات العنقدة لإيجاد العلاقات التطورية . ففي المثال التالي مصفوفة مسافات لأربعة أصناف وكالاتي :

	A	B	C
B	0.4		
C	0.35	0.45	
D	0.6	0.7	0.55

وعند استعمال طريقة المعدل الرياضي غير المرجح تبدأ الطريقة باختيار المفردات التي تكون بينها أقل مسافة وهي A و C (0.35) وتوضع عقدة عند منتصف المسافة

$$\begin{aligned} AC/2 &= 0.35/2 \\ &= 0.175 \end{aligned}$$



وبذا تنشأ مصفوفة مسافات مقلصة (حيث ستعامل كل من A و C كمفردة واحدة) وتكون المصفوفة الجديدة. ومسافة التعنق بين A-C مع B هو

$$(AB+BC)/2 = \frac{0.4 + 0.45}{2} = 0.425$$

والمسافة بين A-C و D يكون

$$(AD+CD)/2 = \frac{0.6 + 0.55}{2} = 0.575$$

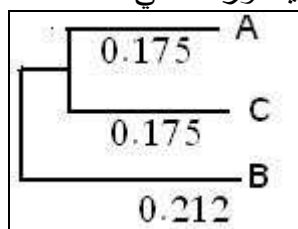
كما موضح في الآتي:

	A-C	B
B	0.425	
D	0.575	0.7

ومن هذه المصفوفة (الجديدة) تكون أقل مسافة هي بين العنقود A-C و B ومقدارها 0.425 والتي تسمح بتكوين عنقود جديد هو A-C-B وان طول الفرع بين B والعنقود A-C يمثل منتصف المسافة

$$\frac{AC-B}{2} = \frac{0.425}{2} = 0.212$$

وبذلك فان مخطط الشجرة يمكن ان يصور كالآتي:



والمجموعة الجديدة الناتجة المكونة من A-C-B تعامل كوحدة تصنيفية واحدة وهذا يعني تقليص المصفوفة الأصلية الى وحدتي تصنيف فقط هي A-C-B و D وبهذا فان المسافة الجديدة تحتسب كما يأتي:

$$(BD+AD+CD)/3 = \frac{0.7 + 0.6 + 0.55}{3} = \frac{1.85}{3} = 0.617$$

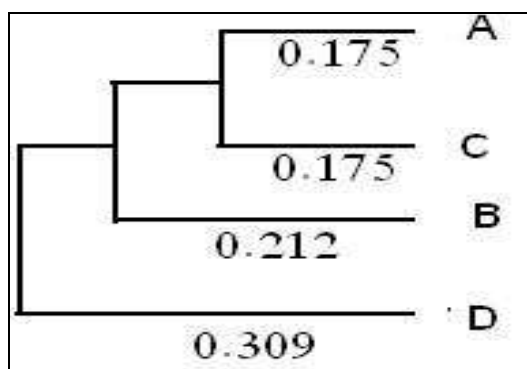
وتكون المصفوفة الجديدة المقلصة كما يأتي:

	B-A-C
D	0.617

وتكون D هي الفرع الأخير الذي يضاف الى الشجرة وان طول الفرع الذي يربطها يحسب كالآتي:

$$\frac{ACB - D}{2} = \frac{0.617}{2} = 0.309$$

ويكون مخطط الشجرة كالآتي:

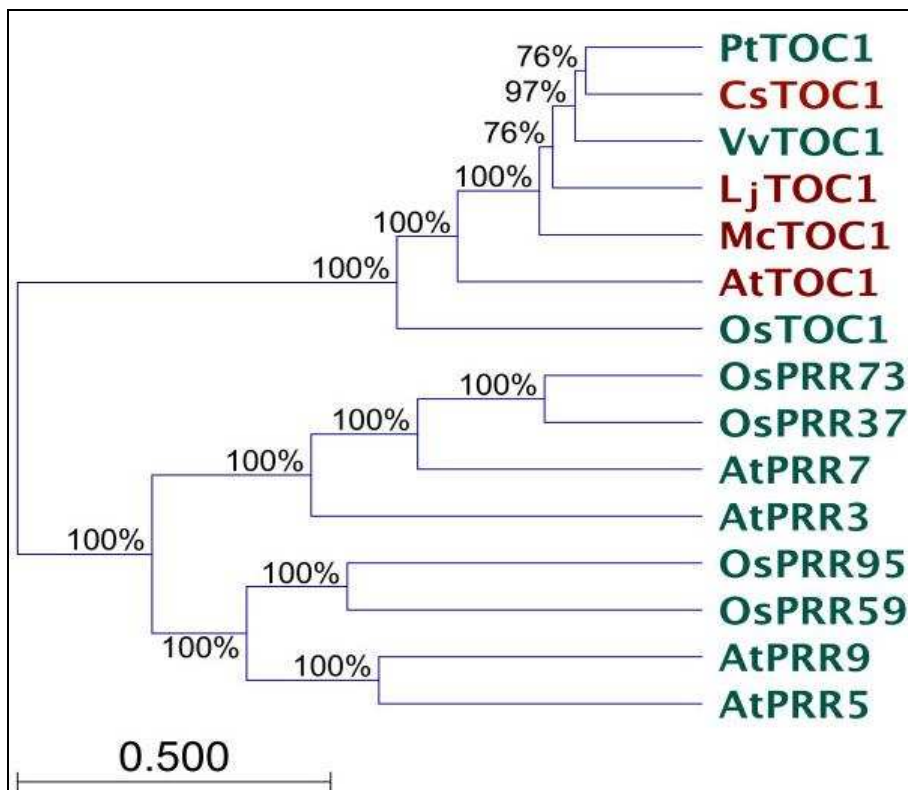


ومن الواضح ان المسافات التقديرية المحسوبة بطريقة المعدل الرياضي غير المرجح لا تطابق مسافات التطور الحقيقية الموضحة في الآتي:

	A	B	C
B	0.425		
C	0.175	0.425	
D	0.617	0.617	0.617

اي ان الطريقة تفشل في تحقيق النتائج التجريبية. وذلك يعود الى عدة أسباب منها ان الطريقة تتبنى فرضية الساعة الجزيئية وبذا تحدد أوزان متساوية (او قيم ترجيحية متساوية) . وكما لوحظ أعلاه فان الطريقة تستعمل اكبر تشابه بين صنفين وتستخرج المعدل لربط الأزواج ولذلك فهي ليست طريقة لتحديد المسافات التطورية اذ ان افتراضها للمعدل الثابت في التغيرات يؤدي في النهاية ان تكون الأفرع متساوية.

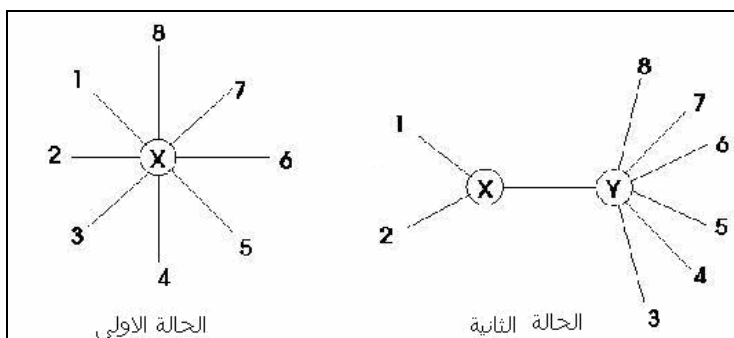
ولكن مع هذا تستعمل الطريقة على مدى ليس بالضيق نظرا لكونها سريعة ويمكن ان تستوعب عددا "كبيراً" من المفردات مثل التواليات، ويمثل الشكل التالي (46.5) أحد الأشجار الناتجة من استعمال طريقة المعدل الرياضي غير المرجح في تحديد العلاقات بين بروتينات رشاد الصخر.



شكل 46.5 : تحديد الشجرة التطورية بين بعض بروتينات رشاد الصخر *Arabidopsis thaliana* بطريقة المعدل الرياضي غير المرجح

طريقة ضم الجوار (NJ) Neighbor joining

طريقة أخرى تعتمد على مصفوفات المسافات ولا تعتمد على الساعة الجزيئية وتستعمل المسافات الوراثية كمقياس للتعنقد. ويبدأ تكوين الشجرة بشكل نجمة، ثم يتم حساب كل زوج ويقيم فيما إذا كان بالإمكان ربطه أم لا، ويتم حساب طول الأفرع لتكوين الشجرة. والزوج الذي يعطي أقل مقدار في الجمع يعد الجار الأقرب لذلك يربط ويقحم كفرع جديد كما في الشكل 47.5 :



شكل 47.5 : إنشاء الأشجار التطورية بطريقة ضم الجوار

ان استبعاد فرضية الساعة الجزئية في هذه الطريقة يعطيها دقة أكثر من طريقة المعدل الرياضي غير المرجح والتي تشترك معها في طريقة التقليل التدريجي لمصفوفات المسافات. وطريقة ضم الجوار NJ لا تفترض ان الوحدات التصنيفية تكون متساوية البعد عن الجذر وعليه تصحح نسب التطور غير المتساوية بين التواليات باستخدام خطوة التحويل وهذه تحتاج الى قيم r وتحويلها باستخدام الصيغة الآتية: $\sum$

$$d'_{AB} = d_{AB} - 1/2 \times (r_A + r_B)$$

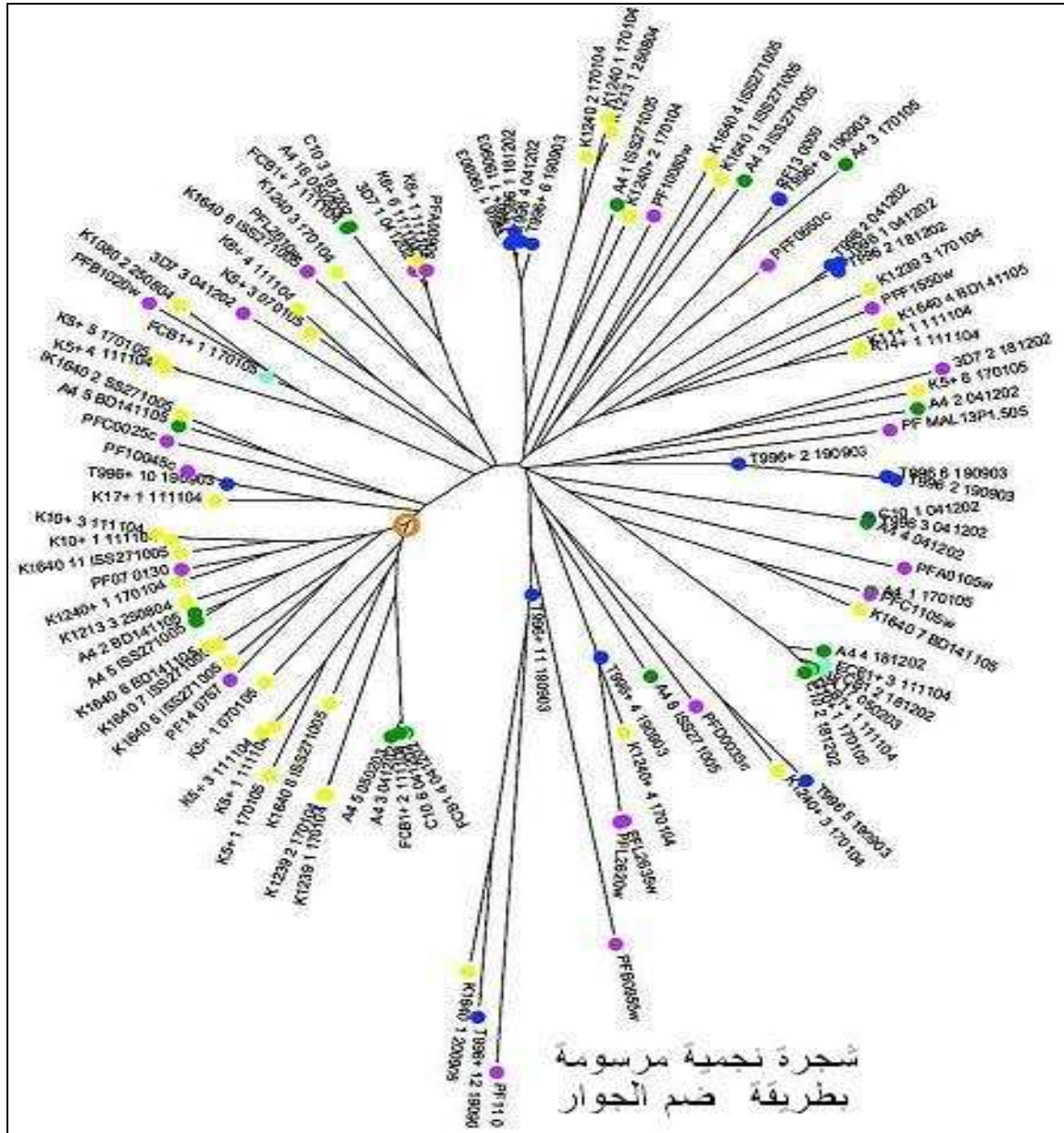
d'_{AB} هي مسافة التحويل بين وحدتي التصنيف A و B وان d_{AB} هي مسافة التطور الحقيقي بين الودعتين A و B ، و r_A و r_B مجموع المسافات لوحدي التصنيف A و B (على التوالي) الى جميع الوحدات التصنيفية الأخرى. ان التعبير العام لقيمة r هو r_i ويحسب اعتمادا على الصيغة الآتية:

$$r_i = \sum_j d_{ij}$$

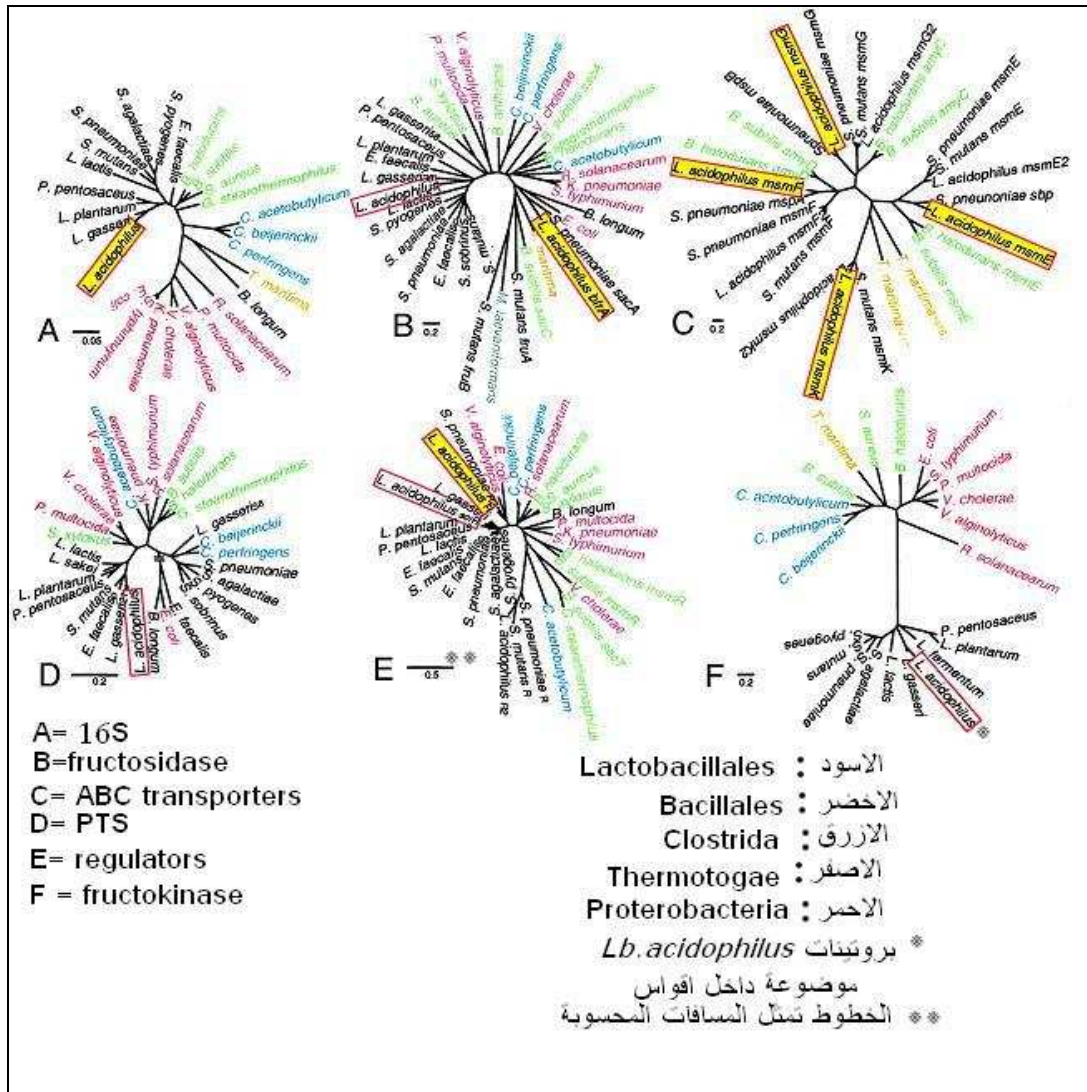
حيث ان i و j وحدتي تصنيف مختلفة. وتحسب قيم r لإنشاء مصفوفات المسافات المعدلة. في حين تستخدم قيم r المحولة (r') لتحديد المسافات لوحداث التصنيف المفردة الى اقرب عقدة.

$$r'_i = r_i / n - 2$$

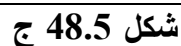
حيث ان n تمثل مجموع أعداد الوحدات التصنيفية قيد الدراسة. وتبدأ طريقة بناء الشجرة بمخطط نجمي كامل وذلك بضم جميع الوحدات التصنيفية الى عقدة مفردة واحدة ثم تجزأ الشجرة تدريجيا باختيار أزواج الوحدات التصنيفية اعتمادا على المسافات المعدلة المذكورة أعلاه. اذ تسمح الطريقة أولا بضم الوحدات التصنيفية ذات المسافات المعدلة الأقصر، وتعاد العملية حتى يتم حل جميع العقد الداخلية وهذه تسمى بتحلل النجمة، وتنتج الطريقة أشجار بدون جذور. وتختلف مخرجات البرامج المعتمدة على طريقة ضم الجوار في الأشكال التي تنتجها كما موضح في الأشكال 48.5 (أ ، ب ، ج) :



شكل 48.5 أ



شكل 48.5 ب



طريقة ضم الجوار العامة

وفضلاً عن ذلك طورت طريقة ضم الجوار المرجح Weighted NJ وتسمى Weighbor التي لها مؤشرات الخاصة الإضافية مثل مسألة الإضافة للأفرع الخارجية وكذلك مسألة الموجيية Positivity للأفرع الداخلية والتي تساعد في التحديد النوعي لعمليات ربط الأزواج وهذه الطريقة تعطي ارجحية أقل للمسافات الطويلة عند إنشاء مصفوفات المسافات وتكون أقل حساسية لبعض أنواع الانحياز ومقاومة للمسافات الطويلة التي تؤثر في طريقة ضم الجوار العادية او غيرها من الطرق. وتستعمل طريقة ضم الجوار مع البرامج المتداولة لصف البروتينات مثل مجموعة Clustal لانها أفضل من قريباتها من الطرق.

1. ان طريقة ضم الجوار ماثلة لطريقة UPGMA لكنها تستخدم خطوة تصحيح نسبة التطوير قبل بناء الشجرة كما في الجدول التالي الذي يمثل مصفوفة المسافات الحقيقية :

	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

تبدأ الخطوة الأولى ضمن طريقة ضم الجوار باحتساب قيمتي r و r' وفق المعادلات :

$$\begin{aligned}
 r_A &= AB + AC + AD = 0.4 + 0.35 + 0.6 = 1.35 \\
 r'_A &= r_A / (4 - 2) = 1.35 / 2 = 0.675 \\
 r_B &= BA + BC + BD = 0.4 + 0.45 + 0.7 = 1.55 \\
 r'_B &= r_B / (4 - 2) = 1.55 / 2 = 0.775 \\
 r_C &= CA + CB + CD = 0.35 + 0.45 + 0.55 = 1.35 \\
 r'_C &= r_C / (4 - 2) = 1.35 / 2 = 0.675 \\
 r_D &= DA + DB + DC = 0.6 + 0.7 + 0.55 = 1.85 \\
 r'_D &= r_D / (4 - 2) = 1.85 / 2 = 0.925
 \end{aligned}$$

2. استنادا الى المعادلة

$$d'_{AB} = d_{AB} - 1/2 \times (r_A + r_B)$$

والى قيم r الناتجة من الخطوة الأولى ، يمكن احتساب المسافات المصححة وكما يأتي

$$\begin{aligned}
 d'_{AB} &= d_{AB} - 1/2 * (r_A + r_B) = 0.4 - (1.35 + 1.55) / 2 = -1.05 \\
 d'_{AC} &= d_{AC} - 1/2 * (r_A + r_C) = 0.35 - (1.35 + 1.35) / 2 = -1 \\
 d'_{AD} &= d_{AD} - 1/2 * (r_A + r_D) = 0.6 - (1.35 + 1.85) / 2 = -1 \\
 d'_{BC} &= d_{BC} - 1/2 * (r_B + r_C) = 0.45 - (1.55 + 1.35) / 2 = -1 \\
 d'_{BD} &= d_{BD} - 1/2 * (r_B + r_D) = 0.7 - (1.55 + 1.85) / 2 = -1 \\
 d'_{CD} &= d_{CD} - 1/2 * (r_C + r_D) = 0.55 - (1.35 + 1.85) / 2 = -1.05
 \end{aligned}$$

3 . يمكن بناء مصفوفة المسافات الجديدة استنادا الى نسب المسافات المحتسبة في الخطوة الثانية وكما يأتي :

	A	B	C
B	-1.05		
C	-1	-1	
D	-1	-1	-1.05

4 . قبل بناء الشجرة تحدد جميع العقد للشجرة النجمية ، اذ تفصل أولاً أزواج الوحدات التصنيفية ذات المسافات الأقصر (في المصفوفة الجديدة) عن الشجرة النجمية. ومن الجدول أعلاه فإن المسافة الأقصر هي بين A و B كما هي الحال بين C و D وقيمتها -1.05 . وبهذا يمكن بناء العقدة الأولى واختيار A-B او اختيار C-D واختيار أي منهم سيؤدي الى نتائج متماثلة.

وبافتراض ان العقدة الأولى هي A-B وتسميتها u كما موضح في الآتي :



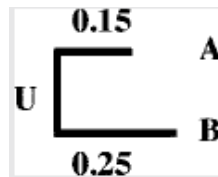
وهو يوضح مخطط اختيار العقدة u المكونة من تفرعين A و B .

5 . حساب أطوال التفرعين A و B الى العقدة u استناداً الى المعادلة وكما يأتي :

$$d_{AU} = [d_{AB} + (r'_A - r'_B)] / 2 = [0.4 + (0.675 - 0.775)] / 2 = 0.15$$

$$d_{BU} = [d_{AB} + (r'_B - r'_A)] / 2 = [0.4 + (0.775 - 0.675)] / 2 = 0.25$$

وبهذا فان الشكل أعلاه سيعدل ليصبح بالشكل الآتي :



الذي يوضح أطوال الفرعين A و B

6. تسمح العقدة الجديدة ببناء المصفوفة المختزلة الجديدة . تبدأ هذه المصفوفة بمسافات حقيقية على خلاف طريقة UPGMA . وتحسب المسافة من وحدة التصنيف الى العقدة كما يأتي :

$$d_{CU} = [(d_{AC} - d_{UA}) + (d_{BC} - d_{UB})] / 2 = [(0.35 - 0.15) + (0.45 - 0.25)] / 2 = 0.2$$

$$d_{DU} = [(d_{AD} - d_{UA}) + (d_{BD} - d_{UB})] / 2 = [(0.6 - 0.15) + (0.7 - 0.25)] / 2 = 0.45$$

وبهذا تصح المصفوفة المختزلة الجديدة الموضحة في الآتي :

	U	B
C	0.20	
D	0.45	0.55

المصفوفة المختزلة الجديدة

7. واستنادا الى هذه المصفوفة المختزلة الجديدة ينم حساب مجموعة قيم كل من r و r' الجديدة وكما يأتي:

$$\begin{aligned}
 r_C &= CU + CD = 0.2 + 0.55 = 0.75 \\
 r'_C &= r_C / (3 - 2) = 0.75 / 1 = 0.75 \\
 r_D &= DU + CD = 0.45 + 0.55 = 1 \\
 r'_D &= r_D / (3 - 2) = 1 / 1 = 1 \\
 r_U &= CU + DU = 0.2 + 0.45 = 0.65 \\
 r'_U &= r_U / (3 - 2) = 0.65 / 1 = 0.65
 \end{aligned}$$

8. ان قيم r تسمح ببناء مصفوفات المسافات الجديدة وكما يأتي :

$$\begin{aligned}
 d'_{CU} &= d_{CU} - 1/2 * (r_C + r_U) = 0.2 - (0.75 + 0.65) / 2 = -0.5 \\
 d'_{DU} &= d_{DU} - 1/2 * (r_D + r_U) = 0.45 - (1 + 0.65) / 2 = -0.375 \\
 d'_{CD} &= d_{CD} - 1/2 * (r_C + r_D) = 0.55 - (0.75 + 1) / 2 = -0.325
 \end{aligned}$$

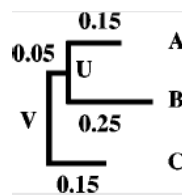
وبهذا تصبح مصفوفة المصححة الجديدة كما في الآتي :

	U	C
C	-0.5	
D	-0.375	-0.325

9. في مصفوفة المسافات المصححة الجديدة أعلاه ، يلاحظ ان المسافة الأقصر هي بين وحدة التصنيف C والعقدة u وقيمتها -0.5 وهذا يسمح بالتهيئة للعقدة الثانية التي يطلق عليها V ويحتسب طول الفرع كما في الخطوة (5) وكما يأتي :

$$\begin{aligned}
 d_{CV} &= [d_{CU} + (r'_C - r'_U)] / 2 = [0.2 + (0.75 - 0.65)] / 2 = 0.15 \\
 d_{UV} &= [d_{CU} + (r'_U - r'_C)] / 2 = [0.2 + (0.65 - 0.75)] / 2 = 0.05
 \end{aligned}$$

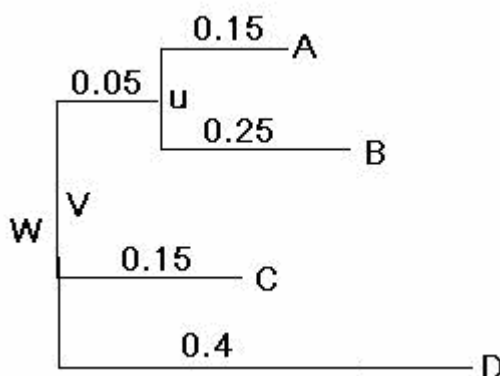
وبهذا فان الشكل سيكون كالآتي :



ويوضح الشكل شجرة علاقة التطور مؤلف من العقدة V (u و C) 10 . ولكون ان D هو الفرع الأخير الذي يحتل من الشجرة النجمية ، فلا يوجد هناك حاجة لتحويل قيم r و r' وذلك لان قيمة r' هي ما لانهاية (كبيرة جدا) عندما $n-2=0$ (n هي V و D) ويحتسب طول الفرع D كما يأتي :

$$d_D = [(d_{DU} - d_{UV}) + (d_{DC} - d_{CV})] / 2 = [(0.45 - 0.05) + (0.55 - 0.15)] / 2 = 0.4$$

وبهذا يصبح مخطط شجرة علاقة التطور كما يأتي :



المخطط يمثل علاقة التطور النهائي المكون من العقدة W (A,B,C,and D) 11 . عندما تترجم جميع أطوال الفروع الى مصفوفة المسافات تستخدم لمقارنتها مع مصفوفة المسافات الأصلية وكما يأتي :

$$A-B=0.15+0.25=0.4$$

$$A-C=0.15+0.05+0.15=0.35$$

$$A-D=0.15+0.05+0.4=0.6$$

$$B-C=0.25+0.05+0.15=0.45$$

$$B-D=0.25+0.05+0.4=0.7$$

$$C-D=0.15+0.4=0.55$$

من الواضح ان المسافات المقدرة تتطابق بالكامل مع المسافات الحقيقية . وبهذا فان هذه الطريقة تؤشر ان لها القابلية على تحقيق القيود الخاصة بالملاحظات التجريبية في هذه الحالة .

الطرق المعتمدة على الامثلية Optimity-based methods

مجموعة أخرى من الطرق المعتمدة على المسافات ولكن دون اتخاذ سبيل العقدة كأساس لإعطاء نتائجها ، ويصعب تحديد أفضلها نتيجة لعدم توفر معايير للحكم. اما الطرق المعتمدة على الامثلية فلها خوارزميات خاصة لمقارنة جميع مخططات الأشجار المحتملة واختيار الشجرة التي لها أفضل تطابق مع مصفوفات مسافات التطور الحقيقية. واعتماد هذه الطرق على البحث الشامل للشجرة الأمثل يؤدي بالضرورة الى إبطاء الحسابات وتكون المشكلة واضحة عندما تكون اعداد البيانات كبيرة ، ومن هذه الطرق:

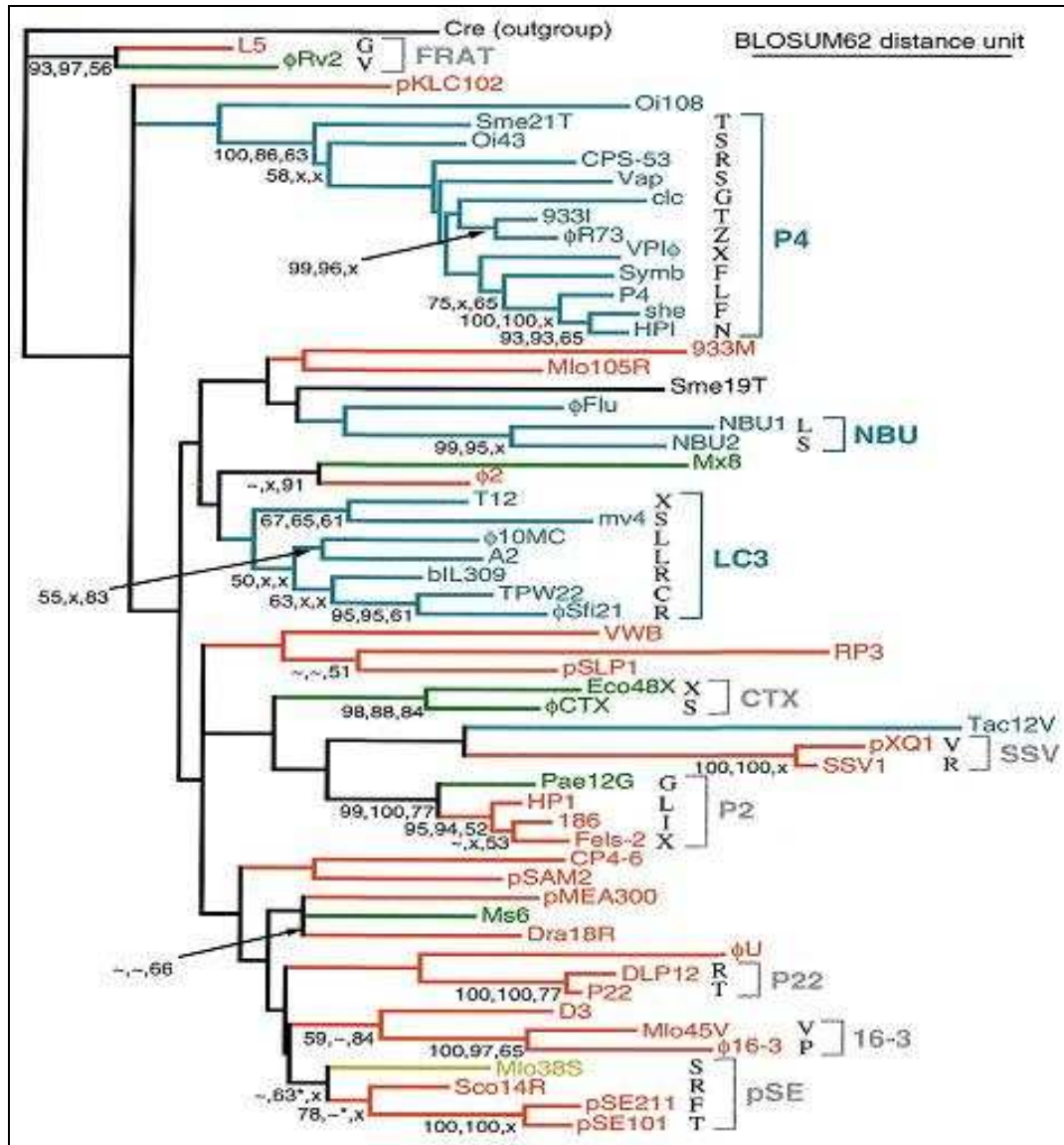
(FM) Fitch-Margoliash

تختار الطريقة أفضل شجرة من بين جميع الأشجار المحتملة بالاعتماد على طريقة المربعات الصغرى الموزونة Weighted least square بين المسافات المحسوبة في جميع التفرعات في الشجرة والمسافات في قاعدة البيانات الأصلية. حيث تبدأ بعقدة عشوائية بين وحدتي تصنيف في عقدة ثم تنشأ معادلات لوصف المسافات، ومن ثم تحل المعادلات الجبرية لأطوال التفرعات غير المعلومة. وعملية عقدة وحدتي التصنيف يساعد في إنشاء مصفوفة جديدة مقلصة ويتم تكرار العملية حتى يتم الوصول الى الشجرة بأكملها. وتبحث الطريقة عن جميع مخططات الشجرة ثم يتم اختيار شجرة واحدة لها مربع الانحراف الأقل للمسافات الحقيقية والمحسوبة لأطوال تفرعات الشجرة وفيما يلي معيار الامثلية التي يعبر عنه بالصيغة الآتية:

$$E = \sum_{i=1}^{T-1} \sum_{j=i+1}^T \frac{(d_{ij} - p_{ij})^2}{(d_{ij})^2}$$

E هو مقدار خطأ التطابق بين البيانات المحسوبة والبيانات الحقيقية ، T عدد وحدات التصنيف، dij المسافة التدريجية بين وحدتي التصنيف، i و j في مجموعة البيانات الأصلية، pij طول فرع الشجرة المناظرة .

ان عملية تصحيح المسافات يكون ضروريا" عمليا عندما يكون هناك اختلافا" في معدل التطور بين الأفرع وحساب مربع الانحراف الأقل يكون أكثر دقة. والشكل التالي (49.5) يوضح العلاقة بين إنزيمات الاندماج Integrases لأنواع من DNA باستعمال مصفوفة البروتينات BLOSUM62.



شكل 49.5 : العلاقات التطورية لإنزيمات الاندماج Integrases للـ DNA مختلفة

التطور الأقل (ME) Minimum evolution

تعمل الطريقة بشكل مشابه لطريقة FM لبناء أشجار التطور ولكنها تستعمل معيار أمثلية مختلف. وتبحث الطريقة عن كل الأشجار المحتملة مع احتساب أقل أطوال للأفرع . ويحدد معيار الامثلية للطريقة بالصيغة الآتية:

$$S = \sum b_i$$

b_i هو طول الفرع i

والبحث عن أقل مجموع لأطوال الفروع هو مؤشر لانجاز أفضل مطابقة لأطوال الفروع مع مجموعة البيانات الأصلية .

الطرق المعتمدة على أساس الرمز Character-based methods

تسمى أيضا بالطرق المتقطعة Discrete methods وهي الأقل تعقيدا من النواحي الحسابية مقارنة بالطرق المعتمدة على المسافات، والصفة أو المفردة تهمل بعد اشتقاق مصفوفة المسافة منها، وتستعمل كل المعلومات المعروفة من النواحي التطورية مثل الاستبدالات الفردية على طول التواليات لتحديد أفضل أصل أو أفضل جد.

فهي تعتمد بصورة مباشرة على حروف التوالي والمسافات التدريجية وبذلك تتجنب ضياع المعلومات عندما تحول الحروف إلى مسافات. وتضم طرق المفردات مجموعة من الطرق منها:

طريقة الاقتصاد الأعظم (MP) Maximum parsimony

تعتمد الطريقة على فرضية أن أبسط الإيضاحات للظواهر التي يتم مشاهدتها هي الأكثر احتمالا وبذا فهي تحتاج إلى أبسط إيضاحات للبيانات وبالتالي تحتاج إلى أقل ما يمكن من عمليات التنبؤ، لأن التفسيرات البسيطة تحتاج فرضيات بسيطة وخطوات منطقية قليلة. وتختار الأشجار التي لها أقل تغيير في العلاقات التطورية والتي تكون بأقل أطوال للفروع بصورة عامة. أن التعامل مع المشاكل التي لها أعداد كبيرة جدا من الحلول يكون اختيار النموذج الأبسط طريقا لتقليل المتغيرات التي تكون غير ضرورية لتوضيح الظاهرة وبذلك تكون العملية أسهل.

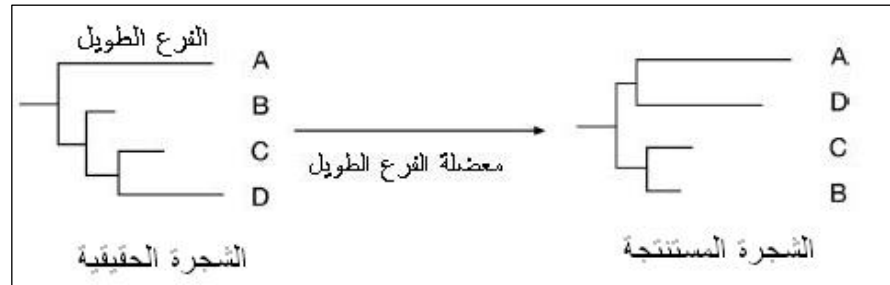
أن فرضية الاقتصاد تبدو جيدة في تحليل العلاقات التطورية بافتراض أن الشجرة التي يكون فيها أقل ما يمكن من الاستبدالات ربما تكون هي الأفضل لإيضاح الاختلافات بين وحدات التصنيف قيد الدراسة، وهذا يبدو واضحا من خلال الحقيقة التي مفهمها أن التغيرات في علاقة التغيرات التطورية تكون نادرة على نطاق زمني قصير، ولذلك يبدو أن هذه الأشجار ذات التغيرات القليلة ربما تكون هي الشجرة الحقيقية وبتقليل التغيرات تقل الأخطاء أو الضوضاء الناتجة عن النشوء المتجانس Homoplasy والتطور المستقل.

ويلاحظ أن الأسس التي تعتمد عليها الطريقة تشابه طريقة التطور الأقل ME ما عدا أن الأخيرة تعتمد على المسافات بدلا من المفردات أو الصفات. وعند إجراء تحليل التطور الأقل للمسافات المعتمدة على مجاميع مفردات مميزة فإنها تعطي الشجرة الاعتيادية للاقتصاد الأعظم للمعلومات نفسها.

وطريقة الاقتصاد الأعظم أسرع من طريقة الاحتمال أعظم (ML) وتعطي معلومات قليلة حول طول الأفرع وتعاني من مشكلة وجود الأفرع الطويلة التي يمكن أن ترتبط بشكل غير حقيقي نتيجة لتجمع التشابهات الداخلية In homologous similarity بالرغم من عدم وجود علاقة تطورية ويمكن أن تعطي أكثر من شجرة ولكن بالدرجات نفسها.

معضلة الفرع الطويلة

تعد مسألة وجود فرع طويل في الشجرة معضلة خاصة ترافق بعض الطرق المستعملة. إذ إنها يمكن أن تشير إلى علاقة تطورية كاذبة وتعمل على جذب الوحدات التصنيفية إلى مواقع أخرى في حالة التفرعات الطويلة في الشجرة ذات التطور السريع والتي قد تكون غير صحيحة كما موضح في الشكل 50.5 :



شكل 50.5 : معضلة الفرع الطويل

وذلك يعود إلى إن بعض طرق الاقتصاد المستعملة في إيجاد الأشجار التطورية تفترض إن جميع خطوط التطور تتطور بمعدل متساوي وإن حصول الطفرات سواء التبادل بين مجاميع القواعد النتروجينية المختلفة أو التبادل ضمن المجموعة الواحدة تساهم بشكل متساوي في تحديد طول الأفرع . أو أنها تكون نتيجة لحدوث عمليات الاستبدالات المتكررة في مواقع معينة بحد ذاتها أو بين المواقع مما يؤدي إلى اختلاف معدل التجانس الذي لا يمكن تصحيحه في بعض الطرق مثل الاقتصاد الأعظم وغيرها .

بناء شجرة الاقتصاد الأعظم

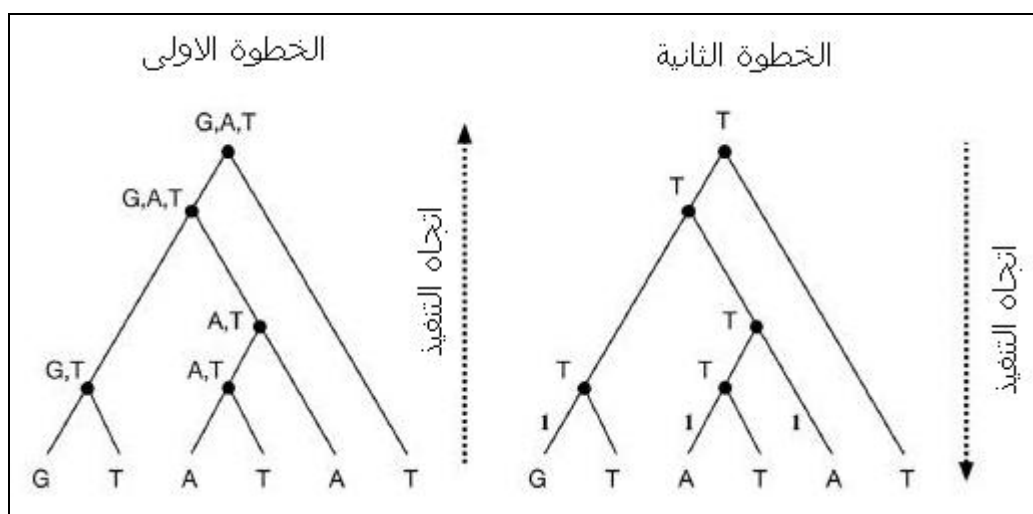
تعمل الطريقة بالبحث عن جميع مخططات الأشجار المحتملة ثم إعادة بناء تواليات الأصل أو الجد الأعلى التي تحتاج أقل عدد من التغيرات لتطور التواليات الحالية. ولغرض تقليص الوقت اللازم للحساب تستخدم عددا "قليلًا" من المواقع التي تمتلك معلومات غنية عن العلاقات التطورية في تحديد الشجرة ، ويطلق على هذه المواقع بمواقع المعلومات وهي التي تتكرر على الأقل مرتين في الوحدات التصنيفية كما موضح في الجدول 5.5 :

جدول 5.5 : مواقع المعلومات المستعملة في بناء شجرة الاقتصاد الأعظم

المواقع sites الوحدة التصنيفية taxa								
	1	2	3	4	5	6	7	8
I	A	A	T	T	A	G	C	T
II	G	G	T	C	G	T	A	G
III	A	A	T	G	C	G	C	T
IV	A	G	T	A	A	G	C	A
V	A	C	T	T	C	G	C	G
VI	A	C	A	T	G	G	C	A

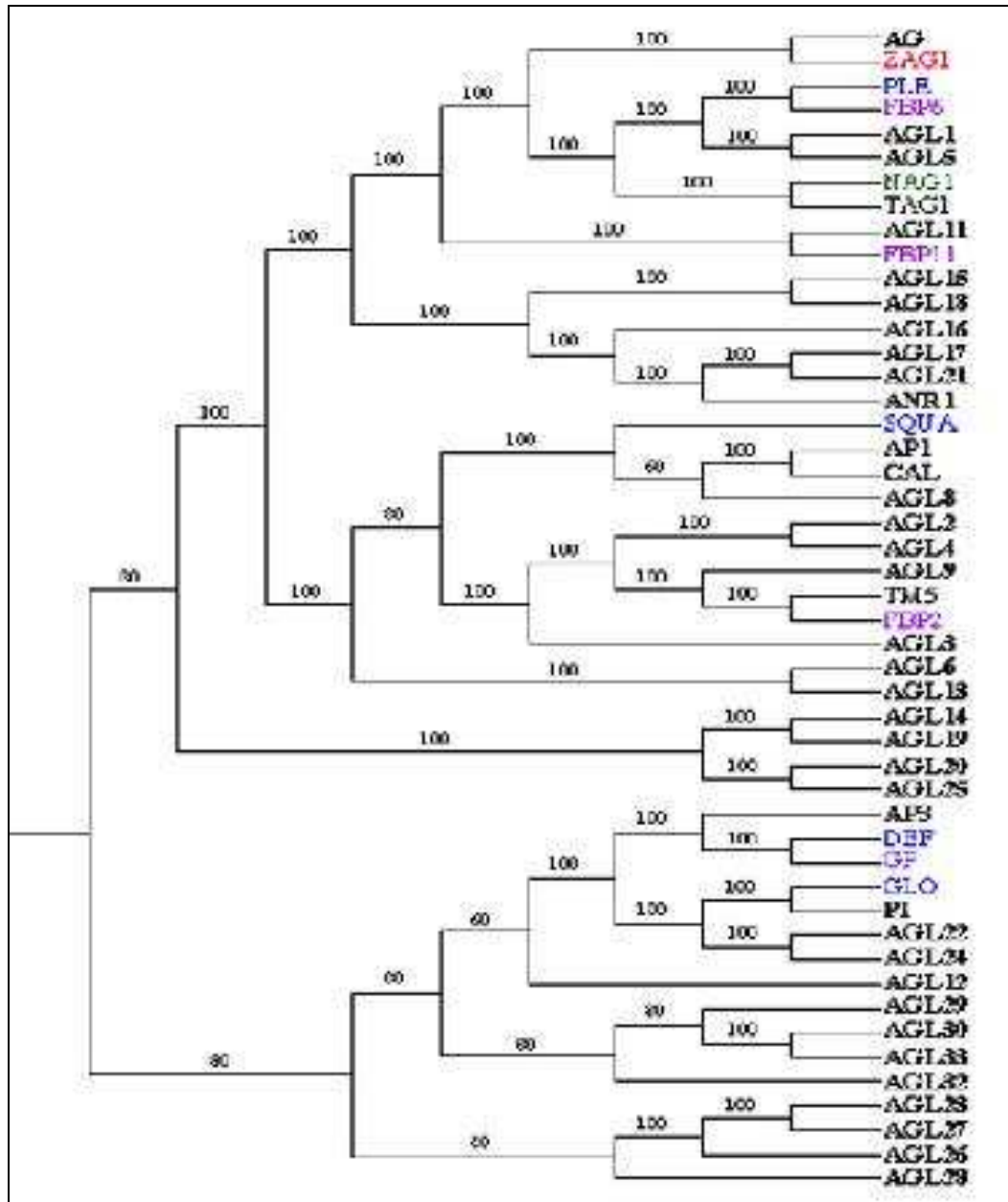
وفي الجدول (أعلاه) تكون المواقع 2، 5، 8 هي مواقع المعلومات، أما المواقع الأخرى فهي إما مواقع ثابتة أو تتغير من وحدة إلى أخرى وهي تكون متماثلة في معظم الوحدات التصنيفية

وتكون عديمة الفائدة في الطريقة الاقتصادية لبناء الأشجار لذلك تهمل وهي التي يحدث فيها التغيير لمرة واحدة . يجمع عدد التغيرات لجميع مواقع المعلومات لكل مخطط شجرة محتمل ويتم اختيار الشجرة التي لها أقل عدد ممكن من التغيرات. ولغرض حساب أقل عدد ممكن من التغيرات في شجرة ما يمكن استخدام المخطط الموضح في الشكل 51.5 :



شكل 51.5 : حساب أقل عدد ممكن من التغيرات في الشجرة عند استعمال طريقة الاقتصاد الأعظم

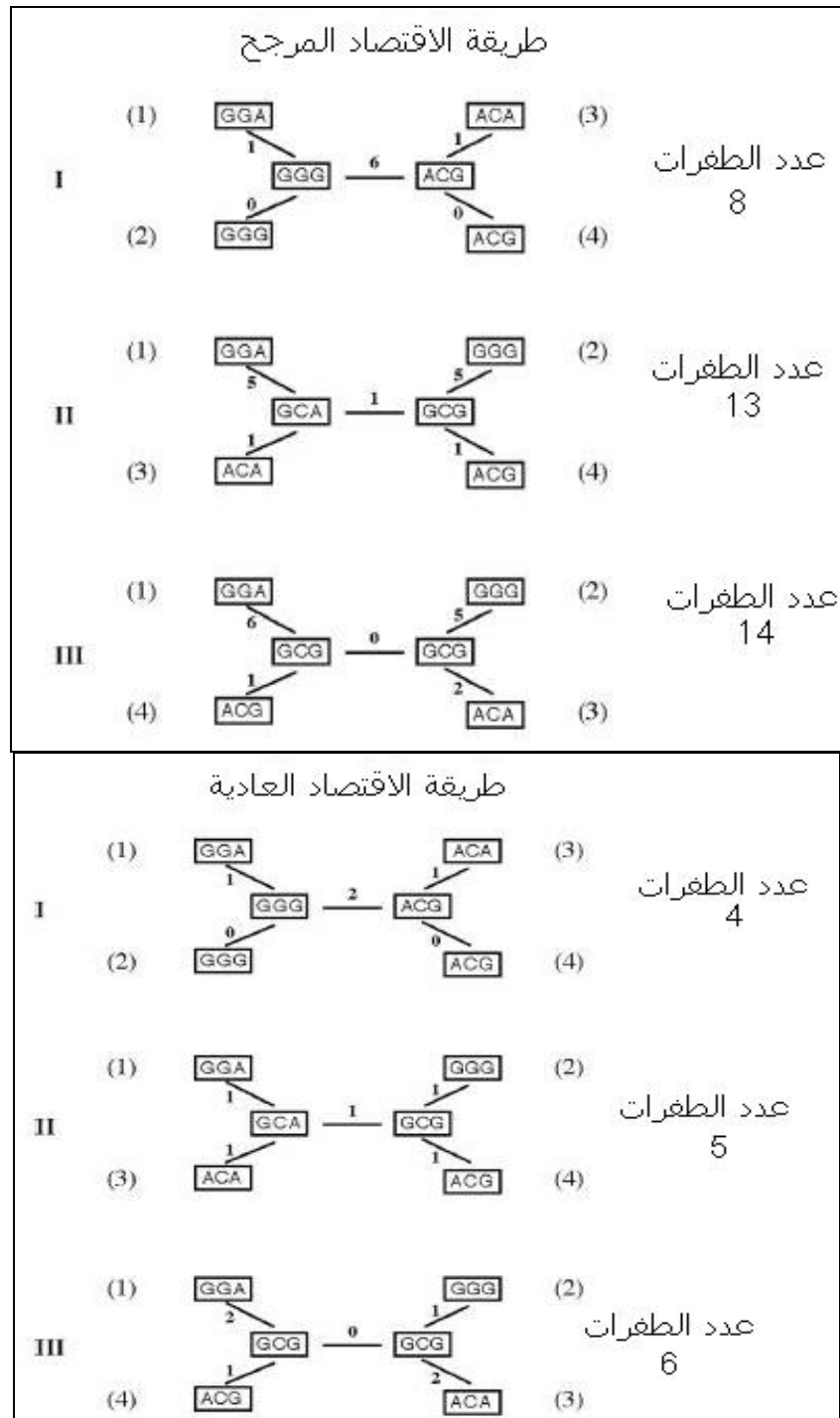
فالطريقة تعتمد الى إيجاد أقل ما يمكن من التغيرات وهناك عدة طرق أحداها الموضحة في الشكل أعلاه ، ونظرا لعدم معرفة مفردة الأصل فتعتمد الطريقة الى المرور على الشجرة المحتملة (اليسرى) من أطراف الشجرة (الأوراق) الى العقد الداخلية ثم الوصول الى الأعلى (الجذر)، ثم العودة من الجذر الى الأوراق لغرض اختيار أقل الاستبدالات الممكنة وفي هذه الحالة تكون 3 كما موضح في الجزء الأيمن في الشكل وان مفردة الأصل هو T نظرا لكونها أكثر تكرارا . والشكل التالي (52.5) يوضح أحد أشجار العلاقة التطورية لمجموعة من جينات رشاد الصخر *A. thaliana* (MADS-box gene family).



شكل 52.5 : شجرة العلاقات التطورية باستعمال طريقة الاقتصاد الأعظم لمجموعة من جينات رشاد الصخر *A. thaliana*

طريقة الاقتصاد الأعظم المذكورة أعلاه هي طريقة غير موزونة أو غير مرجحة لأنها تتعامل مع التغيرات بصورة متساوية . ولكن التغيير في بعض المواقع يكون أكثر تكرارا مقارنة بمواقع أخرى والعكس صحيح ، لذلك يمكن ان تعالج المشكلة باستعمال طريقة الاقتصاد الموزون أو المرجح Weighted parsimony إذ تأخذ هذه الطريقة بالحسبان عدد التغيرات فضلا عن نوعيتها مثل حالة حصول تبادل بين مجاميع القواعد النتروجينية أو حصول تبادل

لقواعد في المجموعة نفسها والشكل (53.5) يوضح الفرق بين طريقة الاقتصاد الأعظم العادية وبين طريقة الاقتصاد الموزون، وبطبيعة الحال فان الأخيرة يمكن ان تنتج أشجار أكثر دقة.



شكل 53.5 : الفرق بين طريقة الاقتصاد الأعظم العادية والمرجحة

طريقة الاحتمال الأعظم (ML) Maximum likelihood

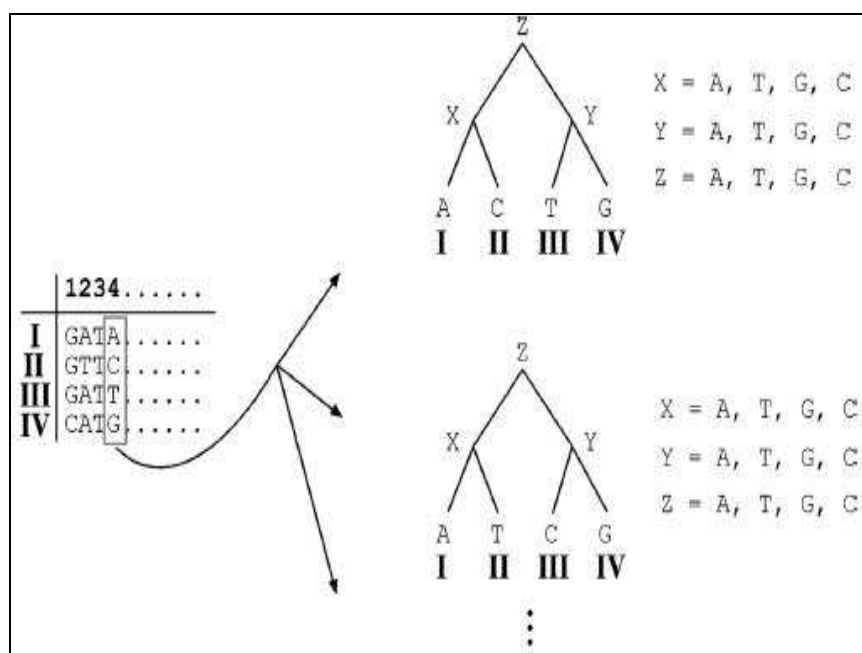
تستعمل الطريقة لإيجاد العلاقات التطورية باستخدام النماذج الاحتمالية لاختيار أفضل شجرة التي توافق البيانات المشاهدة. وتبدأ بنموذج بسيط ثم تتدرج اعتمادا على حدوث التغيرات التطورية في الحوامض النووية أو البروتينات أي ان الشجرة تمثل نمط التغيرات التطورية الحقيقية. وتعد الطريقة شاملة وتبحث في جميع المخططات المحتملة وتدفق في كل موقع من مواقع اصطفاف التواليات المتعددة التي هي بطبيعة الحال امتداد لاصطفاف زوج من التواليات. وتعمل الطريقة باحتساب الاحتمالية لمسلك تطور معين لتوالي معين، فتحدد قيم الاحتمالية باستعمال نماذج الاستبدال اما للحوامض النووية أو الحوامض الامينية ، فمثلا لتواليات الحوامض النووية يستخدم نموذج Jukes-Cantor للتعديل وباستعمال النموذج تبقى الاحتمالية p ثابتة عند مرور زمن مقداره t والتي تحسب بالصيغة الآتية:

$$p(t) = 1/4 + 3/4 e^{-\alpha t}$$

α : نسبة استبدال النيوكليوتيدات في نموذج Jukes-Cantor وتحدد اما تجريبيا او تستنتج من مجموعة من البيانات الخام. والشكل (54.5) يوضح طريقة بناء الشجرة لأربعة وحدات تصنيفية (IV, III, II, I) التي تكون أصولها غير معروفة رموزها X, Y, Z، وبوقت محدد ، يؤشر الرقم 1 الى الوقت المنقضي من انتقال X الى A والوقت 2 يؤشر الى الوقت المنقضي لانتقال Z الى A. وهناك احتمالية تغير الثمالات المشار اليها X, Y, Z الى اي من الثمالات الأخرى (T, G, C, A) في وقت مقداره t فان الاحتمالية لموقع معين تحدد بالصيغة الآتية:

$$p(t) = 1/4 - 1/4 e^{-\alpha t}$$

وتحسب الاحتمالية في الشجرة لمسار معين من ضرب الاحتمالات من الجذر الى أدق طرف في المسار. ونظرا لكون عملية الضرب تنتج أرقام صغيرة لذلك يستعاض عنها بالقيم اللوغاريتمية لتحويل الحالة الى الجمع أي يتم الحصول على اللوغاريتمات الطبيعية للاحتمالات. ونظرا لان الجد الأصلي غير معروف لذا يتم احتساب الاحتمالات لكل من X, Y, Z الموضحة في الشكل الآتي .



شكل 54.5 : طريقة الاحتمال الأعظم لبناء الأشجار التطورية

وبعد احتساب القيم اللوغاريتمية تجمع لكل فرع مفرد في الشجرة بعد احتساب جميع الاحتمالات الممكنة لمسالك الشجرة مع أجزاء التوليفات المختلفة مع الجد الأعلى للتواليات (كما ذكر آنفا)، يصبح للشجرة درجة احتمال عظمى والتي تعد المخطط النهائي عند الموقع. ولكن ان جميع المفردات قد افترض انها تتطور بشكل مستقل عن بعضها البعض فيتم احتساب درجات لوغاريتم الاحتمال الأعظم لمسار شجرة محدد ولكل موقع بصورة مستقلة.

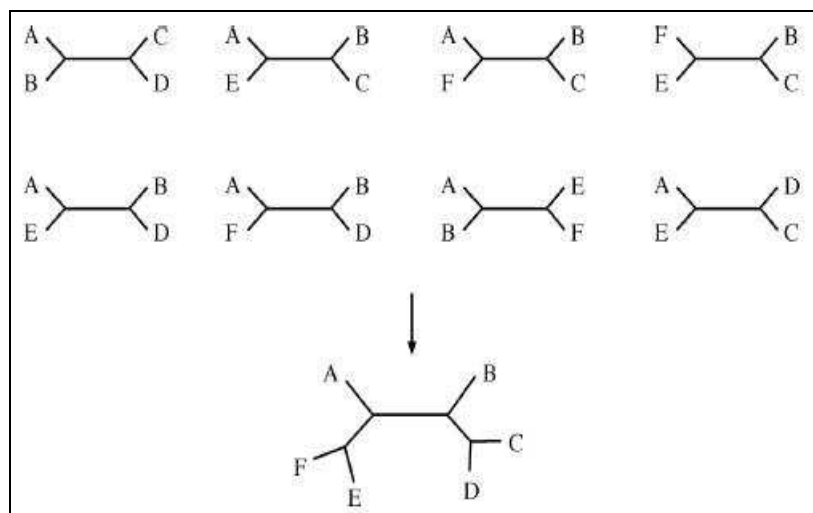
ان درجة لوغاريتم الاحتمال الأعظم لمسار شجرة معينة لتوالي بأكمله هو عبارة عن مجموع لوغاريتمات الاحتمال الأعظم للمواقع المفردة، وتكرر هذه العملية لجميع مخططات الشجرة المحتملة، ثم يتم اختيار الشجرة التي لها أعلى درجة لوغاريتم من بين الأشجار الأخرى كأفضل شجرة وهذه العملية تكون شاملة وتحتاج الى وقت طويل.

ان طريقة الاحتمال الأعظم أسست باستعمال الجوانب الإحصائية وبذا فهي أكثر صرامة من جميع الطرق الأخرى. وتستخدم الطريقة المعلومات الكاملة للتوالي وليس فقط المعلومات الخاصة بالمواقع وهي بالتالي تكون أكثر قوة. ان الطريقة مقاومة لمشكلة الانجذاب الى الفرع الطويل. ونظرا لاعتمادها على نماذج الاستبدال فان الأخيرة قد تكون موقع ضعف للطريقة ان لم يتم اختيار النموذج الأمثل ، اذ انه يمكن ان يؤدي الى إعطاء شجرة غير صحيحة . كما ان الطريقة شاملة بطبيعتها لذا فان زيادة الوحدات التصنيفية يمكن ان يؤدي استحالة استخدامها ولذلك اقترحت عدة طرق بديلة منها الطرق الموجهة مثل :

طريقة الأحجية الرباعية (QP) Quartet puzzling

وهي أكثر الطرق استخداما في مجال تحديد الأشجار بطريقة الاحتمال الأعظم الموجه وفيها يقسم العدد الكلي للوحدات التصنيفية إلى عدة مجاميع كل واحدة بأربع وحدات تصنيفية تعرف بالأرباع ، وتبنى الشجرة من هذه الأرباع وهي عملية بسيطة حيث توجد ثلاث احتمالات ممكنة

لهذه الوحدات الرباعية من الأشجار غير المجذرة . ثم توحد أرباع الأشجار في شجرة ضخمة واحدة تضم جميع الوحدات التصنيفية كما موضح في الشكل 55.5 :



شكل 55.5 : طريقة الأحجية الرباعية لبناء الأشجار التطورية

وبما إنها تشبه لعبة تجميع القطع لذلك اشتق اسمها ، وتبرز مشكلة الاتفاق على رسم معين فقد لا تتفق نماذج التفرعات في الأرباع مع الوحدات التصنيفية المشتركة لذلك تستخدم قاعدة رئيسية لتحديد مواقع التفرعات لغرض إضافتها لبناء شجرة متفق عليها . والسبب في كون طريقة الأحجية الرباعية أسرع حسابيا من طريقة الاحتمال الأعظم الشاملة لأنها تبحث في عدد أقل من مخططات الشجرة . فلو اخذ أربع مجاميع جزئية تصنيفية (C) من n من المجاميع التصنيفية فإنه يوجد عدد كلي بمقدار C_4^n من التوافق . وكل مجموعة جزئية لها ثلاث أشجار محتملة ولذا فإن عدد الأشجار التي بحاجة إلى حساب هي $3C_4^n$ ، فعلى سبيل المثال إن n مساوية إلى 20 وحدة تصنيفية فيكون هناك عدد من المخططات الأشجار المخفية مساويا :

$$3 \times C_4^{20} = 3 \times \frac{20!}{4!(20-4)!} = 14535$$

هذا العدد يمكن البحث فيه باستخدام الطريقة مقارنة بـ 2×10^{20} في حالة استخدام طريقة الاحتمال الأعظم الشاملة . وبذا فإن الطريقة تقلص معنويا الوقت المستغرق للبحث ، ولكنها ليست بقوة طريقة ML ، إلا إنها تستخدم كونها تعمل على تقليل الوقت المستغرق في البحث بشكل كبير وأنه يمكن إن تتعامل مع كمية كبيرة من مجموعات البيانات .

طريقة الخوارزمية الوراثية (GA) Genetic algorithm method

تعد الطريقة إضافة أخرى لطرق الاحتمال الأعظم السريعة وهي طريقة ذات إستراتيجية حسابية مثلى ، وتستخدم المصطلحات الحيوية كقيم مستعارة من العمليات التي تجري لغرض انجاز الطريقة المتضمنة عملية التقاطع الرياضية لتوليد ناتج جديد . وتعمل الخوارزمية على اختيار النتائج الأفضل لعمليات الخلط والتطابق باستخدام عدداً من الحلول العشوائية الموجودة واستخدام المقياس الملائم لمراقبة أمثلة العملية ، وتسجيل قيم المقياس الملائم فأنها تحاكي عمليات الانتخاب الطبيعي وتبادل المعلومات الوراثية . إن القيم الفرعية المسجلة والتي تمتلك أفضل درجة (أفضل عملية ملائمة) يتم اختيارها من الدورة الأولى وتكون بمثابة نقطة بداية للدورة الثانية في حلقة التكرار المثلى . إن استخدام الإشعارات البايولوجية يؤدي إلى إنشاء نتائج التي تكون بمثابة محاولات رياضية مع تحويلات لما سبقها ، ولذلك يسمح بعمليات العبور للكروموسومات لإنتاج حلول جديدة ويتم تكرار العملية إلى حين الوصول إلى الحل .

إن تطبيق الطريقة الوراثية للاستدلال على العلاقة التطورية يعتمد على المقياس الملائم الذي يمثل درجات لوغاريتم الاحتمال الأعظم ، إذ يبدأ البحث في مجموع عشوائي من الأشجار وبأطوال أذرع مختلفة ، لحين الوصول إلى الشجرة التي لها أعلى قيمة من لوغاريتم الاحتمال أعظم.

طريقة Bayesian

طريقة مطورة من طريقة الاحتمال الأعظم ، الأساس فيها الاستدلال على الأشياء غير المشاهدة اعتماداً على قيم معروفة ومشاهدة . أي إنها تعتمد على فكرة مهمة وهي الاحتمالية البعدية Posterior probability والتي تعرف بأنها الاحتمالية التي تحدد بواسطة التوقعات السابقة . وبالصيغة الرياضية تعني احتساب الاحتمالية البعدية من حاصل ضرب حادثتين باستخدام قيم الاحتمالية السابقة والاحتمالية المشروطة بالصيغة الرياضية الآتية :

$$\frac{\text{الاحتمالية المشروطة} \times \text{الاحتمالية السابقة}}{\text{الاحتمالية الكلية}} = \text{الاحتمالية البعدية}$$

$$\frac{\text{Conditional probability} \times \text{Prior probability}}{\text{Total probability}} = \text{Posterior probability}$$

ومن المهم معرفة إن الطريقة يمكن إن تستخدم للاستدلال على أشجار العلاقات التطورية باحتمالية بعدية عظمى . وفي عملية اختيار الشجرة الأفضل فأن الاحتمالية السابقة أو القبلية هي احتمالية لجميع مخططات الأشجار الممكنة قبل إجراء التحليل بطريقة Bayesian والتي تكون متساوية لكل المخططات . أما الاحتمالية الشرطية فهي تمثل تكرار عمليات الاستبدال للثملات الملاحظ من اصطفاغ التواليات . وهذه المعلومات تستخدم للبحث عن الأشجار الأكثر احتمالية والتي تمثل أفضل تمثيل للبيانات الحقيقية .

إن الطريقة يمكن إن تتجز بصورة مماثلة أو أفضل من طريقة الاحتمال الأعظم ولكنها بسرعة أكبر ، ولها القابلية للتعامل مع قواعد البيانات الكبيرة جداً ، وذلك لان طريقة الاحتمال الأعظم تبحث عن أفضل شجرة مفردة في حين إن طريقة Bayesian تبحث عن أفضل مجموعة من الأشجار .

تقيم طرق بناء الأشجار التطورية

من الصعب إيجاد طريقة من طرق إنشاء الأشجار التطورية التي يمكن أن تعكس الحقيقة كاملة ، كذلك فإن الطرق والنماذج المستعملة يمكن أن تهتز مصداقيتها عند تناقل المواد الوراثية بين الأحياء بطرق الانتقال الأفقي . لذلك فعند تفسير النتائج يجب الأخذ بنظر الاعتبار النموذج المستعمل والافتراض الذي تعتمد عليه الطريقة فهناك العديد من الفرضيات التي تكون خاصة بكل طريقة .

كما إن بعض التغيرات تكون غير محسوسة في بعض الطرق كما في حالة الطفرات الراجعة التي تمحي دليل التغير وهذه الحالة عامة بكل الطرق خاصة الطرق المعتمدة على المسافات وحسابها بين التوالين ، وإن كانت هناك بعض الطرق التي تستفيد من هذه الأدلة الخفية الموجودة في بعض الوحدات التصنيفية بمقارنتها مع وحدات أخرى . وهناك عدة طرق أو وسائل يمكن بواسطتها تقييم أداء طرق بناء الأشجار منها :

- إعادة النمذجة وإعادة تحليل صف واحد من المفردات والبحث عن التجمعات التي تظهر بشكل متكرر وبذا تشير إلى مواقع الثقة في بعض الأشجار الخاصة .
- استعمال اختبار Jackknife والذي يهدف إلى إزالة واحدة من التواليات في كل مرة وإعادة التحليل وملاحظة الفرق .
- التقييم المعتمد على البيانات وذلك باستعمال البيانات المميزة Data signals مقابل البيانات العشوائية
- حساب بعض المؤشرات الإحصائية مثل

إجراء فحص الانحياز Skewness test :

يستعمل مع الأشجار المنشأة بطريقة الاقتصاد الأعظم إذ تحسب قيمة g_1 التي تمثل الانحراف الإحصائي، وحساب القيمة الحرجة له التي تختلف مع عدد الوحدات التصنيفية والمواقع المتغيرة في التوالي ولكن الفحص لا يحدد مصداقية بعض المخططات الخاصة ولكنه يكون حساس جدا عند إشارات ضعيفة في مجموعة البيانات ، فإذا كانت الوحدات التصنيفية من المجاميع التي يكون حذفها واضحا" يستعمل الفحص لإيضاح فيما إذا كانت إشارة العلاقة التطورية لازالت موجودة ، و يعطى على الأقل 10 من المتغيرات للمفردات عند فحص 5 من الوحدات .

فحص التبديل Permutation test

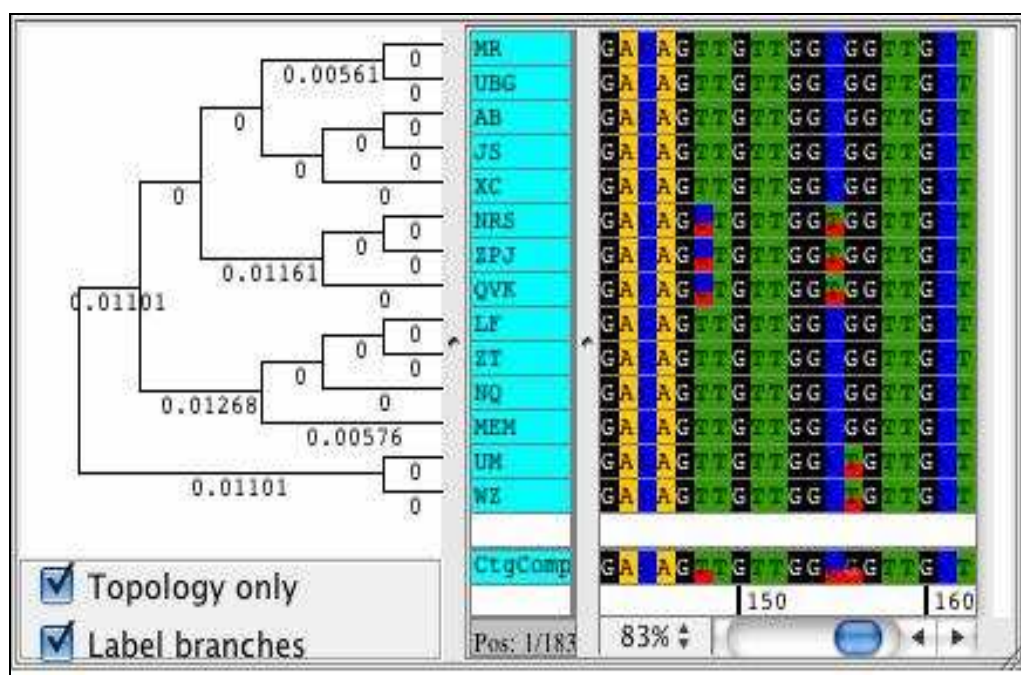
يستعمل الفحص لتحديد فيما إذا كانت الشجرة أو جزء منها نشأ عن طريق الصدفة ، فالبيانات الحقيقية لا توزع عشوائيا ولكن يحصل تبديل في كل عمود من الاصطفافات ولذلك يزال التباين المرافق Covariation بين التواليات الأولية وبالتالي فإن النتيجة لا تكون تواليات عشوائية وهناك فحص احتمالية التبديل Permutation probability test (PTP) يستعمل لإيجاد الإشارة للعلاقات التطورية في البيانات الأصلية . كما إن هناك فحص يعتمد على النمط الطوبوغرافي Topology dependent test (T-PTP) الذي يحدد فيما إذا كانت الاختلافات ناتجة عن الصدفة ومثل هذه الفحوص تدمج مع بعض البرامج المستعملة في إيجاد العلاقات التطورية.

طريقة او فحص Bootstrapping

أسلوب إحصائي يستخدم لفحص واختبار أخطاء المعاينة للشجرة عندما يكون تحديد التوزيع صعبا . وتعتمد الطريقة على إعادة النمذجة لتقييم الأشجار التي تعتمد في إنشاءها على المسافات مثل طريقة الاقتصاد الأعظم او الاحتمال الأعظم لتعطي إشارات تؤكد أن الغصن موحد الأصل وتشمل الطريقة خطوتين ، إذ تقوم بإنشاء مجاميع بيانات جديدة من البيانات الأصلية وإجراء الحسابات التي تعطي نسبة المرات التي يظهر فيها فرع معين في الشجرة وتمثل عدد المرات Bootstrap value

برامج رسم الأشجار

هناك العديد من البرامج المستعملة في رسم الأشجار التطورية وهذه البرامج توفر العديد من الإمكانيات في عمليات الرسم وإمكانية إظهار طول الأفرع أو أشكال مختلفة من الأشجار فضلا عن إظهار القيم الإحصائية ، وكذلك توفر الإمكانيات الفنية للأشجار المرسومة والشكل التالي (56.5) يوضح نموذج في هذه المخرجات



شكل 56.5 : مخرجات احد برامج إظهار العلاقات التطورية بإمكانيات مختلفة

التنبؤ بالتطور المستقبلي

هناك بعض التوجهات في هذا المجال كما هو الحال في التوجهات في مختلف الفعاليات للإنسان . والتوجه المعني هنا هو علم الحياة الصناعي Synthetic biology وهو محاولة بناء جينومات وخلايا صناعية وفي هذا المجال تكون الحاجة إلى تعرية الجينومات الموجودة الآن إلى أقل ما يمكن من التعقيدات المحيطة بها ، ويعتمد هذا التوجه على تحديد تواليات الجينات الأساسية وبالتالي يقوم الحاسوب بتقليل الجينات أي إن الحاسبة هي التي

تقوم باختزال الجينوم وبالتالي دراسة ما يمكن أن ينتج عنها ، أي دراسة حالة التطور المستقبلي . وهذه تكون الأساس في بعض التوجهات المستقبلية في مجالات صحة الإنسان التي سيمر ذكرها لاحقا" .

المعلوماتية الحيوية والجينوم

يتعامل مجال المعلوماتية الحيوية مع عناصر ومركبات الأنظمة الحيوية على أصعدة مختلفة منها ما هو عام مثل التعامل مع الجينوم ومنها ما هو خاص مثل التعامل مع الشفرات أو مواقع التنظيم. وفي هذا الجزء من الكتاب سيتم عرض علاقة المعلوماتية الحيوية مع DNA وما يتبعه. ويرجى ملاحظة ان بعض المعلومات سيتم تكرارها للضرورة القصوى وبالحدود الدنيا للحفاظ على وحدة الموضوع.

اختلافات جينومات حقيقية وبدائية النواة

نظرا لكبر كروموسومات الخلايا حقيقية النواة وامتلأها بالمكررات فان عمليات تحديد تواليات الكروموسوم كاملا تكون صعبة ان لم تكن مستحيلة. والمكررات تكون مشكلة خاصة عند استعمال قطع صغيرة لذلك كان هناك عدد قليل من جينومات حقيقية النواة التي تم تحديد جينوماتها الكاملة، ففي نهاية عام 2007 تم تحديد جينوم أحد الطحالب *Cyanidioschyzon merolae* التي تقوم بعملية التخليق الضوئي فضلا عن ما ذكر في الفصل الأول ، اما الخمائر فعدد بسيط جدا حدد توالياتها اذ لا تزال مناطق اطراف الكروموسومات التي تعد من المكررات الكثيرة غير محددة.

ومقارنة تواليات DNA الكروموسومية يمكن ان يعطي أنماطا" مهمة حتى عند غض النظر عن الجينات المشفرة. منها مقارنة حجم الجينومات والمحتوى العام من AT وغيرها . ومن الجدير بالذكر ان أبسط الجينومات التي يتم التعامل معها هي جينومات الخلايا بدائية النواة والتي يمكن ان تحوي على جينات مثالية وأخرى غير مثالية، فالمثالية هي التي تصل أطوال نواتجها بين 100-500 حامض أميني مع توزيع نيوكلويتيدات ملائم ، اما الجينات غير المثالية فهي اما ان تكون أطول او اقصر وهذا يعني الحاجة الى أكثر من نموذج لبرامج الحاسوب ونماذج إحصائية للوصول الى اقرب ما يمكن من صحة التنبؤ.

دراسات الجينوم المقارنة Comparative genomics

وتشمل شطري الدراسات المقارنة التركيبية والأخرى الوظيفية

الدراسات الجينومية التركيبية Structural genomics

الدراسات تهدف الى مقارنة جينومات أحياء مختلفة وتشمل:

- عدد الجينات

- مواقع الجينات

- محتوى الجينوم من الجينات اي الكثافة التشفيرية

وهذه المقارنات تظهر مدى ثبوت الجينوم المساعد في معرفة الآليات التي تمت وتمت بها عمليات التطور وعمليات انتقال الجينات بين الجينومات وتوضح نمط اكتساب الغريبة أثناء عمليات النقل الأفقي ، وكذلك معرفة الجينوم المركزي القياسي Core genome والتي تكون شائعة بين الجينومات المختلفة وتمثل الجينات التي تكون أساسية لبقاء الكائنات والمهمة في

هندسة عمليات الايض وتركز هذه الدراسات ايضا على تراكيب الجينومات مع التأكيد على الخرائط الجينومية والتجميع والتهميش فضلا عن المقارنات الأخرى . وهذا يعني إجراء عمليات اصطفاف الجينومات ومقارنة ترتيب الجينات في الجينوم ومحاولة بناء الجينوم الأقل Minimal genome . فزيادة وجود تواليات الجينومات المحددة التي أصبحت جاهزة للمقارنات حتم فهم ثبوت التواليات بين الجينومات والتي تظهر ان هناك عناصر وظيفية ثابتة. وعمليات الاصطفاف على مستوى الجينوم الكلي تشبه عمليات الاصطفاف العادي ولكنها أدت الى ظهور بعض التعقيدات. فعمليات الاصطفاف العادية تظهر فيها بعض الأخطاء وتكون غير كفوءة عند مقارنة امتدادات طويلة من DNA تحوي مئات وآلاف الجينات، والمشكلة الأخرى هو إمكانية الإظهار Effective visualization لعمليات الاصطفاف وهذه تحتاج الى إيضاحات بالرسم لتفسير النتائج لذلك احتاجت هذه العمليات الى برامج خاصة لتوضيح عمليات اصطفاف الجينومات الكاملة.

وفي هذا المجال توجد برامج مستعملة لاصطفاف الجينومات الطويلة مثل تلك الخاصة بالقاعدة TIGR التي تقارن اثنين من الجينومات ومقارنة الجينات المتناظرة . وكذلك BLAST الذي يعد المحور فهناك إصدارات في هذا المجال تقوم في خطوة البذار (السابقة الذكر) على إيجاد أطول التطابقات الممكنة التي تشمل ايضا إيجاد حالات عدم التطابق بدلا من إيجاد تطابق K-mer كما هو الحال مع BLAST العادي. وتظهر النتائج على شكل رسم نقطي وخطوط تربط النقاط لتشير الى توافق Collinearity للجينات .

ومن هذه الإصدارات BLASTZ الذي يستعمل كلمات تصل 12-19 نيوكلوไทيد اي أطول من BLAST العادي، وهذه تستعمل في مرحلة البذار ثم إجراء امتداد لعمليات الاصطفافات بالاتجاهين الى حين انخفاض تسجيل النقاط او الدرجات الى الحد الحرج ويحور البرنامج لزيادة حساسيته وخوارزمية BLAT تعمل على اصطفاف التواليات القصيرة بحدود 40 قاعدة نتروجينية المشابهة بدرجة 95% كأساس للعمل وتكون المخرجات رسوم توضح النسبة المئوية لتشابه وتمائل التواليات.

وهناك برامج تميل الى استعمال ما يسمى بالمراسي Anchors والتي تمثل إيجاد تواليات قصيرة متطابقة تماما، والمناطق التي توجد فيها هذه التطابقات المتماثلة تماما تعد المرسى لتبدأ منه عمليات الصف باستعمال طريقة Needlman-Wunsch للاصطفاف الشامل ومثل هذا التوجه يأخذ بنظر الاعتبار ظاهرة اختلاف الشفرات لذا يمكن استعمالها في التعامل مع التواليات البعيدة.

والبرامج العاملة في هذا المجال يمكن ان تتخذ اي وسيلة لغرض الوصول الى أفضل حالات التطابق التي تنتج عنها معرفة العديد من المعلومات. والبعض منها يعمل على بناء قواعد بيانات خاصة به ، كما ان بعضها يتخصص بمقارنة جينومات خاصة من الأحياء مثل تلك المتخصصة بمقارنة جينومات الإنسان والفيضان والجرذان والحشرات وغيرها ، كما ان معطياتها تكون بأشكال مختلفة البعض منها يعتمد الى رسم الأشجار التطورية واستعمالها كأشجار دالة.

مقارنة ترتيب الجينات

من مهام دراسات الجينوم المقارنة الأخرى هو مقارنة ترتيب الجينات. وعندما يكون ترتيب عدداً من الجينات المرتبطة ثابتاً بين عدد من الجينومات يطلق على الحالة التوافق Synteny. وبصورة عامة فإن ترتيب الجينات يكون أقل ثباتاً مقارنة بثبات تواليات الجينات. وتجري مقارنة العلاقات التوافقية Syntenic relationships بين الأحياء المتقاربة نوعاً ما، ويمكن أن تشذ بعض الأحياء بدائية النواة في هذا المضمار حيث توجد علاقات توافقية لترتيب الجينات بين الأحياء البعيدة عن بعضها، وهذا يعطي مؤشراً مهماً للعلاقات الوظيفية للجينات المعنية، فمثلاً الجينات العاملة في مسار معين تميل للتعقد في الأحياء المختلفة.

إن المحافظة على ترتيب الجينات يكون نتيجة للضغط الانتخابي للسماح للجينات بأن تنظم بشكل متزامن وتعمل بشكل أوبرون، وعلى كل حال فإن التوافق بين الجينات في أحياء متباعدة عادة ما يكون مرتبطاً بالتداخل الفيزيائي لنواتج الجينات المشفرة.

وهناك بعض البرامج التي يمكن أن تقارن ترتيب الجينات بين توالين يمكن أن يصل طولها إلى 2 ميكاقاعدة. والبرامج في هذا المجال تظهر نتائجها بشكل رسوم نقطية وتظهر فيها خطوط قطرية تشير إلى توائم الجينات، أما الخطوط خارج القطر الأساسي فتشير إلى إعادة الترتيب للجينات الذي حصل في الجينوم. وتختلف البرامج بطبيعة الحال من حيث السرعة والدقة.

الدراسات الجينومية الوظيفية Functional genomics

يتضمن الحقل دراسة وظائف الجينات على مستوى الجينوم الكلي بالاعتماد على الطرق السريعة High throughput وهذا يعني التحليل المتزامن لكل الجينات في الجينوم. وهذه النقطة أساسية للفصل بين علم الحياة الجزيئي التقليدي الذي يدرس جين واحد في وقت واحد، في حين أن الطرق السريعة التي تسمى أيضاً تحليل مكنون النسخ أي تحليل كل التعبير عن الجينات بالبحث عن كل نسخ RNA التي تنتج في الخلية تحت ظروف محددة وفي هذه الحالة فإن أكثر الجزيئات دراسة هي mRNA.

وفي هذه الدراسات تتضح الكيفية التي تتعاون بها مجموعة من الجينات لعمل المسارات الأيضية ونقل الإشارات داخل الخلية أي أنها تظهر نمط التعبير المتزامن وعمليات التنظيم المرافقة للتعبير الجيني وتحديد وظائف الجينات التي لم تكن موصوفة من قبل.

الطرق والتوجهات المعتمدة على التواليات

هناك عدداً من الطرق لدراسة الجينومات دون الاعتماد على اصطفايف التواليات وإنما تستعمل بعض الواسمات الخاصة Markers ومنها:

استعمال طمغات التواليات المعبر عنها (ESTs)

هي تواليات قصيرة يتم الحصول عليها من cDNA وتستخدم كمعرفات للجينات الكاملة الطول وتكون بطول 200-400 من النيوكلووتيدات وتكون من النهاية 5' أو 3' للـ cDNA . وتحضر بشكل رئيس بواسطة الطرق السريعة ومكتبات cDNA التي تحضر بواسطة النسخ العكسي لجزيئات mRNA باستعمال بوائى قصيرة Oligo(dT) primer التي تنهج مع ذيل الأدينين (A) Poly لجزيئة mRNA ثم ربط قطع cDNAs الى نواقل ملائمة والاحتفاظ في مضاف ملائمة.

ولإيجاد معلومات ESTs فان نساءل في cDNA يتم اختيارها عشوائيا لغرض تحديد التوالي من اي من النهايتين لقطع cDNA . والمعلومات المستخلصة من استعمال ESTs تعطي نتائج محتملة اي غير دقيقة للجينات التي تكون فعالة ويعبر عنها في الجينوم تحت ظروف فسلجية معينة، وذلك لان تردد طمغات معينة منها يشير الى وفرة mRNA الخاصة بها في الخلية والذي يكون ذا علاقة مع التعبير الجيني تحت تلك الظروف والاختيار العشوائي المذكور آنفا يمكن ان يساعد في اكتشاف جينات جديدة.

معوقات استعمال الطمغات ESTs

يوجد عدد من المعوقات لاستعمال هذه الطمغات فبعض الأحيان تكون ذات نوعية غير جيدة نظرا لطريقة إنتاجها وبدون مصادقة وبذا فهي تحوي على معدلات عالية من الأخطاء. وبعض الأحيان يشار الى عدد من القواعد بشكل غامض مثل 'Ns'. ومن الأخطاء الشائعة هو وجود أخطاء في الأطر Frameshift error ووجود شفرات وقف غير حقيقية كما انه بعض الأحيان يحصل تلوث من تواليات الناقل المستعمل ووجود تواليات الانترونات في جزيئات RNA غير المخيطة بشكل صحيح، فضلا عن وجود rRNA او tmRNA ، وغيرها من المعوقات ، وكلها تؤدي الى الفشل في التعبير عن التواليات. وقد وجد ان تواليات ESTs عند النهاية 3' تكون أكثر فائدة وتمثيلا مقارنة بالتواليات المشتقة من النهاية 5' وفي الوقت نفسه تكون التواليات في 3' أكثر إنتاجا للأخطاء وقد وجد ان 11% من نساءل cDNA تكون مختلطة Chimeric.

والمشكلة الأخرى في استعمال ESTs هو حضورها الكبير بالنسبة للجينات التي يعبر عنها بشكل عالي مما يغطي على ما يشترك من الجينات ذات التعبير الواطئ وبذا يكون المسح عن ESTs لمثل هذه الجينات صعبا.

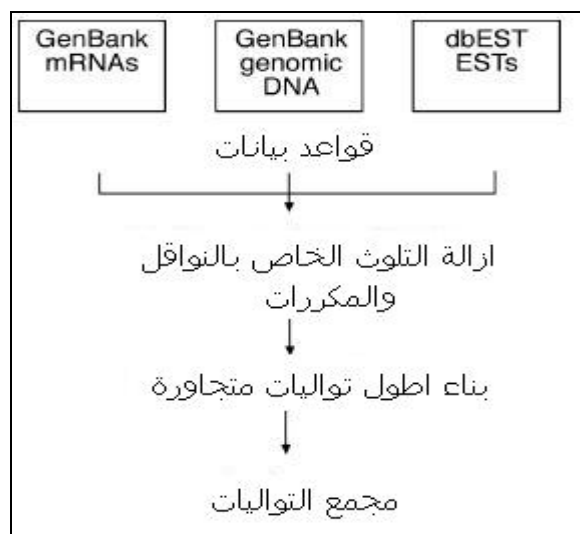
ولكن بالرغم من المعوقات المذكورة آنفا حول ESTs الا انها لا تزال تستعمل بكثرة، نظرا لسهولة تحضير مكتبات منها من مختلف خطوط الخلايا والأنسجة والأعضاء وبمراحل تطويرية مختلفة، كما ان مجمع ESTs اي استعمال عددا منها يفضي الى معلومات مفيدة. والتجمع والتراكم السريع لتواليات الطمغات شجع انشاء قواعد بيانات عامة وخاصة لأرشفتها فالقاعدة البيانية الكبيرة بنك الجينات تحوي على قواعد خاصة EST database تضم مجمعات للـ EST لأكثر من 250 كائن وتحديث وتجدد بشكل منتظم لتعكس التطور الحاصل في دراسة هذه الطمغات المختلفة، وكل EST يودع في قواعد البيانات يتم إخضاعه لعدة اختبارات قبل

الإيداع ومن هذه الاختبارات إيجاد التشابه الشديد لجينات معروفة موجودة في القاعدة وغيرها مما يؤدي إلى إمكانية تهميشه ووضع المعلومات المرافقة له.

ومن أهم أهداف إيجاد قواعد للطمغات هو إيجاد كشف خاص بها EST index وذلك بجمع المعلومات الكثيرة المتوفرة حولها لتحسين نوعية المعلومات وتسهيل الوصول إليها لتستخدم في استخلاص cDNAs الكاملة الطول وتشمل عمليات بناء الكشافات معالجات أولية منها :

- إزالة التلوث من النواقل باستعمال برامج خاصة لمسح تواليات النواقل البكتيرية المستعملة.
- حجب المكررات
- عنقدة ESTs بالجينات الخاصة بها (أي الجينات الحاوية على ESTs معينة)
- اشتقاق التواليات المتوافقة من ESTs المتداخلة والكثيرة وتصحيح الأخطاء الناتجة من إزالة الأطر أو زحف الأطر .
- وكل هذه الخطوات ستؤدي إلى إنتاج متجاورات من EST طويلة وهي طريقة مشابهة لعملية تجميع المقروءات عند استعمال طريقة التشظية لتحديد الجينومات (التي سيأتي ذكرها لاحقاً والمذكور بعضها في موضوع تحديد تواليات الجينومات).
- في الخطوة النهائية تستعمل بعض الخوارزميات لإيجاد المناطق المشفرة.

وبمثل هذه الطريقة سيتم التخلص من الانترونات والمناطق غير المشفرة مثل 3'-UTRs . وعند الحصول على المناطق المشفرة يتم تهميشها بعد ترجمتها إلى بروتينات ومن ثم إيجاد مشابهاتها في قواعد البيانات. وكذلك يمكن أن تستعمل في اصطافاف التواليات الجينومية إذا كانت متوفرة لمعرفة المواقع الجينومية للجينات المعبر عنها، وكذلك تستعمل في معرفة مواقع تلاقي الانترونات والاكسونات باستعمال برامج خاصة بهذه المهمة وهذه الخطوات موضحة في الشكل 59.5 :

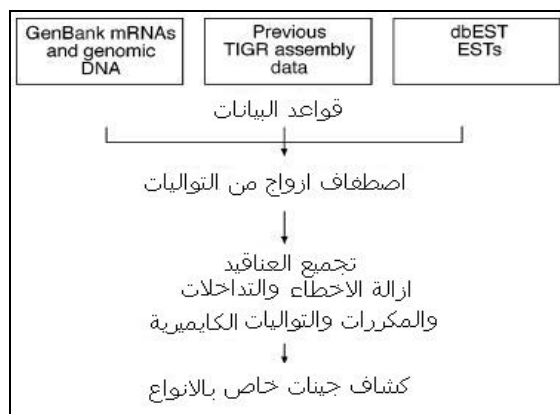


شكل 59.5 : خطوات تحسين البيانات الخاصة بـ ESTs

ان عملية العنقدة المذكورة أعلاه تؤدي الى تقليص عدد ESTs وتنتج بيانات غير مكررة او زائدة تساعد في تهميش وتعريف بالـ EST وهذا بدوره يمكن من إنشاء كشافات خاصة بالجينات Gene index . ومثل هذه الكشافات الجينية موجودة في بعض قواعد البيانات مثل UniGene الموجودة في المركز الأمريكي (NCBI EST cluster database)، وكل عنقود فيها يمثل مجموعة من التواليات ESTs المتداخلة التي عولجت بالحاسوب لتمثل جين واحد معبر عنه تم تهميشه لبيان وظائفه وتحديد نوع الأنسجة التي يوجد فيها في حالة التعبير عنه ، ومن الجدير بالذكر ان قواعد البيانات الخاصة بـ ESTs الموجودة ضمن قواعد بيانات عامة وكبيرة تحوي فقط تلك الخاصة بالنهايات 3' للتقليل من مشكلة الوهمية Chimerism او حالة الالتباس والخلط وذلك لان التواليات ESTs 3' تعطي تمثيل اكبر للنسخ المستعملة. والتواليات المشدبة والمدققة يتم استعمالها للبحث مقابل قواعد البيانات للجينات المعروفة باستعمال BLAST.

كشافات الجينات Gene indices

يمثل جانب آخر من مهام المعلوماتية الحيوية المعتمد على التواليات، وهناك مواقع الكترونية خاصة بالكشافات، وتعتمد بشكل أساس على قواعد بيانات ESTs الموجودة ضمن قواعد بيانات عامة مثل بنك الجينات وتستخدم مختلف أنواع المعلومات والقواعد الخاصة ليتم عنقدة التواليات ، وفي مجال الكشافات تتم العنقدة اذا كان هناك تشابه يصل الى 95% او أكثر من التماثل لمناطق تشمل 40 نيوكلوثيريد في حالة استعمال اصطفايف زوج من التواليات لمنع التوهم او الالتباس وتستخدم فيها أشهر البرامج مثل BLAST و FASTA لمعرفة تداخل التواليات، ومن ثم في مرحلة تجميع التواليات تستخدم برامج تجميع كفوءة لإنتاج متجاورات تخمينية. ثم يتم تحديد وظائف المتجاورات بالاعتماد على البحث المكثف باستعمال BLAST في قواعد بيانات البروتينات المعتمدة، ويحدد معامل خاص ومن معرفة قيم المعاملات الناتجة يمكن الإعلان عن الوظائف ومدى التشابه لنتائج البحث والتحري . وخطوات بناء كشافات الجينات موضحة في الشكل 60.5 :

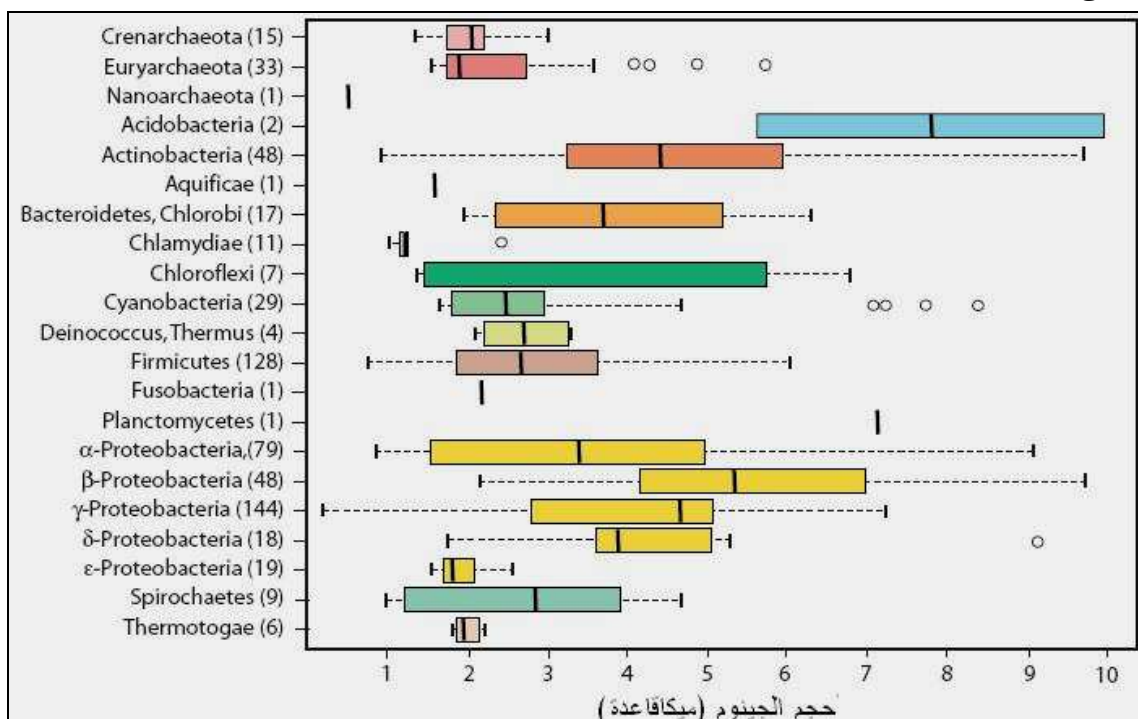


شكل 60.5 : إنشاء كشافات الجينات

حجم الجينومات

من أهم المعلومات الأولية التي يتم الحصول عليها لتحليل الجينوم هو الحجم ومقارنة حجم الجينومات تعد من الخطوات الأولى في ذلك. وتحديد حجم الجينوم أو الكروموسوم ومحتواه من AT يكون بعمليات حسابية بسيطة وفي بعض الأحيان لا تعتمد في تهميش الجينات كما في ملفات القاعدة البيانية واسعة الانتشار بنك الجينات ، وذلك لأن حجم الكروموسوم على وجه الخصوص في الخلايا حقيقية النواة لا يعكس تعقيد الكائن ولا عدد الجينات المشفرة. فمثلا بعض انواع الاميبا وهي كائن أحادي الخلية تحوي على DNA يزيد بمئات المرات عن خلية الإنسان وهذا يحدد بـ C-value والتي لا تشير الى عدد الجينات او غيرها من المعلومات وتمثل القيم الظاهرية C-value paradox وتعكس الظاهرة المذكورة آنفا وهي ان كمية DNA لا تعكس مدى تعقيد الكائن الحي. وهناك عددا من المشاكل مع تحديد حجم الجينومات عند التعامل مع قواعد البيانات التي لا تحدد او تخصص ان الكائن يحوي على البلازميدات خاصة في البكتيريا او كروموسومات متعددة لذلك كان لابد من تصحيح الوضع لكل حالة وذلك بتحديد المشروع Project ID الخاص بكل كائن حي .

ولتحديد حجوم الجينومات يمكن وضع القيم في جداول محددة يمكن منها استخراج المعدل والمدى وقيم الوسيط . ويكون التمثيل بالرسوم هو الأفضل مقارنة بالجدول كما موضح في الشكل 61.5 :



شكل 61.5 : تمثيل مقارنة حجم جينومات أحياء مختلفة باستعمال طريقة رسم الصندوق

Box-and-Whisker

فمن الشكل يمكن استنتاج ان حجم الجينومات يختلف وفق الشعب والأصناف فضلا عن استنتاج الاختلافات ضمن الشعبة الواحدة. ومن أهم الطرق الملائمة للتمثيل الرسمي Box-and-Whisker plot. وعند البحث في قواعد البيانات عن اي جينوم يجب الأخذ بنظر الاعتبار قاعدة البيانات المستعملة وكيفية تحليل بياناتها. وتشير البيانات المستخلصة ان حجوم الجينومات تختلف باختلاف قواعد البيانات لذلك كانت هناك محاولات لتوحيد الجينومات وفرض السيطرة النوعية ، وهذه المهام صعبة ولكنها ضرورية وتبين ان اكبر الجينومات المسجلة بالنسبة لخلايا بدائية النواة في قواعد البيانات تعود الى *Sorangium cellulosum* السلالة So ce56 التي تعود الى مجموعة البكتريا *Myxobacteria* وجينومها 13 ميكا من القواعد النروجينية (13,033,77 قاعدة) ورقم التسجيل لها Acc. No. AM746676 والبكتريا تنتج عدد من المواد ذات الفعاليات الدوائية. والبكتريا الأخرى المسجلة *Burkholderia xenovorans* السلالة LB400 المهمة جدا في تفكيك بعض الملوثات العضية وحجم جينومها يصل الى 9.8 ميكاقاعدة ويمثل ثلاث كروموسومات هي CP0002170, CP0002171, CP0002172. وأصغر الجينومات المسجلة هو للـ *Carsonella ruddii* التي تتطفل على الحشرات وحجم جينومها 160 كيلوقاعدة وتكون بذلك اصغر بثمانين مرة من اكبر جينوم بكتري مسجل ، وجينوم *Pelagibacter ubique* الذي يصل جينومها 1.3 ميكاقاعدة وهو يمثل 10:1 من جينوم *Sorangium cellulosum* المذكورة أعلاه. وجينومات البكتريا المتطفلة هي أثرية وينقصها العديد من الجينات الأساسية لذلك لا تستطيع النمو بشكل حر او خارج خلايا المضيف، وهذا يوضح ما حصل للمايوتكوندريا والكلوروبلاست التي تطفلت في الخلايا حقيقية النواة. وهذه المعلومات توضح انه كلما قلت الجينات في جينوم الكائن أصبح معتمدا على البيئة المحيطة للتزود بالمقومات الأساسية للحياة. والجينومات الصغيرة تؤخذ كقاعدة لبناء شكل مشروع الحياة الصناعي Artificial life form وتضاف اليها الجينات تباعا في محاولة للوصول الى الحالة الحرة من الحياة اي دون الاعتماد على المضاييف .

المكررات في الجينومات

تكثر المكررات Repeats بمعدلات عالية من التردد في جينومات الخلايا حقيقية النواة وهذه يمكن ان تؤدي الى اضطراب بعض عمليات اصطفاف التواليات سواء زوج من التواليات او تواليات متعددة.

ويستعمل الحاسوب لتحديد المكررات طريقة النوافذ المنزلقة Sliding windows لتحليل تكرار توليفة معينة من النيوكلووتيدات اي الكلمات Words وهذه مكونة من اثنين او ثلاثة من النيوكلووتيدات لمقارنة التكرار المتوقع في تواليات DNA . ان توالي القواعد في DNA ليس عشوائيا. وقد تم دراسة تكرار اثنين او ثلاثة من النيوكلووتيدات (الكلمة) بشكل كبير، وهذه الدراسات تعتمد على معرفة كمية A, T, G, C في الجينوم ومنها يمكن حساب عدد المرات التي تتكرر بها اثنين من النيوكلووتيدات مثل AA او GC او غيرها التي يمكن ان تحدث ، وبعد حساب التردد يمكن مقارنتها بالتردد المتوقع. ان تحليل تردد النيوكلووتيدات الثنائية يمكن ان يعطي بعض الملاحظات المهمة مثل الصدفة لوجود اثنين من البيورينات A او G واحدة تلو الأخرى والتي يمكن ان تحسب بسهولة:

فعدد القواعد النتروجينية 4 ← اثنين منها G, A

← اثنين منها C, T

فان وجود اي البيورينات هو 2 من 4 او 0.5

وعليه فان امتداد او خيط مكون من 8 بيورينات او 8 بريمدينات يجب ان تكون هناك 1 لكل 256 نيوكلوتيد بالصدفة (4⁴). بغض النظر عن المحتوى من AT وذلك لان صدفة او احتمالية وجود بيورينات هي لا تزال 1 من 2 وبشكل غير معتمد على جزء AT.

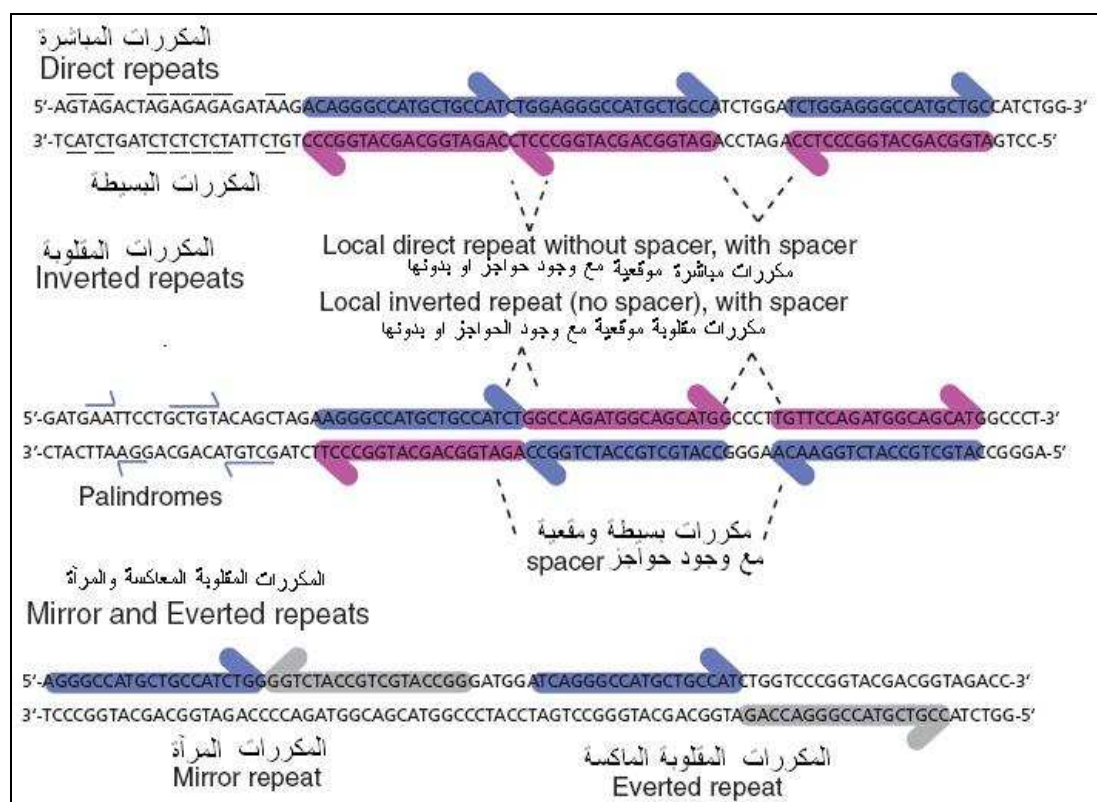
وبالعودة الى الملاحظات الفعلية او المشاهدة وبعيدا عن التوقعات فان بعض كروموسومات البكتيرية وخاصة المحبة للحرارة يلاحظ ان هناك قطع مكونة من البيورينات أكثر من المتوقع ومثل هذه الحالة موجودة في مجموعة البكتيريا Firmicutes مثل *Clostridium tetani* وفي البكتيريا الأخيرة يلاحظ ان هناك انحرافا نحو الادنينات والكوانينات لوجودهم على الشريط نفسه ، لذلك فان إمكانية وجود امتدادات من البيورينات يكون اكبر من حالة التوزيع العشوائي. وهناك جينومات أخرى تظهر ميل نحو الكوانينات في شريط واحد والادنينات في الشريط الآخر والذي يؤدي الى قلة في البيورينات اقل من المتوقع.

وفي الخلايا حقيقية النواة يكون الانحياز الى البيورينات اكبر بكثير والبرامج الخاصة - لمثل هذه الدراسات- تفشل في حدس الترددات بشكل صحيح. وهناك عدة طرق إحصائية لحساب الانحرافات عن المديات المتوقعة وكل طريقة تؤدي غرض بشكل جيد في دراسة التواليات الجينومات ومنها يمكن تحديد النقل الأفقي للجينات الذي يحصل في البكتيريا وعلاقة البلازميدات بجينومات الخلايا المضيفة.

اما تكرار ثلاثة نيوكلوتيدات فيكون ذات أهمية خاصة لانها تمثل شفرة وراثية في التواليات المشفرة ولذلك تكون تحت ضغط انتخابي، اما النيوكلوتيدات الثنائية فهي تمثل جزءا من الشفرة لذا فان ترددها يتأثر بنمط استعمال الشفرات .

وهناك تردد آخر للنيوكلوتيدات الرباعية وهذه تكون مهمة من الناحية البايولوجية لانها يمكن ان تمثل مواقع مهمة للقطع بالإنزيمات ولذلك يتم التغاضي عنها عند وجود الضغط الانتخابي وكذلك بالقطع بالإنزيمات القاطعة الآتية من عناصر DNA المتحركة.

اما المكررات الأطول والمكونة لكلمات معقدة وتتردد بترادف تسمى المكررات المباشرة وتكون بالاتجاه نفسه وعليه فان المكررات المذكورة أعلاه تكون جزءا من المكررات المباشرة وتكون كلمات قصيرة تتردد بكثرة وبعض الأحيان توجد فواصل بين المكررات ويطلق عليها المكررات المباشرة الموضوعية اي انها قريبة من بعضها وعلى الشريط نفسه والاتجاه نفسه، كما موضح في الشكل 62.5 :



شكل 62.5 : أنواع المكررات

والمكررات يمكن ان توجد في عموم الجينوم ومفصولة بامتدادات من DNA قد يصل الى عدة كيلوات من القواعد.

وهناك مجموعة أخرى من المكررات وهي المكررات المقلوبة Inverted repeats وقد توجد بينها فواصل أو بدون فواصل، وتواليات أخرى هي تواليات المراة Mirror repeats والتواليات المقلوبة المعاكسة Everted repeats وهذه تقرأ بالاتجاه غير المألوف في مجال البايولوجي اي تبدأ قراءتها من 3' الى 5' والأهمية البايولوجية لهذه هي لتكوين تراكيب خاصة في كروموسومات خلايا حقيقية النواة وهي غير شائعة في الجينومات البكتيرية.

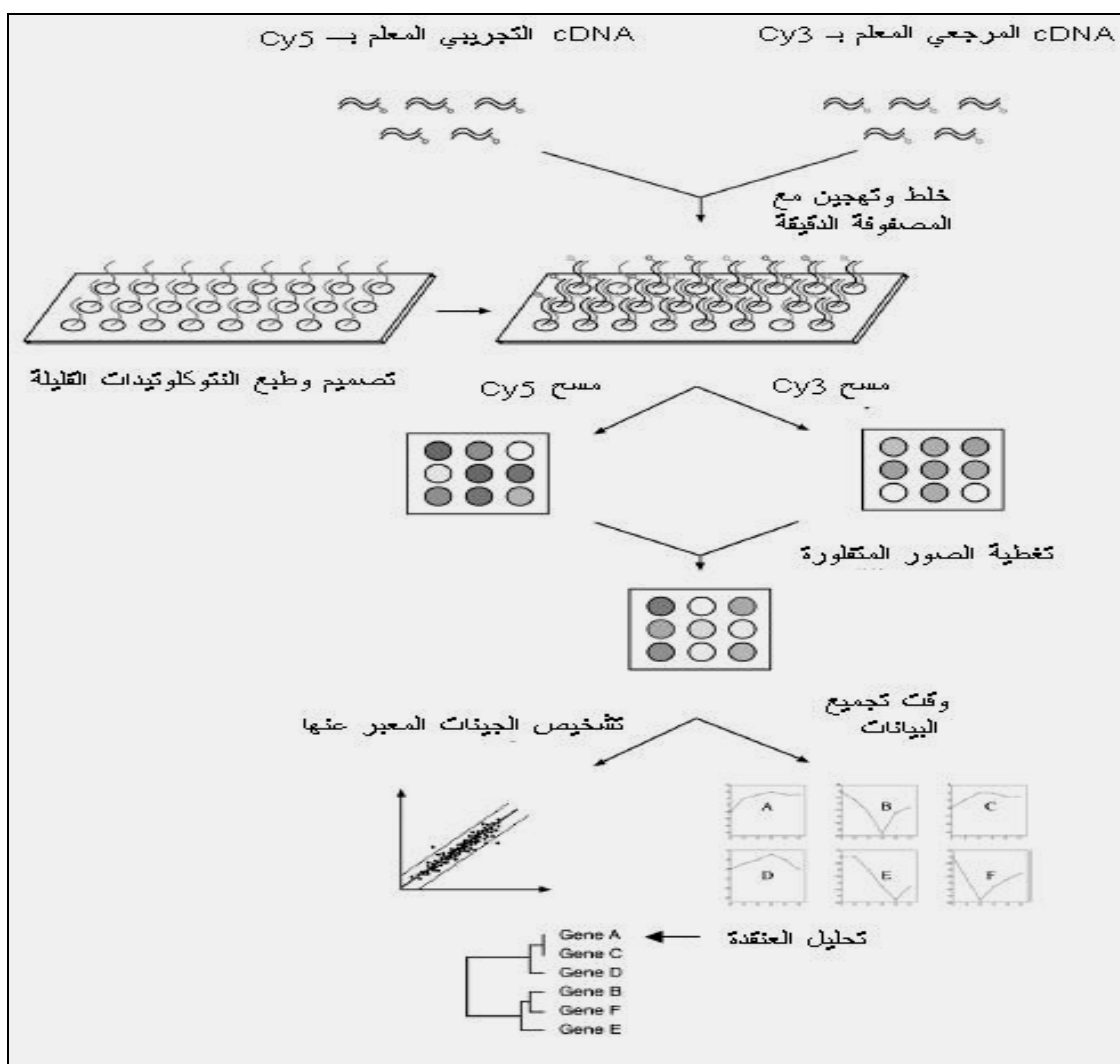
تجميع قطع الجينوم

نظرا لما ذكر في مواقع أخرى من ان الكروموسومات في الخلايا حقيقية النواة تكون كبيرة وملينة بالمكررات فلا يمكن في بعض الأحيان تحديد توالياتها بالطرق العادية وبذا يتم اللجوء الى استعمال طريقة التشظية وفيها تقطع الجينومات أو الكروموسومات الى ملايين القطع الصغيرة التي تصل أطوالها الى مئات من القواعد التي يمكن تحديد توالياتها بكفاءة ودقة، وحتمًا يحصل تداخل بين هذه القطع وتوضع القطع مع بعضها بطريقة تشبه لعبة الأحجار المقطعة Jigsaw puzzle ومثل هذا الوضع غير مقبول لان تواليات DNA بعيدة عن العشوائية والعبثية في الأنظمة الحيوية ، وتحتوي على العديد من المكررات والقطع المتشابهة التي يمكن ان تؤدي الى تجميع خاطئ ، فضلا عن ان عملية تحديد التواليات قد تكون حاوية

على مستوى معين (واطي) من الأخطاء، كما ان بعض أجزاء DNA لا يمكن تحديد توالياتها ، لذا يكون في بعض الأحيان التجميع الكامل للجينوم غير ممكن واغلب المنشور هو مجموعة من المتجاورات التي تمثل قطع الجينوم والتي تعاني من قصور بوجود حالة تضاعف الجينات ووجود المكررات وقد طورت بعض البرامج لتجميع قطع كبيرة من DNA تصل الى 100 ميكاقاعدة. وأكثر الطرق المستعملة لتجميع الجينومات هي الطرق المعتمدة على مصفوفات DNA الدقيقة او رقائق الجين.

طرق تحديد التواليات المعتمدة على مصفوفات DNA الدقيقة

تعد هذه الطرق الأهم المستعملة في الوقت الحاضر للتعرف على التعبير الجيني في الدراسات الجينومية، وتتكون الرقائق المستعملة من سطوح تربط او تقيد عليها قطع من مكوثرات قليلة من DNA وبعض الأحيان تمثل البقع الجينوم الكلي للكائن قيد الدراسة ولكل توالي صغير موقع محدد على الشريحة ليعمل كمجس ليرتبط بشكل دقيق للشريط المكمل له، والمجموع الكلي للـ cDNA المراد البحث عنه يكون معلما بصبغات متفلورة او مواد مشعة اما cDNA المرجعي فيعلم ايضا بصبغات او مواد مشعة وبعد حصول عملية التهجين فان شدة الصبغة او المادة المشعة تعكس كمية mRNA المقابلة في الخلية. وباستعمال طريقة التحليل هذه يتم الحصول على نمط للتعبير الجيني الكلي في الخلية الذي يمكن فحصه، وكذلك يكون بالإمكان تحديد الجينات العاملة في مجال تنظيم المسارات الايضية والأساس موضح في الشكل 63.5 :



شكل 63.5 : تحديد التواليات بالاعتماد على مصفوفات DNA الدقيقة

تصميم النيوكلوتيدات

يتم تحضير رقائق DNA المذكورة أعلاه بثنيت تواليات من النيوكلوتيدات القليلة على مادة سائدة مثل الشرائح الزجاجية باستعمال وسائل ريبونية ، والشرائح تضم الآلاف من الجينات التي يتم اختيارها مسبقا من الكائن الحي، وطول كل قطعة يتراوح بين 25-70 قاعدة وتسمى هذه بالمجسات التي ستتهجن مع cDNA المعلم. وتكون القطع القصيرة متخصصة أكثر في عمليات التهجين لأنها أكثر دقة في تمييز القطع المكمل لها من بين القطع التي لا تتطابق معها. وهناك بعض المؤشرات التي يمكن استخدامها لزيادة كفاءة عملية التهجين وتصميم مصفوفات مثالية منها:

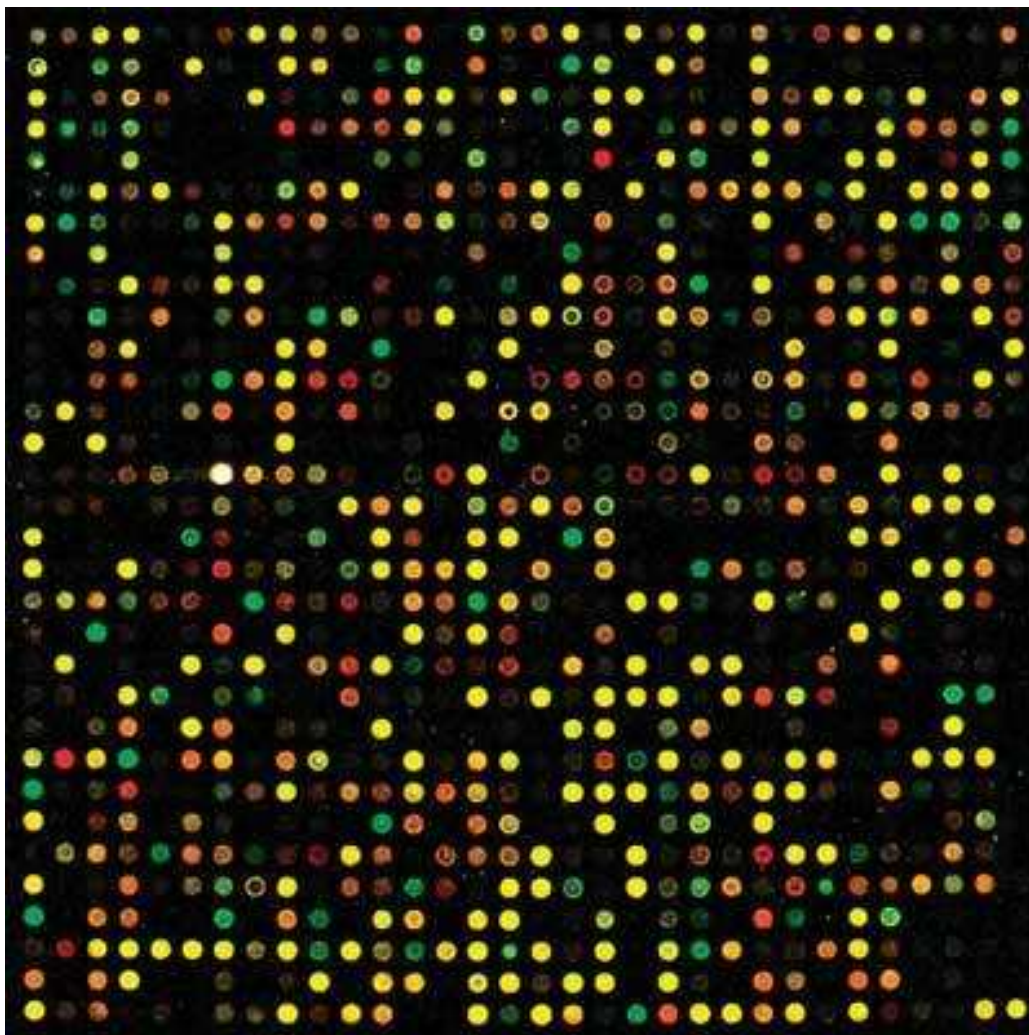
- المجس يجب ان يكون متخصصا" جدا للتقليل من التهجين المتداخل مع الجينات غير المتخصصة وهذا يحتاج البحث في قواعد بيانات للجينومات باستعمال BLAST لإيجاد مناطق من التواليات التي لها اقل ما يمكن من التشابه مع الجينات غير المستهدفة.
- المجسات المستعملة يجب ان تكون حساسة وفيها اقل ما يمكن من المناطق المشوشة اي Low complexity regions لذلك تستعمل معها برامج ترشيح لحجب هذه المناطق عند البحث عنها .
- يجب ان تكون المجسات المنتجة من النوع الذي لا يكون تراكيب ثانوية داخلية مثل تراكيب ماشة الشعر اذ ان هذه المناطق ستتداخل مع عمليات التهجين وفي هذه الحالة يمكن استعمال برامج ملائمة لإيجاد DNA/RNA folding عند البحث .
- يجب ان تكون التواليات المنتخبة كمجسات اقرب الى النهاية 3' من الجين لان اغلب cDNA المجمعة تعتمد على هذه المناطق.
- يفضل ان تكون المجسات لها درجة انصهار Tm متساوية تقريبا ومحتواها من GC بحدود 45-65%.

وقد طور عددا" من البرامج للمساعدة في تحديد تواليات المجسات لاستعمالها في عملية Microarray spotting منها على سبيل المثال لا الحصر الذي يستخدم برامج Java ويعمل مسح أي ويسمح للمستخدم الاتصال بمواقع أخرى لأداء عملية التحليل، وتصمم النيوكليوتيدات القليلة بالاعتماد على عدد من المؤشرات منها التماثل الجذري ودرجة حرارة الانصهار وقلة المناطق المشوشة وغيرها من المؤشرات. وبرامج أخرى تعتمد على خصوصية الجين والتركيب الثانوي للتواليات المستعملة كمجسات ومؤشرات الحرارة الديناميكية.

جمع البيانات

ان نتائج التعبير باستعمال المصفوفات الدقيقة يعتمد عند قياس التعبير الجيني على إشارات التهجين للـ cDNA مع المجسات الخاصة الموجودة في الرقائق الدقيقة. لذلك يتم استخلاص RNA الكلي او mRNA من الأنسجة او الخلايا وخلطها مع الصبغات المتفلورة أثناء عمليات تخليق cDNA. وأكثر الطرق المستعملة تستعمل رقائق بلونين وتعلم cDNA التي يتم الحصول عليها من التجارب باللون الأحمر Cy5 والأخرى المرجعية او السيطرة تعلم باللون الأخضر Cy3 وعند خلط هذه الجزيئات (بكميات متساوية) وتركها لعمليات التهجين مع مجسات DNA على الرقائق فان نمط التعبير الجيني للنماذج يمكن ان يقاس في الوقت نفسه. ويتم التقاط الصور باستعمال الماسح الليزري الذي يمسخ ويمر على كل بقعة في المصفوفة وتستعمل أطوال موجية معينة لتهييج الصبغات المتفلورة وتسجل على انها إشارات حمراء او خضراء والتي تشير الى التعبير النسبي لكل جين ولذا فالصورة تعكس شدة الألوان. فاذا كان الجين معبر عنه بمستويات عالية تحت الظروف التجريبية (احمر) مقارنة بجين السيطرة (اخضر) فان البقعة ستظهر اللون الأحمر، وفي حالة إعطاء الجين مستويات اقل من التعبير مقارنة بالسيطرة فتظهر البقع خضراء . اما الجينات غير المتغيرة من حيث التعبير فيكون تفلور صبغاتها متساوي من الأخضر والأحمر مؤدية الى ظهور اللون الأصفر كما موضح في الشكل

: 64.5



شكل 64.5 : نتائج الصبغات المتفلورة لإظهار مستويات التعبير الجيني

وتخزن الصور في الحاسوب لإجراء الدراسات والمعالجات الأخرى.
معالجة الصور

تتم معالجة الصور لتحديد موقع وكمية بقع التهجين ولفصل إشارات التهجين الحقيقي من تشوش الخلفية والبقع الكاذبة التي تنتج من التهجين غير المتخصص وكذلك عدم الانسجام ووجود التلوث ووجود تغاير في إبعاد بقع التهجين التي تجعلها ذات أشكال غير منتظمة. ويستعمل الحاسوب لتصحيح مواقع وحدود البقع وقياس شدة إضاءة صور البقع بعد طرح قيم الخلفية Background pixels، وبعد عملية الطرح فالإشارات تحول إلى أرقام وتكتب على شكل نسب بين الصبغات Cy3 و Cy5 لكل بقعة وهذه النسب تمثل التغيرات النسبية للتعبير وتعكس عدد مرات التغيير في كميات mRNA في التجربة مقابل معاملة السيطرة. والمعلومات بعض الأحيان تمثل كألوان كاذبة بشدة إضاءة مختلفة لكل من الأحمر والأخضر فيما إذا كانت النسبة

أكبر أو أقل من 1 على التوالي . ويوجد عدد من البرامج الخاصة بتحليل الصور وتساعد في تسوية عمليات التحليل وإظهار التوزيع الطبيعي للجينات المعبر عنها.

التحليل الإحصائي لتحديد الجينات المعبر عنها

لتحديد الجينات المعبر عنها فإن معظم الدراسات تستعمل قيم حد الضعف الفاصل 2fold cutoff وهذا الحد الفاصل هو كافي والذي يمكن ان يعد عالي جدا او واطئ جدا اعتمادا على تغيرات البيانات فضلا عن استيراث التغيرات في البيانات الذي لا يؤخذ بنظر الاعتبار، وقيمة النقطة التي تكون أعلى أو أوطأ من الحد الفاصل قد تكون صدفة أو ناتجة عن خطأ، ولتأكيد ذلك لابد من إعادة التجربة أو إجراء عدة تجارب لمعرفة الجينات المعبر عنها ثم فحصها إحصائيا. فالتجارب المكررة تعطي مكررات لنقاط البيانات التي توفر معلومات حول التغيرات في بيانات التعبير عند ظروف محددة، ولكن أكثر معوقات الحصول على مكررات من مجموعة من البيانات هي الكلفة وذلك لان تجارب المصفوفات الدقيقة تكون مكلفة جدا للمختبرات العادية.

وعلى العموم وفي حالة توفر مجاميع البيانات المكررة فإن هناك فحوص إحصائية جيدة يمكن إجراءها مثل فحص $t\text{-test}$ و ANOVA للإشارة الى ان قيم النقاط او قيم البيانات لا تختلف عن معدل توزيع البيانات، وفي هذه الفحوص تستخدم قيم $P\text{ value}$ وكحد فاصل 0.05 والتي تمثل قيم ثقة 95% Confidence value لتمييز مجاميع البيانات. والملاحظ ان الانحرافات أقياسية تكون ذات معنى اذا كانت البيانات موزعة بشكل طبيعي والذي يجعل عملية التطبيع أو التعديل (السابقة التي تجري على البيانات) أكثر قيمة.

وهناك بعض البرامج الإحصائية التي تستعمل بيانات المصفوفات الدقيقة باستعمال أسس إحصائية مختلفة للوقوف على مدى التغيرات ومقاييس التعبير الجيني لتحديد الجينات التي يعبر عنها بشكل تفاضلي وبعضها توفر إمكانية تحديد النتائج الموجبة الكاذبة والسالبة الكاذبة في البيانات التجريبية.

ويمكن تصنيف بيانات المصفوفات الدقيقة في مصفوفتين تمثل الصفوف Rows فيها الجينات والأعمدة تمثل ظروف أخذ النماذج التجريبية مثل وقت أخذ النماذج أو تركيز الأدوية وكل نقطة في المصفوفة تمثل نسبة شدة Cy5/Cy3 والتي تشير الى مدى التعبير عن الجين تحت الظروف الخاصة المطبقة، ثم تستعمل وسائل مختلفة لتصنيف قيم المصفوفات لغرض مقارنة التعبير الجيني ومنها برامج العنقدة بمختلف أشكالها مثل العنقدة الهرمية. وفي هذا المجال يمكن استعمال أكثر من برنامج البعض منها لتحليل الصور وآخر لتعديل البيانات وأخرى لتحليل العناقيد وإظهارها وأخرى لإدارة البيانات، وحزمة البرامج هذه توفر مخططات كثيرة للتعديل وخيارات العنقدة.

التحليل المتتالي للتعبير الجيني (SAGE) Serial analysis of gene expression

إحدى الطرق السريعة لتحديد أنماط التعبير الجيني الشاملة، وتستند الطريقة الى تقدير كميات نسخ (mRNA) باستعمال كميات كبيرة من cDNA ذات القطع الصغيرة التي تصل أطوالها الى 15 قاعدة تمثل كل الجينوم وهي تكون أفضل في الكشف عن الجينات ذات التعبير الضعيف مقارنة بالطرق المعتمدة على EST. ويطلق على هذه القطع الصغيرة الطمغات Tags وحجمها بحدود 15 قاعدة التي تجمع مع بعضها ويتم تحديد تواليها ثم يجري تحليل النسخ بالحاسوب

لمعرفة انتشارها في الجينومات وترددتها يشير الى مدى التعبير الجيني، فعند وجود قطع كبيرة تصل الى 700 قاعدة فانها يمكن ان تحوي على حوالي 45 طمغة والذي يعني ان الطريقة أكثر كفاءة حوالي 45 مرة من طريقة استعمال EST.

وعملها هي تشابه طريقة مصفوفات DNA الدقيقة وان نتائجهم تكون متناغمة الى حد ما ولكن هناك بعض الاختلافات الواضحة بينهما منها:

- طريقة التحليل المتتالي للتعبير الجيني لا تحتاج الى معلومات أولية عن تواليات النسخ في حين ان تجارب الرقائق تحدد الجينات التي توضع على البقع في الرقائق.
- وبما ان التحليل المتتالي للتعبير الجيني يستطيع ان يقيس كل mRNA المعبر عنه في النموذج فهي تمتلك قابلية للكشف عن الجينات الجديدة.
- قياس التحليل المتتالي للتعبير الجيني لمستويات mRNA هي مطلقة ولا تحتاج الى قياسات مرجعية لذلك تكون نتائجها أكثر قابلية لإظهار المقارنات عند الظروف المختلفة.
- عند خطوة التضخيم بالـ PCR للتحضير لعملية التحليل المتتالي للتعبير الجيني تحتاج الى كميات صغيرة فقط من نماذج mRNA بينما تكون الحاجة كبيرة من mRNA لإجراء تجارب الرقائق الدقيقة.
- جمع مكتبات التحليل المتتالي للتعبير الجيني تكون صعبة وتحتاج الى جهد وتكون مكلفة مقارنة بطريقة الرقائق.
- تحديد الجينات من بيانات التحليل المتتالي للتعبير الجيني تكون أكثر صعوبة من طرق الرقائق وذلك للحاجة الى استخلاص رقع mRNA ثم تجميعها والتعرف عليها بالحاسوب.
- طريقة التحليل المتتالي للتعبير الجيني غير ملائمة لعمليات المسح السريعة للخلايا في حين تكون عملية المسح عند استعمال رقائق DNA ملائمة .

ومن الملاحظ في طريقة الرقائق الدقيقة تكون هوية المجسات معروفة أصلا ولكن في حالة التحليل المتتالي للتعبير الجيني تكون الحالة ليست مشابهة. وقد طورت برامج كثيرة تلاؤم طريقة الرقائق التي أصبحت في الوقت الحاضر تتم بشكل آلي . ومع هذا فقد أنشأت قواعد بيانات خاصة بالتحليل المتتالي للتعبير الجيني لها قابلية اكبر لتوفير البيانات في مقارنة البيانات بين الظروف القياسية والظروف التجريبية.

ومما تبين أعلاه حول استعمال ESTs والتحليل المتتالي للتعبير الجيني او مصفوفات DNA الدقيقة انها طرق تتعامل مع مكنون النسخ اي مع جزيئات RNA المستنسخة من DNA والتي تكون صلب موضوع الدراسات الجينومية الوظيفية، ولكنها ليست لها علاقة بالتعبير الجيني على مستوى البروتينات التي بالتالي تمثل النمط المظهري، وهذا يواكب الحقيقة التي تشير الى ان ليست كل الجزيئات المنتسخة من DNA ممكن ان تترجم الى بروتينات، لذلك فان الدراسات الجينومية الوظيفية لا تعطي صورة كاملة للعمليات الحيوية وربما أعطت ما يقابل النصف او اقل.

التنبؤ وإيجاد الجينات والممهدات Gene and promoter finding and prediction

يعد إيجاد الجين وتهميشه من المهام الرئيسة في المعلوماتية الحيوية عند تسجيل تواليات جديدة. ونظرا لأهمية الموضوع وزيادة المعلومات حول توالي الجينومات فقد وضعت برامج عديدة وأخرى طورت من برامج قديمة لتعطي إصدارات جديدة تطوع للاستخدام مع أحياء خاصة، ولا يوجد برنامج واحد يعطي كل متطلبات العمل مع الجينات المختلفة لأنه لا يوجد نوع واحد من الجينات فضلا عن أن الجينات تختلف في مستويات التعبير عنها وكمية البروتينات الناتجة وكذلك تختلف في موقع التعبير عنها داخل الخلايا، وهذا يعني عدم وجود نموذج جيني موحد. وهذه تؤخذ بنظر الاعتبار عند بناء قواعد بيانات خاصة التي يجب أن تكون بحجم كافية للأغراض الإحصائية الملائمة للأنظمة الحيوية. وتكون عملية التنبؤ أو إيجاد الجينات بالبحث في قواعد البيانات العادية العالمية أو ببناء قواعد البيانات التي ينشأها المستخدم بعد أن يطوعها بالتدريب أفضل عند استعمال عدة قواعد من استعمال واحدة.

الأغراض من إيجاد والتنبؤ بالجينات

الأغراض متعددة منها الحاجة إلى تهميش وتوضيح الجينات فهي في الأحياء حقيقية النواة لا تكون بشكل توالي مستمر وإنما تكون مقسمة إلى أكسونات وانترونات . وقبل وضع التواليات الجديدة في قواعد البيانات يجب أن تحلل لإيجاد المظاهر الحيوية وتزويد قاعدة البيانات بالتعليقات المفيدة حول التوالي. وهذا يحتاج إلى حدس أو تنبؤ Prediction للتوالي الجديد وتحديد وظائفه والتصديق العملي له وذلك بمقارنته بما موجود من التواليات في قواعد البيانات الموجودة وتحديد أطر القراءة المفتوحة والبروتينات التي ينتجها ووظائف البروتينات الناتجة بعد البحث عن الدومينات والقطيقات فضلا عن الاعتماد على البحوث المنشورة. وتطوير طرق حاسوب كفوءة في حدس الجينات يقلص بشكل كبير حجم العمل التجريبي . ونظرا لزيادة المعلومات حول تواليات الجينومات بشكل تزايد كانت الحاجة ماسة لتطوير طرق أوتوماتيكية فضلا عن كونها طرق موجهة .

الطرق المستعملة للتنبؤ بالجينات

تختلف الطرق المستعملة للتنبؤ بالجينات وكذلك الممهدات وهي تقسم أساسا إلى:

- 1- الطرق الخارجية Extrinsic methods وهي التي تعتمد على استعمال التماثل الجذري لتوالي الاستعلام مع جينات معروفة وهناك حوالي 50% من الجينات يمكن إيجادها بهذه الطريقة وتسمى هذه الطرق غير المباشرة.
- 2- الطرق الداخلية Intrinsic methods وهي تمكن من إيجاد 50% من الجينات بالاعتماد على طبيعة تركيب التوالي. وهي من الطرق المباشرة إذ يتم البحث فيها عن ESTs أو cDNA أو البروتينات في الكائن الحي نفسه أو الأحياء القريبة منه .
- 3- الطرق الهجينة وهي التي تستعمل الطريقتين المذكورة أعلاه.

المؤشرات المستعملة في طرق التنبؤ

تختلف المؤشرات المستعملة لإيجاد الجينات بين نوعي الخلايا الرئيسة وهي البدائية النواة والأخرى حقيقية النواة. والمعروف أن المناطق المشفرة للبروتينات هي التي تترجم يكون عدد قواعد قابلة للقسم على 3 (أي عدد القواعد في الشفرة الواحدة) وكل منها سيعطي حامض أميني معين في البروتين الناتج. في الخلايا بدائية النواة تكون الجينات متواصلة فيكون من

السهل البحث عن أطر القراءة المفتوحة ORFs ويتم البحث عن شفرات البدء وشفرات الوقف اما في الخلايا حقيقية النواة فتكون المناطق المشفرة قصيرة وغير متواصلة والبحث عن شفرة وقف يكون غير مجديا.

ومن المؤشرات الأخرى التي تستعمل لتوصيف المناطق المشفرة هو استعمال مكونات المنطقة من القواعد خاصة G+C ، فالانترونات تكون غنية بـ A/T أكثر مما هو عليه الحال مع الاكسونات خاصة في النباتات. لذلك يتم البحث عن مكونات الشفرات وتكرارها Hexamer frequency واستعمال المكوثر السداسي (اي استعمال كلمات للبحث بطول 6 نيوكليوتيدات) هي طريقة جيدة للتمييز بين المناطق المشفرة وغير المشفرة وقد استغلت هذه في وضع العديد من البرامج باستعمال طرق مختلفة.

الطرق الخارجية

التوجه المعتمد على المعلومات الخارجية اي باستعمال عمليات التشابه والتماثل لإيجاد الجينات والممهدات وتسمى ايضا بطرق تقفي الأثر التطوري Phylogenetic footprinting وهي طرق معقدة وتستعمل أكثر من خوارزمية للاصطفاف وتعتمد أساسا على خوارزمية Smith-Waterman التي تعتمد على فتح الفجوات او غلقها وتكون بدون جزاءات ويمكن وصفها بانها برامج اصطفاف متخصصة والبرامج المستعملة في الطرق الخارجية يمكن ان تصنف الى أصناف تعتمد على نوعية الجزيئات او التواليات المستعملة في الاصطفاف مثل:

- DNA الجينومي / بروتين
- DNA الجينومي / cDNA
- DNA الجينومي / DNA الجينومي

وبعض البرامج يمكن ان تتعامل مع أكثر من واحد من الاصطفافات المذكورة مع الأخذ بنظر الاعتبار إزاحة الأطر في تواليات DNA الجينومي او تواليات cDNA. وفي حالة استعمال تواليات DNA مع البروتينات فتأخذ البرامج كل الاحتمالات لتجميع الاكسونات (اي تأخذ بنظر الاعتبار الخياطة البديلة وما يمكن ان ينتج عنها)، ثم تصفها مع البروتين المستهدف باستعمال المصفوفات الخاصة بالبروتينات مثل PAM بمختلف درجاتها لتحديد حالة عدم التطابق.

فمثلا اذا كانت ترجمة تواليات DNA مشابهة لبروتين معروف او عائلة بروتينية معروفة في قواعد البيانات فان هذا سيكون دليلا "قويا" على ان المنطقة المشفرة للبروتين متشابهة ، وبعض هذه البرامج تبني قواميس للكلمات من البروتينات الخاصة بها وتستعمل معها قواعد بيانات EST ثم تستعملها لإيجاد قطع من توالي الاستعلام المتشابه مع القواميس التي أنشأتها. وتكون القواميس المعدة معتمدة على الطريقة وتلحق بهذه القواميس الحسابات الإحصائية لتحديد صلاحيتها.

وهناك برامج تستعمل تواليات DNA جينومي (توالي الاستعلام) وتقارنها مع قواعد البيانات للـ cDNA وهذه تهدف الى تحديد الاكسونات بغض النظر عن حالة تشفيرها خاصة اذا كان مصدري DNA من الكائن نفسه. فاذا كانت الاكسونات في منطقة DNA الجينومي تطابق منطقة من cDNA فهذا يوفر الدليل التجريبي على وجود المنطقة المشفرة. وفي هذه الحالة يمكن ان تنشأ بعض المصاعب من وجود المناطق غير المشفرة UTRs كجزء من الجين

وبالتالي تؤثر هذه في تحديد مواقع بدء الترجمة وشفرات الوقف. والبرامج المستعملة لإيجاد التشابه تختلف في توجهاتها عند استعمال مؤشرات تركيب الجين فالبعض يعتمد على ESTs والتي تؤدي الى عدة صعوبات لان ESTs كثيرة وتؤثر في بعض البرامج المعتمدة على العقدة الأولية ، فضلا عن انها توجد في بعض المناطق غير المشفرة، كما ان العديد منها ينتج بصورة خاصة ، ولكن استعمال خوارزمية Smith-Waterman المذكورة أعلاه يمكن ان تستوعب هذه الصعوبات . والجانب الآخر الذي يخص ESTs هي ان اغلبها يشتق من ESTs-3' لذلك تكون ملائمة للكشف عن المناطق التي هي في النهاية 3' وتعطي نتائج موثقة فقط لهذه النهاية (3') ، لذا لا تساعد في تحديد تركيب الجين كاملا. كما ان بعض البرامج التي تقع تحت واجهة التشابه اي التوجه الخارجي تتخصص بنوع واحد من الكائنات باعتماد فرضية ان مناطق الانترون — اكسون تكون ثابتة في النوع الواحد.

ومن الجدير بالذكر ان بعض البرامج المستعملة في هذا التوجه تستعمل الاصطفاف الشامل وهذه البرامج أظهرت ان نصف الجينات التي تم التحري عنها كانت لها متشابهات مع الجينات الموجودة في قواعد البيانات العامة .

الطرق الداخلية

تعني الطرق المعتمدة على مواصفات الجين نفسه ومن أهمها طرق Ab initio [ولغويا تعني Ab initio من البداية ، لذلك سيستعمل مصطلح طرق البداية للسهولة] التي يمكن ان تستعمل لوحدها او تربط مع الطرق الخارجية المعتمدة على التماثل وتسمى بعض الأحيان طرق الكرة الجديدة de novo prediction methods والتي تعمل مسح على التوالي المعطى. وتستعمل الطريقة تركيب الجين كقالب للكشف عنه وتعتمد على نوعين من المعلومات حول التواليات المكونة للجين هي :

• صفات متعلقة بالإشارات Signal sensors او Gene signals

• صفات متعلقة بمحتوى الجين Content sensors

والمجموعة الأولى تمثل وجود القطيفات من التواليات القصيرة وغيرها من المناطق الفعالة في الجين مثل مواقع الانفلاق او نقاط التفرع او وجود مكررات البريميدونات وكذلك وجود شفرات البدء او الوقف والانحياز في مكونات المادة الوراثية من القواعد والانحياز في استعمال الشفرات. وكذلك استعمال مواقع ارتباط عوامل الانتساخ والرايبوزومات وتركيب الشفرة الثلاثي الذي يحدد الإطار المشفر وتكون من مضاعفات 3 وغيرها من الإشارات. وتستعمل مع هذه الطرق البرمجة الديناميكية لزيادة دقتها والتقليل من الأعداد الهائلة التي يمكن ان تنتج لذلك تكون محتويات الجين هي الأساس وتكون بمثابة وصف إحصائي للمناطق عادة المشفرة منها. وتستعمل برامج الطريقة في التحري عن الاكسونات كما سيأتي ذكره لاحقا. وتعترى هذا التوجه من الطرق بعض حالات عدم الدقة والتي يمكن ان تحدد من المؤشرات الإحصائية ، وعدم الدقة يعود الى عدة أسباب منها:

• عدم انضباط تحديد المناطق المشفرة (في حالة الخلايا حقيقية النواة) بشكل دقيق جدا

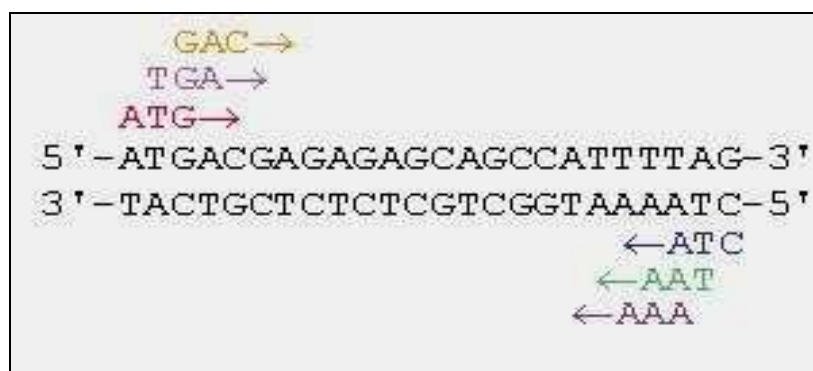
• اعتماد الطرق على فلق الجين الى عدة قطع ثم تجميعها مرة أخرى.

اما المجموعة الثانية اي قياس المحتوى وهي تقسم المناطق الى مناطق مشفرة مقابل مناطق غير مشفرة وتستنثر التشابه الكافي بين تواليات الجينوم او البروتين مع ما يقابلها في قواعد البيانات لتحديد فيما اذا كانت تواليات الاستعلام متشابهة ام لا ، وبطبيعة الحال فان هذه تعتمد الى استعمال الاصطافات الموضعية كما في استعمال خوارزمية Smith-Waterman التي تعد مثالية لهذه الطرق. وتصلح الطريقة لأنواع التواليات البايولوجية الثلاثة ، تواليات DNA او تواليات RNA او تواليات البروتينات ، وأكثر التواليات المستعملة هي البروتينات لتقارن مع تواليات البروتينات في قواعد البيانات العامة مثل قاعدة البروتينات السويسرية و PIR وهذه القواعد تغطي 50% من الجينات. ولكن مع هذا وحتى عند وجود تطابق جيد فان التشخيص الكامل لتركيب الجين يبقى صعبا لان البروتينات المتشابهة يمكن ان لا تشترك في كل دوميناتها كما ان مشكلة UTRs لا تزال قائمة.

اما التواليات الأخرى المستعملة ضمن نطاق قياسات المحتوى فهي النسخ التي يحدد تواليها بعد إنتاج cDNA بعد نسخها من mRNA وتستعمل كتواليات طويلة (قدر الإمكان) لتغطي الجين بأجمعه. او تستعمل ESTs التي يمكن جمعها من مكتبات cDNA libraries.

تحديد مكنون قراءة الأطر المفتوحة ORFome

يمثل المصطلح ORFome كل أطر القراءة المفتوحة ويشمل كل القطع القابلة للترجمة في الجينوم. وتحديد يمكن ان يلحق بالطرق الداخلية المعتمدة على تركيب الجين نفسه، فإطار القراءة المفتوح (والذي يقع بطبيعة الحال ضمن المناطق المشفرة) يشمل توالي من الشفرات التي تترجم الى حوامض امينية محددة في البروتين الناتج. ويبدأ الإطار بشفرة ATG (شفرة البدء) في اغلب الأحياء وينتهي بإحدى شفرات الوقف TGA، TAG، TAA ولذلك كانت إحدى الطرق لإيجاد الجينات هو البحث عن هذه الشفرات وتكون هذه مهمة في الأحياء بدائية النواة وتتعدد في الأحياء حقيقية النواة. ولغرض مسح قطعة من DNA للبحث عن أطر القراءة المفتوحة يجب ان تتم عملية المسح 6 مرات وذلك لان توالي DNA يمتلك 6 أطر للقراءة، ثلاثة منها في اتجاه وثلاثة في الاتجاه المعاكس للأشرطة المزدوجة المتكاملة وتقرأ بالاتجاه من 5' الى 3' وباعتبار ان الجينات يمكن ان توجد في اي من الشريطين كما موضح في الشكل 65.5 :



شكل 65.5 : اطر القراءة المتعكسة الاتجاه

ومن العودة الى مسألة الشفرة المكونة من ثلاث قواعد ننتروجينية فهناك ستة إمكانيات لقراءة الشفرات وان (المفروض) كل الجينات تبدأ بشفرة الميثانونين ATG ولكن ليس كل ATG تعني موقع بدأ ، ولكن يكاد يكون هناك اتفاق حول شفرات الوقف الثلاثة وهذا يعني احتماليات شفرة الوقف هي 20:1.

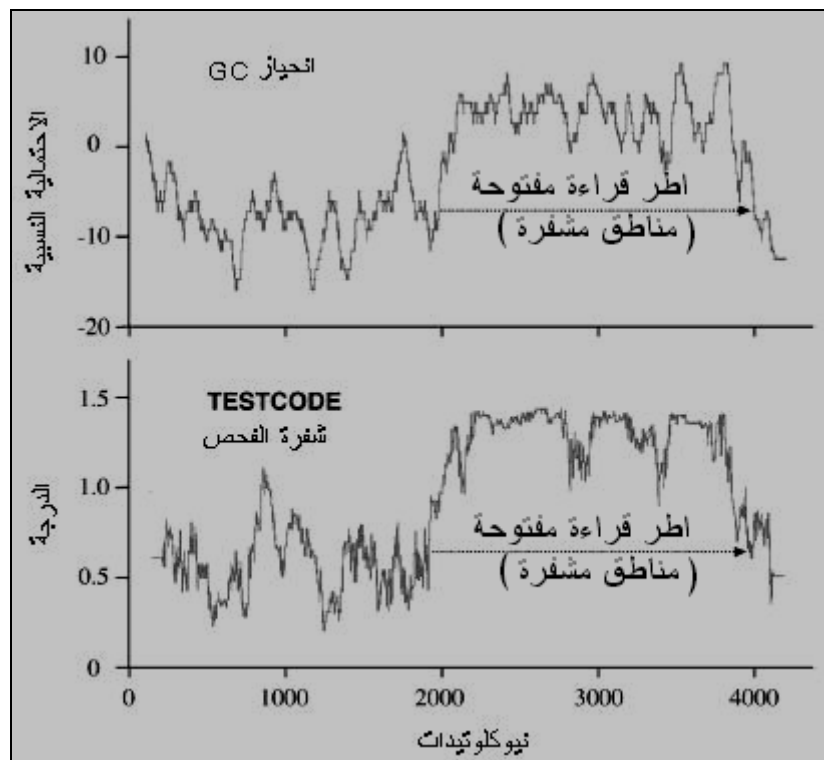
ومن الجدير بالذكر ان نقطة البدء ونقطة الانتهاء لإطار القراءة المفتوحة لا تكافئ نهايات mRNA المنتسخ منها وانما يكون ضمن جزيئة mRNA ، فضلا عن وجود تغيرات في شفرات البدء والوقف في أطر القراءة بين الكائنات. وتعتمد البرامج التي تتخذ من أطر القراءة كأساس لعملها هو ان أطر القراءة الطويلة دليل واضح على وجود الجين فمثلا وجود 30 شفرة متتالية بدون شفرة وقف دليل على وجود منطقة مشفرة وبالرغم من هذا فان الحد الحرج او حد العتبة Threshold لأطر القراءة المفتوحة تحدد بقيم أعلى من 50-60 شفرة والذي يخضع الى عمليات أخرى للتأكد منها ، مثلاً "بالإضافة الى البحث عن شفرات البدء والوقف يتم البحث عن توالي Shine-Dalgarno في الخلايا بدائية النواة. ومن المشاكل الأخرى إمكانية وجود بعض أطر القراءة المفتوحة خارج الجينات ولكن هذه تكون قصيرة وتنتهي بعد عدد من الشفرات لذلك لا تؤخذ بنظر الاعتبار. ومن جهة ثانية فان أطر القراءة المفتوحة تختلف في الخلايا حقيقية النواة ويمكن ان تستمر عبر الانترونات (في بعض الحالات) واستعمال أطر القراءة في حقيقية النواة يكون باستعمال mRNA الذي يكون حاملا لرسالة واحدة ولذلك يكون هنالك إطار واحد ولكن المشكلة تظهر عند استعمال pre-mRNA فهناك امتدادات طويلة من الانترونات ضمن الإطار التي لا تترجم لذلك كان من الضروري في هذه الحالة التعامل مع جزيئات mRNA الذي جرت عليها عمليات الخياطة المذكورة توا".

وفي الخلايا حقيقية النواة يتوقع ان تكون شفرة البدء في الاكسون الأول وشفرة الوقف في الاكسون الأخير وبذا تكون أطر القراءة المفتوحة هي الاكسونات. في حين ان الحالة تختلف في الخلايا بدائية النواة اذ تكون الجينات المشفرة بمثابة أطر قراءة مفتوحة وبعض الأحيان يحصل تداخل بين أطر القراءة كما في حالة الاوبرونات . وبعد عمليات البحث عن الأطر فان الجين الذي تم تحديده بالحاسوب يعرض لعملية ترجمة الى بروتينات التي يستعمل تواليها للمقارنة مع قواعد بيانات البروتينات.

استعمال الشفرات في التنبؤ

طرق بدأت بالاعتماد على ترتيب النيوكليوتيدات اللاعشوائي او تفضيل بعض الأحياء لاستخدام شفرة معينة . ففي طريقة استعمال الشفرات لإيجاد الجين يتم فيها البحث عن شفرات معينة في المناطق المشفرة وغير المشفرة وتستخدم برامج متخصصة بها لتقارن الشفرة مع جداول مرفقة مع البرنامج.

اما الطرق الأخرى فهي تعتمد الى دراسة الانحياز في GC خاصة في الموقع الثالث من الشفرة، اذ لوحظ ميل استعمال G او C أكثر من استعمال A او T ولذلك يرسم مخطط لمكونات GC والدرجات او الاحتمالية النسبية بطرق مختلفة، والحصول على قيم عالية في مناطق أعلى من التوزيع العشوائي يشير الى وجود أطر القراءة المفتوحة وبالتالي تميز المناطق المشفرة عن غيرها كما موضح في الشكل 66.5 :



شكل 66.5 : تحديد المناطق المشفرة بالاعتماد على انحياز المحتوى من GC

ولكن هذه الطرق تكاد تكون محصورة في الجينات المثالية وتتناسى الجينات الأخرى التي لا يكون فيها انحياز في القواعد، لذلك احتاجت عمليات التنبؤ إلى برامج إحصاء معقدة للوصول إلى دقة عالية.

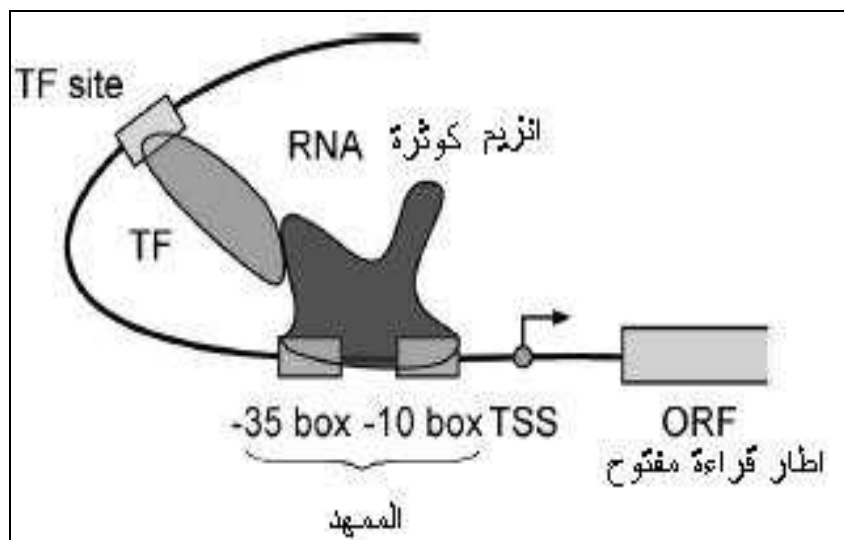
التنبؤ بالممهدات وعناصر التنظيم

الممهد عنصر أو قطعة من DNA تقع في منطقة قريبة من بداية الجين المشفر وهي ليست موقع بدء الترجمة. ويعمل الممهد كموقع لارتباط ماكينة الانتساخ المكونة من إنزيمات كواثر RNA وعوامل الانتساخ وهذا يعني أنها تنظم التعبير الجيني للجينات التي تقع أمامها وعلى الشريط نفسه باتجاه النهاية 5' من الشريط الحساس الذي يحوي على مواقع ارتباط عوامل الانتساخ وكذلك تحوي المنطقة على عناصر الاستجابة التي تساعد وتؤكد ارتباط RNAP وتوثيقه بواسطة عوامل الانتساخ. والممهدات وعناصر التنظيم الأخرى يتم تحديدها عادة بالتحليل التجريبي وإجراء مثل هذه التجارب يحتاج إلى وقت طويل، لذلك كانت عمليات الحرس أو التنبؤ بالحاسوب من العمليات الواعدة لأنها تمتلك القدرة على الإحلال محل التحليل التجريبي وتختلف الخلايا بدائية النواة عن حقيقية النواة في هذا المجال بشكل واسع.

أهمية إيجاد الممهدات

تكون مهمة إيجاد الممهدات في الأحياء حقيقية النواة الراقية صعبة والتي تكون بمثابة جزر صغيرة في خضم بحر من التواليات غير المشفرة. والسيطرة على الانتساخ تكون لها علاقة

بنمو الخلايا الذي يعد أحد الأسباب المهمة في السرطانات وكذلك لها علاقة بالربو والتهلاسيما نوع بيتا واعتلالات أخرى مثل متلازمة Rubinstein-Taybi وغيرها .
ممهّدات بدائية النواة
 يوضح الشكل التالي (67.5) مخطط بسيط لممهّد في خلايا بدائية النواة



شكل 67.5 : ممهّدات بدائية النواة

TSS: موقع بدء الانتساخ

ORF: إطار القراءة المفتوح

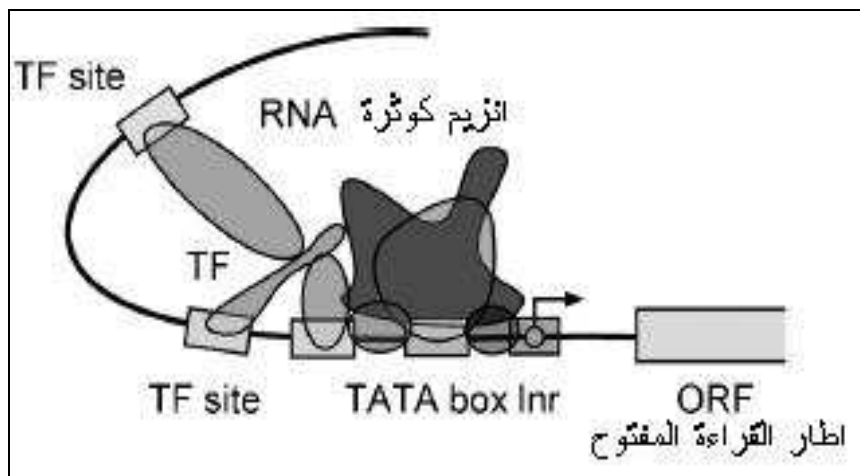
Pol: إنزيم كوثرّة Polymerase

TF: عامل انتساخ

ويرتبط إنزيم RNAP إلى منطقة الممهّد حيث يبدأ عملية الانتساخ وذلك بتداخله مع عوامل الانتساخ التي ترتبط إلى مواقع مختلفة. وفي الشكل تكون منطقة -35 مكونة من توالي TTGACA والمنطقة -10 مكونة من TATAAT (TATA box). وتوالي الممهّد يمكن أن يحدّد التعبير عن جين واحد أو عدد من الجينات المرتبطة به إلى جهة اليمين (Downstream) في حالة الاوبرونات التي يسيطر عليها بممهّد واحد. وتميّز الممهّدات بعوامل سيكما المختلفة التي تساعد في جلب النوع الوحيد من RNAP إلى المنطقة المراد انتساخها.

ممهّدات حقيقيات النواة

تكون ممهّدات هذه المجموعة معقّدة ويوضح الشكل التالي (68.5) مخطط لممهّد في الخلايا حقيقية النواة:



شكل 68.5 : ممهدات حقيقية النواة

ويوضح الشكل مدى تعقيد الممهد الذي ترتبط اليه إنزيمات كثر RNA وعوامل الانتساخ، وتكون هناك درجات إضافية من التعقيد في هذه الخلايا حيث يوجد ثلاثة أنواع من إنزيمات الكثر RNAPs هي I، II، III، والنوع الثاني هو الذي ينتسخ الجينات المشفرة للبروتينات. وفي حقيقيات النواة يكون لكل جين الممهد الخاص به كما ان ماكنة الانتساخ تحتاج الى عدد اكبر من البروتينات لبدء الانتساخ تصل الى 7 بروتينات. وفضلا عن ذاك فان هناك مناطق تنظيم قد تكون بعيدة عن الممهد مثل المشجعات وكذلك وجود المسكتات ومناطق العزل لتحديد مستويات الانتساخ. element/insulators Boundary.

مشاكل تحديد الممهدات

تحديد الممهدات وما يلحق بها من عناصر التنظيم الأخرى هي مهمة ليست بالسهلة وذلك لعدة أسباب منها:

- ان هذه المناطق لا تحدد بشكل دقيق وبعض الأحيان يكون فيها اختلافات كبيرة .
 - ان كل جين له تشكيلة خاصة من قطيقات التنظيم التي تحدد التعبير عنه في الزمان والمكان التي تقرر الانتساخه .
 - الممهدات وعناصر التنظيم لا تنتسخ وبالتالي لا تترجم الى بروتينات وبذا تقل حساسية الطرق الخاصة بالكشف عنها.
 - الممهدات وعناصر التنظيم تمثل امتدادات قصيرة وعادة تصل الى 6-8 من النيوكلووتيدات ومثل هذه يمكن ان توجد بالصدفة في اي توالي مما يؤدي الى إنتاج معدل عالي من النتائج الموجبة الكاذبة.
- لذلك احتاجت برامج حدس الممهدات الاعتماد على صفات الممهد المتعددة واستعمال برامج حاسوب معقدة.

عناصر الممهدات المستعملة في التنبؤ

يتكون الممهد من عدة أجزاء يمكن ان تستهدف عند وضع البرامج ومنها الممهد المركزي Core promoter ويمثل الحد الأدنى من تواليات الممهد التي تكون كافية لعملية انتساخ

ناجحة. وفيه ايضا مواقع بدء الانتساخ وتكون في الخلايا حقيقية النواة من الموقع 34- قاعدة من بداية موقع الانتساخ اي موقع ارتباط الإنزيم RNAP. والعنصر الآخر هو الممهد الأقرب Proximal promoter ويقع الى يسار موقع البدء ويصل طوله الى 250 قاعدة والتي ترتبط به بعض عوامل الانتساخ. والمكون الآخر هو الممهد البعيد Distal promoter وهي تواليات بعيدة عن موقع البدء وتقع ايضا الى يسار الجين ويمكن ان تحوي على مواقع ارتباطات إضافية ولكن تأثيرها اقل من تأثير الممهد الأقرب وهي تشمل اي تواليات بعيدة ولكن لا تشمل المشجعات التي يكون تأثيرها موضعي ومستقل عن الترتيب Orientation independent.

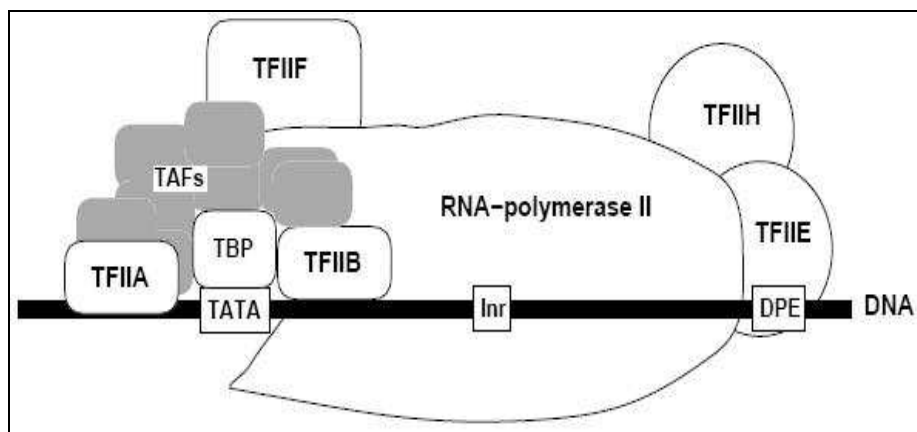
وفي بدائية النواة فالعناصر او الأجزاء المهمة هي 10- و 35- الواقعة الى يسار مواقع بدء الترجمة. ويطلق على الموقع 10- Pribnow box او العنصر 10- ويتكون من 6 قواعد هي TATAAT ولذا يكون هذا الصندوق أساسي في عملية الانتساخ. اما المنطقة 35- وتتكون من 7 قواعد TTGACAT فوجودها يؤدي الى زيادة معدل الانتساخ. وهذه التواليات تكون ثابتة في معظم الأنواع التابعة لبداية النواة التي تم تحديد تواليها ولكن ليس بشكل مثالي وفق الصيغ المذكورة وانما توجد بعض الاختلافات. وبعض الممهدات تحوي على عنصر يقع على بعد 50- ولذلك يبدو ان الموقع 35- غير مهما في عملية الانتساخ في الممهدات الحاوية على عنصر 50- . وكل ممهدات بدائية النواة يتم تمييزها بواسطة عوامل سيكما التي تتداخل مع RNAP وتشجعه للارتباط بالممهد المركزي.

اما ممهدات الخلايا حقيقية النواة فتكون متنوعة بشكل كبير ومن الصعوبة توصيفها بشكل مثالي والصفة العامة لها انها تقع الى يسار الجين الذي تنظمه ويمكن ان تحوي على عناصر تنظيم تبعد عدة كيلوات من القواعد عن موقع بدء الانتساخ TSS كما في حالة المشجعات . وفي هذه الخلايا فان تكون معقد الانتساخ يؤدي الى انحناء DNA ويؤدي ذلك الى مساعدة التواليات البعيدة للاقتراب من موقع الانتساخ.

وهناك حوالي 10-20% من جينات حقيقيات النواة تحوي على TATA box التي ترتبط الى بروتين رابط TATA binding protein (TBP) يساعد في تكوين معقد بدء الانتساخ مع RNAP، ويقع هذا الصندوق قريب جدا من موقع بدء الانتساخ وفي العادة يكون ضمن 50 قاعدة.

ومن الاختلافات في ممهدات الخلايا حقيقية النواة التي تؤدي الى عدم فعالية برنامج واحد للنتيؤ بها هو ان بعض الخلايا تحوي على ما يسمى بالممهدات الدنيا Minimal promoter الذي يمثل اقل توالي يكفي لتجميع معقد البدء الأولي وهذا النوع من الانتساخ يسمى الانتساخ الأساسي Basal transcription وهناك أكثر من صنف من هذه الممهدات الصغيرة التي تسبب مشكلة للبايولوجيين العاملين في مجال الحاسوب للتعرف عليها، فضلا عن وجود أكثر من ممهد أدنى بالصفات المذكورة آنفا مما يزيد المشكلة تعقيدا. ومن هذه الحالات ان بعض الممهدات تتكون من TATA box فقط ويقع على بعد 30 قاعدة من الجين الذي ينظمه بعد ارتباطه الى TFIID الذي يعرف بالبروتين الرابط TBP فضلا عن وجود عوامل ارتباط أخرى. اما النوع الآخر من الممهدات الدنيا فهي التي تخلو من TATA (TATA-less) وفي

وهناك مميزات تحوي على الـ Inr و TATA box وكذلك توجد مميزات لا تحوي على الـ Inr. وهناك عنصر جديد اكتشف في الإنسان والـ دروسوفلا ويكون موجودا في المميزات الحاوية على كل من Inr و TATA box ويقع على بعد 30 قاعدة يمين نقطة بدء الانتساخ ويساعد Inr في السيطرة على بدء الانتساخ ويسمى (Downstream promoter element) DPE الموضح في الشكل 69.5 :



شكل 69.5 : عناصر الممهدات في بعض الأحياء

كل ما ذكر أعلاه من عناصر في تركيبة الممهدات لا يمكن ان تؤدي الى التنبؤ بالممهد بشكل صحيح وذلك لعدة أسباب منها مثلا ان Inr الكامل وجد انه يمكن ان يتكرر كل 512 قاعدة في تواليات عشوائية وكذلك أظهرت بعض البرامج المستعملة لتحديد ممهدات اللبائن ان هناك TATA box واحد 120 قاعدة اعتمادا على المصفوفات الوزنية او الترجيحية وهذا يشير الى حتمية وجود عوامل أخرى في الممهدات بجانب تواليات الممهد المركزي وهذه تكون فاعلة في داخل الأنظمة الحية وتعتمد على التركيب الثلاثي الجسم للـ DNA داخل النواة الذي له تأثير كبير في عملية الانتساخ، ووجود زيادة في مواقع الارتباط التي ترتبط الى بروتينات معينة وكون الأخيرة يمكن ان ترتبط الى أكثر من موقع فضلا عن تكوينها كموثرات ثنائية او مزدوجات كل هذه لا يمكن ان تفسر ببرامج تعتمد على تواليات DNA . فالممهدات تعمل اعتمادا على الموقع الذي تكون فيه فهي اما ان تعمل بصورة مستمرة كما في جينات الإدامة التي تحوي على TATA box او نتيجة للاستجابة للمؤثرات الخارجية. وكذلك فان عوامل

الانتساخ بصفاتها بروتينات تتعرض هي الأخرى لعمليات تنظيم وبذلك تكون العمليات بمثابة شبكات تؤدي إلى حصول عمليات سيطرة رجعية. وبعض الأحيان تقع العناصر المنظمة إلى يمين الجينات التي تنظمها بدلا من الموقع اليساري التقليدي، كما أنه في حالات عديدة تتجمع مواقع ارتباط عوامل الانتساخ بشكل مجموعة أو عناقيد تنتشر على مدى واسع لتعمل بشكل تآزري لتشجيع عملية بدء الانتساخ. وما ذكر أعلاه مهم ويوافق هوى البيولوجيين ولكنه يشكل مشكلة كبيرة عند التنبؤ أو حدس الجين. فعوامل الانتساخ كثيرة تصل إلى الآلاف والعديد من التواليات التي تميزها أو تتعرف عليها في DNA غير موصوفة لحد الآن، كما أن بعض التواليات المرشحة لارتباط عوامل الانتساخ قرب الممهد قد لا ترتبط أبداً نظراً لعدم وجود عامل الانتساخ المعني تحت ظروف خاصة عند انتساخ الجين.

ومما يزيد من صعوبة حدس الممهدات والعناصر المنظمة هو قصرها وعدم ثبوتها بشكل كبير من حيث التوالي وكذلك من حيث توزيعها على الجينوم واعتماداً على هذا فإن التحليل الإحصائي يشير إلى أن عناصر التنظيم المعروفة في الوقت الحاضر لا تحوي على معلومات كافية للتمييز بين مناطق الممهد والمناطق الأخرى. وهذا يبدو مثبطاً للهمم ولكن من الضروري تذكر أنه داخل الخلايا يتم تمييز الممهدات بشكل صحيح بواسطة آلة الانتساخ. وهذا يعني أن المعلومات الضرورية يجب أن تكون موجودة ولكن لا بد من وجود أسباب وراء مثل هذا الإخفاق.

علاقة الممهدات بعملية التطور والظروف الخارجية

التغيرات في الممهدات أكثر أهمية من التغيرات في الجينات المشفرة على مدى الزمن، وأهمية الممهدات تأتي من قدرتها على إدماج الإشارات الداخلية والبيئة للتغير في التعبير الجيني والعديد من التغيرات في البيئات الداخلية والخارجية لها تأثير في التعبير الجيني المعتمد على هيئة الممهد Promoter configuration وهذه التي يجب أن تؤخذ بنظر الاعتبار عند وضع برامج التنبؤ. وكذلك التوليفات وترتيب لتواليات خاصة والعناصر التي ترتبط إلى الممهد في نقطة زمنية محددة. فعند ورود إشارات فسلجية أو مرضية أو دوائية للخلايا يتم تحويل عدد من البروتينات بتفاعلات تسمى شلالات الإشارة ويتغير تركيب بعضها ويكتسب القابلية للدخول إلى النواة للارتباط بممهدات DNA أو ترتبط ببروتينات هي أصلاً مرتبطة بالـ DNA والمعقدات المتكونة لها القابلية على تغيير مستويات التعبير الجيني إما بالزيادة أو النقصان. والمصدر OMIM المدرج ضمن الموقع NCBI يعطي معلومات عن العلاقة بين الطفرات والتغايرات الطبيعية في الجينات ومدى إمكانية الإصابة بالأمراض، وهذه تحتاج إلى إستراتيجيات معقدة ومتطورة لاستخلاص ومعرفة الأمراض ذات العلاقة بعملية السيطرة على الانتساخ التي يشارك فيها الممهد بشكل مباشر خاصة في الأمراض متعددة الأسباب ويعزز هذا معرفة الأعراض والاستجابة أو عدم الاستجابة للعلاج والذي يشير إلى اختلاف الأصول الجزيئية للأمراض التي تختص بدراستها Pharmacogenomics، ومن هذه الأمراض بعض أنواع السرطانات التي تحدث نتيجة لتطور جينات متباينة الأصول Chimeric genes نتيجة لحركات الكروموسومات.

مثيلة DNA وتداخلها مع التنبؤ بالممهدات

تكثر حالة مثيلة DNA في جزر CpG وهذه الجزر يصل امتدادها الى حوالي 200 قاعدة ويكون أكثر من 50% من مكوناتها من G+C اي ان PG+PC اكبر من 0.5. اي ان تردد هذين النيوكليوتيدين CpG على الأقل 0.6 من المتوقع على أساس المحتوى النيوكليوتيدي في المنطقة أي ان

$$\text{PCpG} > 0.6 \times \text{pG} \times \text{pC}$$

(P على يسار القاعدة تمثل الاحتمالية اما على يمينها فتمثل الأصرة Phosphodiester bond)

ومعظم هذه النيوكليوتيدات الثنائية في الفقرات (98%) تكون حاوية على المثل على الموقع ذرة الكربون الخامسة لحلقة القاعدة النروجينية السايترز، وهذا يعاكس الحال مع اللافقرات متعددة الخلايا حقيقية النواة الذي يكون فيها المثل غائبا او موجودا في مناطق محددة. ان نمط المثيلة في الفقرات يحدد من المراحل الجنينية المبكرة ويستورث في الخلايا البنية بعد الانقسام ، وقد مر في موقع آخر ان الإنزيم المسئول عن المثيلة يعمل على الأشرطة الجديدة التي تقابل CpG الحاوية على المثل. وتشكل CpG حوالي 25% من مكونات القواعد، والقلّة تكون ناتجة من إزالة الأمين من السايترز لتكوين الثايمين عند حدوث الطفرات والذي لا يمكن ان يميز عن الثايمينات الأخرى بواسطة أنظمة إصلاح DNA وبمرور الزمن تتحول CpG الى TpG وفي حالة التنبؤ بالممهدات فان الأجزاء غير الحاوية على المثل هي التي تستهدف بالبرامج ومثل هذه تصل نسبتها الى 2% من الجينوم وتوجد عادة في النهاية 5' ويمكن ان تتداخل مع أول اكسون من الجين.

ان العلاقة بين CpG والممهدات ربما تكون تاريخية أكثر مما هي وظيفية ولكنها يمكن ان تكون مهمة للتنبؤ بالممهدات فهي تكون بمثابة إشارات شاملة ولا يوجد في الوقت الحاضر برامج ذات استعمال واسع تستعمل هذه العلاقة في حدس الممهدات، ولكن هناك برامج تستعمل نماذج ماركوف الخفية وتعتمد هذه العلاقة للتعرف على الممهد المركزي وكذلك في برامج قليلة أخرى.

وهناك العديد من البرامج المعتمدة على تحديد جزر CpG تستعمل في التنبؤ بممهدات حقيقية النواة خاصة عن استعمال مؤشرات أخرى والأخذ بنظر الاعتبار حالة التعاونية Cooperativity بين عوامل الانتساخ التي تكثر كثافتها في ممهدات هذه الخلايا لغرض زيادة حساسية البرامج في إيجاد الممهدات.

التركيب الكروموسومي وإحباط الانتساخ وعلاقته بحدس الممهدات

ان المادة الوراثية DNA معبأة في تراكيب هي كروموسومات الموجودة في النواة، وفي الإنسان فان هذه التعبئة تجعل من الممكن تعبئة 2 متر من DNA في النواة التي قطرها يصل الى القليل من المايكرومترات ، والجسيم النووي الحاوي على حوالي 146 قاعدة ملفوفة حول مكوثر ثماني من الهستونات مكونا الجسيم النووي، والتركيب الأعقد يتكون من التقاف الجسيمات النووية التي يثبت بالهستون H1 . هذه التعبئة تقلل من DNA المتاح لأجهزة الانتساخ الأساسية وتمنع الانتساخ داخل الكائنات الحية مقابل إمكانية انتساخ DNA العاري،

ووجود الكروماتين (DNA + البروتينات) يمكن ان يمنع او يحبط الانتساخ وهذا قد يكون وراء عدم فعالية الممهدات الخفية في الخلايا الحية وهي من وراء التنظيم الصارم للجينات في الكائنات الحية. والملاحظ ان التنشيط بالمنشطات لقالب DNA العاري يكون خارج الأنظمة الحية بمعدل 10 مرات ولكنه يصل الى آلاف المرات داخل الأنظمة الحية. ولذلك يلاحظ ان إزالة الكبح عند بدء الانتساخ بعمليات فك طويات الكروماتين يكون جزءا مهما في تنظيم فعالية الجين وهناك العديد من عوامل الانتساخ والعوامل المشجعة للانتساخ تعمل على اضطراب تركيب الكروماتين.

والحقيقة الواضحة ان تركيب DNA يؤثر في عملية الانتساخ وهذا يشير الى ان تحليل تركيب DNA ومواقع الجسيمات النووية يمكن ان تكون مساعدة في حدس الممهدات. والملاحظ ان التركيب الثلاثي الأبعاد وبالتالي مواقع الجسيمات النووية داخل الأنظمة الحية تتأثر بشكل مباشر بتواليات النيوكليوتيدات وهذا يشير الى إمكانية معرفة الصفات او المظاهر الرئيسية حتى تلك الخاصة بالصفات اللاجينية من عمليات تنظيم الانتساخ وذلك بتحليل تواليات DNA ، ومثل هذه الإشارات تكون مفيدة في عمليات التنبؤ بالممهدات. فضلا عن ذلك فان حالات تقوس DNA (DNA Curvature) التي تختلف في المناطق المحيطة بـ TATA box يمكن ان تستعمل في حدس الممهدات.

وعند دراسة العديد من الممهدات التي لا علاقة بينها تبين ان المناطق اليمنى لمنطقة البدء تكون تراكيب مقوسة بشكل يشابه للـ DNA في الجسيم النووي، ونظرا لكون طول المنطقة القابلة للتقوس هي مساوية للقطعة من DNA التي تلفت حول المجمع الثماني من الهستونات في الجسيم النووي فان هذا يحد على افتراض ان إشارة موقع الجسيم النووي تكون مباشرة عند نقطة بدء الانتساخ. ان وجود الجسيم النووي قرب نقطة بداية الانتساخ قد يشير الى العلاقة القوية لتنظيم التعبير الجيني التي تلاحظ داخل الكائن الحي. ويحدث تقوس الجسيم النووي على مدى 7 قواعد هي PyPy +₁ N [TA] PyPy وتكون ذروة التقوس عند الموقع +1 اي (PyA +₁ N) والتي سيرد ذكرها لاحقا في موضوع أطلس الجينومات .

وتحتوي الكروموسومات على تواليات تربطها بالأرضية النووية التي يطلق عليها منصات الارتباط Scaffold attachment regions (SARs) او تسمى Matrix attachment regions (MARs) ويكون الارتباط بتكوين عروات من DNA التي ربما يكون لها دخل في التعبير الجيني اي انها المناطق التي تتداخل فيها كل من المشجعات والمحبطات لتمارس فعلها. وهذه الدومينات لها علاقة بالسيطرة على الانتساخ بشكل مشابه لعملية او ظاهرة Position effect variegation (PEV) وهي الظاهرة التي تكون واضحة عند نقل الجينات التي تقم في مناطق مختلفة من DNA في الخلية وتؤدي الى إسكاته ، وأحد التفسيرات ان الاختلاف في التعبير يعود الى انتشار الكروماتين المكثف من مناطق الكروموسوم المجاورة الى DNA الجين المنقول. ويعتقد ان التعاونية بين المعقدات البروتينية التي تتداخل مع الجسيمات النووية هي الأساس في تكوين الكروماتين المتباين وبطريقة مماثلة لتكوين معقدات الإسكات التي تشمل مجموعة من البروتينات مثل تلك المتكونة عند نهاية الكروموسومات في الخمائر. وبعض الأحيان فان إدخال الجين في بعض المناطق المحيطة بجينات حقيقيات النواة تكون له القابلية

على إزالة تأثير الموقع في إسكات الجين وتحد من انتشار الكروماتين المتباين. ومن هنا يتضح ان هذه الدومينات (MARS) فضلا عن كونها حدود مكانية الا انها يمكن ان تكون دومينات وظيفية وقد تكون مناطق عزل وظيفي، لذلك كان لها دورا في إيجاد الممهدات اذ يتطلب إيجاد التواليات التي تحد هذه الدومينات المشاركة في التعبير الجيني وهذه الإشارات التي تكون على جانبي الجينات تكون بمثابة إشارات شاملة ذات فائدة في أداء البرامج الباحثة عن الممهدات اي ان الكشف عن MARS تكون مفيدة في الكشف عن الممهدات.

طريقة إيجاد الممهد اعتمادا على نسق التعبير الجيني

التطورات الحديثة في الطرق السريعة لتحليل نسق الانتساخ مثل استعمال مصفوفات DNA الدقيقة أتاحت إمكانية تسجيل التعبير الجيني لمئات او آلاف الجينات في وقت واحد ، ويعتقد ان التعبير المترامن ناتجا من ممهدات وعناصر تنظيم مشتركة، واذا كان هذا الافتراض صحيحا فان التواليات اليسارية للجينات المترامنة التعبير يمكن صفها مع بعضها لغرض إيجاد عناصر التنظيم المشتركة التي يمكن تميزها بعوامل انتساخ متخصصة. ان العمل في هذا الاتجاه للإيجاد يعتمد أساسا على العمل التجريبي ويكون مقدمة لإيجاد مواقع ارتباط عوامل الانتساخ وذلك لان عناصر التنظيم للجينات المترامنة التعبير هي عادة قصيرة وضعيفة ونمطها يصعب تمييزه باستعمال عمليات الاصطفاف التواليات المتعددة البسيط، لذلك احتاجت هذه الى إيجاد طرق لعمليات إيجاد النسق مستقلة او غير معتمدة على الاصطفاف لإيجاد التفاصيل الدقيقة لتواليات القطيفات.

ومن المعوقات الأخرى في استعمال هذا التوجه ان تحديد نسق التعبير عن الجينات المترامن يعتمد على طرق العنقدة التي تشتهر بحدوث الأخطاء، وهذا يعني ان نوعية وجودة البيانات المستعملة كمدخلات يمكن ان يكون موضع شك خاصة عند استعمال جينات ليست لها علاقة وظيفية او متباعدة نوعا ما والتي يمكن ان تتعقد سوية .

وفضلا عن ذلك فان الافتراض الذي يمكن ان تؤسس على ضوءه البرامج والمذكورة أعلاه وهو ان الجينات مترامنة التعبير تحوي على ممهدات وعناصر تنظيم مشتركة ليس صحيحا دائما. فالكثير من الجينات المترامنة التعبير موجود في مسارات متوازية وتكون تحت سيطرة عمليات وآليات تنظيم غير متشابهة لذا وجب أخذ الحيطة والحذر عند استعمال هذه البرامج. وفي الخلايا حقيقية النواة يستعمل نسق التعبير ويخزن في قاعدة بيانات تسمح فيما بعد بمقابل التواليات غير معروفة لإيجاد المناطق والأنماط الثابتة.

الوراثة اللاجينية وإيجاد الممهدات

مر سابقا ان تحديد التواليات في قطع DNA التي ترتبط الى عوامل الانتساخ وغيرها من المعلومات حول تواليات DNA لا تكون كافية لتحديد الممهدات وهذا ما أشارت اليه التحليلات الإحصائية التي أوضحت ان كثافة مواقع ارتباط عوامل الانتساخ تكون غير كافية وغامضة ولا تستطيع التمييز بين الممهدات والمناطق الأخرى، وهذه قد تكون غريبة من الناحية البيولوجية لان الأحياء تستطيع السيطرة على آلاف الجينات، وهذا يشير الى مشاركة الوراثة الخارجية او اللاجينية في السيطرة على الانتساخ ، اي ان عمليات التنظيم لا يمكن ان تكون معتمدة على تواليات DNA لوحدها ، فهناك حالة طوي الكروماتين الذي يؤثر بشكل كبير في عملية الانتساخ وكذلك مثيلة DNA التي وردت بعض جوانبها سابقا. والحقيقة تشير الى ان عملية بدء

الانتساخ تعتمد على مدى كبير من التداخلات التي ترتبط الى الممهد والعوامل الأخرى التي ترتبط الى مناطق بعيدة، لذلك كان لابد من استعمال إشارات شاملة أخرى فضلا عن الإشارات الموضوعية التي تستعمل في البرامج الموجودة في الوقت الحاضر مثل TATA، Inr و CCAAT وغيرها من التواليات.

والإشارات الشاملة المعنية تشكل سؤالاً كبيراً ولكن ما مدرك منه في الوقت الحاضر هو استعمال جزر CpG وكذلك حدود الدومينات الكروموسومية او عوازل الدومينات والتي مر مناقشتها آنفاً.

وعليه فيفضل استعمال عدة إشارات لوضع برامج الكشف عن الممهدات (كما سيأتي ذكره لاحقاً)، منها الشاملة والموضوعية فضلا عن استعمال المظاهر الكاشفة المستعملة. فاستعمال الإشارات الشاملة يكون مفيداً لأنها توجد نوع من المرونة في الحدود الفاصلة Cutoff للإشارات الموضوعية وتبعد العديد من حالات الالتباس .

مناقشة وضع البرامج الخاصة بالكشف عن الممهدات

بعد مناقشة بعض جوانب إمكانيات الكشف عن الممهدات كان الهدف منها الوصول الى أفضل إمكانيات لوضع البرامج، وكانت المناقشة السابقة ضرورية لان معظم البرامج في هذا المجال لا يزال أداءها سيئاً ، فبدائية يجب تحديد الهدف هل هو الكشف عن نقطة بدء الانتساخ المضبوطة ام الكشف عن المنطقة التي يوجد فيها الممهد. والأفضل الجمع بين هذين التوجيهين. وذلك لان استهداف نقطة البدء ستعرقل من حدس الممهد فهي قد تكون قرب توالي الممهد وليس موضعاً آخر، وفي بعض الأحيان فان بعض الممهدات تكون لها أكثر من نقطة بداية، فضلا عن احتمالية وجود نقطة بدء واحدة لعدد من الممهدات الموجودة في قواعد البيانات. ولكن مع هذا فالبرامج الموجودة في الوقت الحاضر هي أفضل من سابقتها ولكنها لا تزال بعيدة عن الكمال نظراً لوجود مستوى عال من النتائج الموجبة الكاذبة. فمثلاً بعض البرامج تحدد 1 ممهداً كيلوقاعدة ولكن وجد في جينوم الإنسان ان المعدل يشير الى وجود جين واحد لكل 30-40 كيلو قاعدة. كما ان المعلومات الحيوية المهمة في الجوانب اللاجينية واشتراكها في التنظيم تضعف من موقف برامج إيجاد الممهد مقابل الوراثة اللاجينية ، اي ان آلة النسخ مصممة للتعامل مع الكروموسومات وليس مع توالي DNA المفرد كقالب في حين ان التوجيهات الحاسوبية تعمل للبحث عن الممهد في توالي شريط مفرد.

ومشكلة أخرى تنبثق من السؤال هل ان الخوارزمية الموضوعة يجب ان تميز كل الممهدات وغير الممهدات؟ وهذا بطبيعة الحال كبيرة على خوارزمية واحدة، فضلا عن الحاجة الى إيجاد الممهدات الثانوية والتي تحتاج الى خوارزميات ثانوية تبحث عن إشارات خاصة للكشف عنها، وان كان في هذا الموضوع حل مثل هذه المشاكل يكون بضم الخوارزميات سوية .

ومن جهة ثانية هناك بعض عناصر تنظيم الانتساخ لا يمكن استنتاجها من توالي DNA (كما مر في حالة تركيب الهستونات وتحويراتها) كما ان هناك حقيقة أخرى هي ان التواليات المشفرة ليس لها تأثير في عمليات الانتساخ، وان كانت هناك بعض المناطق اليمينية الموقع يمكن ان تساعد في تشخيص بعض الممهدات في البكتريا حيث تقع الممهدات في المناطق البينية .

ويتضح من فشل برامج الكشف عن الممهدات انها لا تأخذ بنظر الاعتبار الكفاية من المعلومات البايولوجية ذات العلاقة ، ولتحسين أداء البرامج هناك بعض النقاط الواجب أخذها بنظر الاعتبار عند حدس الممهد وحالة العموم مقابل الخصوص ، فهناك فرق بين الكشف عن كل الممهدات مثلا في حقيقات النواة وبين الكشف عن بعض الممهدات النشطة في بعض أنواع الخلايا عند وقت محدد أثناء التطورات الجنينية مثلا.

ولحل المشكلة الأولى وهو إيجاد كل الممهدات لابد من تشخيص بعض الصفات او المزايا المشتركة بين كل الممهدات وغير موجودة في باقي الجينوم وإلا انقسمت عمليات إيجاد الممهدات الى عدة مشاكل ثانوية وهذا يعني الحاجة الى تصميم خوارزميات خاصة للكشف عن كل صنف خاص من الممهدات وفي هذه الحالة اجري بعض التحسين في تطوير برامج تكون خاصة بالبحث عن الممهدات الموجودة في العضلات، وكذلك تطوير برامج كشف الممهدات من نوع واحد او مجموعة أنواع تعود الى وحدة تصنيفية خاصة.

وهذا يعني ان الاختيار الأفضل هو الأهم، فالبرامج الحالية تعتمد الى تمييز مواقع ارتباط عوامل الانتساخ الموجودة في قواعد البيانات الخاصة وهذا قد يكون مقبولا لحد ما، ولكن العديد من مواقع الارتباط المسجلة في قواعد البيانات ينقصها الدليل التجريبي وهذا يؤثر بشكل كبير في أداء البرامج المعتمدة عليها ولذا يجب التأكد من مواقع الارتباط من المراجع الأصلية وهذا سيكون عملا مجهدا خاصة عند ازدياد عدد مواقع الارتباط. فضلا عن المشاكل البايولوجية الحقيقية عند استعمال قوائم مواقع الارتباط فمثلا عامل انتساخ واحد يرتبط الى مجموعة من المواقع في 50 ممهد مختلف وهذا يعني ان التواليات في هذه المواقع لابد ان تظهر نوعا من الاختلاف او التنوع ولا يتوقع ان يكون هناك توالي منضبط بشكل كامل او عام Canonical sequence ، ففي حالة ارتباط مواقع عوامل الانتساخ قد يكون هناك توالي واحد يمكن ان ترتبط به عوامل الانتساخ بشكل قوي تحت ظروف خلوية خاصة ، ولكن في الانتخاب الطبيعي فان الارتباطات ذات الطاقة القليلة هي المفضلة ولذلك تستعمل التواليات الموجودة في الأنواع الطبيعية لتكون هي الممثلة . وذات الشيء يكون صحيحا في حالة التعبير عن القوة (طاقة الارتباط) للتدخلات المقابلة للبروتينات مع DNA. فمثلا البروتين الذي يرتبط الى بروتين آخر ليساعد في الارتباط الى DNA يكون ارتباطا "ضعيفا" وعليه لو ان 50 موقع ارتباط موجودة في قاعدة البيانات استعملت لتمييز الموقع فان المعلومات الموجودة في المواقع الضعيفة ستكون مؤكدة عليها بشكل كبير مما يؤدي الى ان طريقة الحدس تعطي عدد كبير من النتائج الموجبة الكاذبة. ولذلك كان لابد من استعمال طاقات الارتباط لتدخلات البروتينات مع DNA لترجيح تواليات محددة عند وضع الخوارزميات. وبذلك يتضح ان طرق تنبؤ وحدس الممهدات تفشل في تحديد المواقع الضعيفة.

كما ان نوعية قواعد البيانات المستعملة تكون مهمة فهناك بعض القواعد التي تضم نقاط بدء الانتساخ المستقاة من التجارب وهذا يعني ان كل معلومة فيها قد تم مقارنتها بالمراجع الأصلية. ونقطة أخرى يمكن ان تثار وهي كمية التوالي (طوله) لمعرفة نقطة بدء الانتساخ فلا بد من استعمال التواليات حول نقطة البدء من الجانبين فمثلا عند استعمال جزر CpG او MARs يجب استعمال كميات اكبر من DNA التي تحيط هذه المناطق عند إدخالها للبرامج. فضلا عن

وجوب الأخذ بنظر الاعتبار التركيب ثلاثي الأبعاد للمناطق التي تحيط بمناطق البدء التي نبحث عنها.

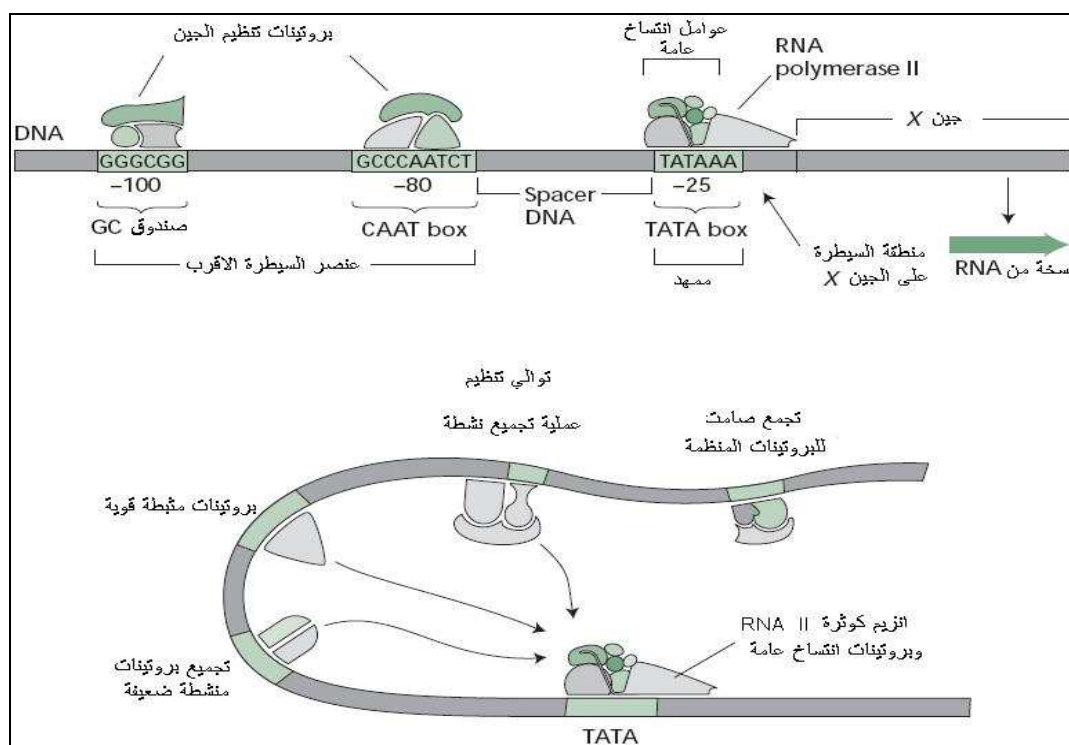
ومن هنا لابد من إيجاد توجهات جديدة في تصميم البرامج الخاصة بحدس الممهدات وقد بدأت بعض البرامج تأخذ بنظر الاعتبار العلاقة بين التراكيب الدقيقة والوظائف مثل استعمال طرق الشبكات العصبية ونماذج ماركوف الخفية التي تستطيع إدماج معلومات عن التواليات المجاورة للممهدات. ولزيادة التخصيص فان بعض البرامج تحذف التواليات المشفرة قبل بدء العمل وتركز على المناطق الواقعة التي تمتد الى 0.5-2 كيلو قاعدة والتي يتوقع ان تقع فيها الممهدات وبذا يكون حدس الممهد والجين المرافق له عملية متلازمة.

التنبؤ بجينات الخلايا حقيقية النواة

فضلا عن ما ذكر آنفا من طرق إيجاد الجينات والممهدات بصورة عامة الا ان هناك بعض الجوانب الخاصة بالخلايا حقيقية النواة تتعلق بطبيعة تركيب جيناتها الخاصة من حيث وجود الانترونات والاكسونات واحتمالات الخياطة البديلة وغيرها من الجوانب.

وبدءا لابد من التذكر ان التعريف بالجينات في هذه الخلايا ليس بالمهمة السهلة وذلك لكبر الجينومات التي تصل أطوالها من 10 ميكاقاعدة الى 670 ميكاقاعدة ، فضلا عن ان الكثافة الجينية فيها منخفضة جدا ففي الإنسان مثلا هناك 3% او اقل من الجينوم هي جينات اي بمعدل جين واحد 100 كيلوقاعدة. والمسافة البينية بين الجينات تكون كبيرة جدا فضلا عن احتوائها على تواليات مكررة وعناصر قافزة، كما ان الجين الواحد يكون موزائكي لاحتوائه على الاكسونات والانترونات.

هناك عدد كبير من جينات حقيقيات النواة التي تم تحديد تواليها وتوفرت الخرائط الوراثية والفيزيائية في تحديد مواقعها واتجاهات DNA وبذلك أمكن تهميشها وتوضيحها بطريقة مشابهة للكروموسومات البكتيرية. وتحتاج جينات حقيقيات النواة الى برامج معقدة للتعرف عليها. وعلى العموم فان تحديد تواليات الجينومات يحتاج الى عمليات تنظيم وترتيب كبيرة وتحتاج الى تمويل مالي كبير، فبالنسبة للكائنات التي لم تحدد توالياتها بشكل كامل فالطريقة العادية هي وضع الخرائط والتعرف على الجينات بشكل فردي. ويوضح الشكل التالي (70.5) خريطة عامة لجينات الخلايا.



شكل 70.5 : خارطة عامة لجينات خلايا حقيقية النواة

وللمعلوماتية الحيوية مهام ضرورية في هذا المجال عند القيام بمشاريع تحديد التواليات على مقياس كبير وذلك لمليء الفجوات بين التواليات والوظائف التي تقوم بها وهناك عدد من البرامج التي طورت لمسح وتهميش التواليات الخام ولكن المعروف انه اي من البرامج لا يعطي دقة 100% لذلك وكما هو الحال في جميع الأوضاع هو التوصية باستعمال أكثر من توجه وأكثر من برنامج.

تمييز جينات حقيقية النواة

يفترض في الخوارزميات المستعملة ان تكون قادرة على التمييز بين DNA الجيني و DNA غير الجيني وان تحديد أطر القراءة والترجمة المفتوحة تكون غير كافية وانما هناك جوانب أخرى يجب أخذها بنظر الاعتبار ومنها الإشارات ومحتوى الجينات والتماثل وغيرها. الإشارات: وهي قطيقات تمتاز بتواليات خاصة ومن هذه الإشارات عناصر الممهّدات ومواقع ذبول الأذنين ومواقع معطي الانفلاق والموقع المكتسب بين الاكسونات والانترونات. والخوارزميات التي تعمل في هذا المجال هي قياسات الإشارات Signal sensors (المارة الذكر) والتي تبحث عن إشارات معينة بشكل منعزل او مجموعة من الإشارات.

وتستعمل مع هذه الإشارات مصفوفات الترجيح Weight matrices لتحديد الدرجات لكل موقع من مواقع الإشارة المخمّنة وكذلك لتحديد الدرجات لكل قطعة او قطيفة وكل هذه تستعمل لتحديد فيما اذا كانت القطعة المعنية هي إشارة حقيقية ام لا. ومن الجدير بالذكر ان تحديد الإشارات وحده لا يكفي للتعرف على الجينات في حقيقيات النواة وذلك لان هذه الإشارات

مختلفة بشكل كبير جدا. فمثلا TATA box من العناصر المسيطرة على عمليات الانتساخ ويقع في العادة أمام الجين على بعد حوالي 30- قاعدة وهذا يشير الى انها إشارة ملائمة للتنبؤ ولكن الملاحظ ان هذا الصندوق يوجد في حوالي 70% من الجينات وليس كلها، الإشارة الأخرى هي ذيل الأدنين الذي يوجد في الجينات ولكن في عدة حالات فان التوالي يمكن ان يختلف عن التوالي التقليدي (AATAAA). الإشارة الأخرى هي مواقع الانفلاق بين حدود الانترونات والاكسونات والتي تكون ثابتة جدا في التركيب ولكن مرة أخرى هذه توجد في الجينات الحاوية على الانترونات فقط.

محتوى الجينات: من المؤشرات التي تؤخذ بنظر الاعتبار عند تصميم البرامج والملاحظ ان المحتوى هي ليست قطيفة قصيرة وانما هي امتدادات من التواليات المختلفة الطول وتختلف عن محتوى DNA المجاور، فمثلا DNA المشفر وغير المشفر يختلف بشكل كبير في تردد المكوثر السداسي . وهناك اختلافات ايضا في تردد التواليات مثل وجود نيوكلوتيدات معينة في مواقع مختلفة في الوقت نفسه. كل هذه تحسب إحصائيا عند تحديد حجم النوافذ الخاصة Specific window size عند تصميم البرامج وتتم مقارنة التردد والاعتمادية بين النوافذ.

ان تردد المكوثر السداسي وغيره من المؤشرات ضمن محتوى الجين تعتمد بدورها على أفضلية استعمال الشفرات من قبل الكائن، فالأحياء المختلفة تفضل شفرات مختلفة للحامض الاميني نفسه يقع ضمن ظاهرة تشتت الشفرات Codon degeneracy والذي يعكس وجود جزيئات tRNA مختلفة للحامض الاميني الواحد. لذلك فان برامج الحاسوب يجب ان تدرب باستعمال جداول خاصة بتفضيل الشفرات للكائن الذي أخذ منه التوالي وبذا كانت هناك برامج متخصصة بالكائن او النوع.

وفي مجال المحتوى الجيني يمكن ان تستعمل مكونات الجين الأخرى مثل البحث عن ترددات عالية غير طبيعية للنيوكلوتيدات الثنائية CpG (اي جزر CpG) في جينومات الفقريات او استعمال مكررات للنيوكلوتيدات الثلاثية CNG في جينومات النباتات.

ومن الواضح ان الخوارزميات التي تعتمد على استعمال الإشارات ومحتوى الجينات تقع ضمن مجموعة طرق البداية (الواردة سابقا) المعتمدة على مكونات الجين نفسه ولا تدخل فيها المعلومات من الجينات الأخرى المعروفة. لذلك قد تكون دقتها قليلة وعليه فهي تعتمد على القيم الإحصائية وذلك لان تحديد مواقع المناطق المشفرة لا يكون مضبوطا بشكل كامل كما ان الطرق تعتمد على فلق الجين الى عدة قطع ثم تجميعها مرة أخرى. وبعض هذه البرامج تؤكد على اكسونات الطرف 3' والأخرى تؤكد على اكسونات الطرف 5' وهذه تشكل مشكلة لان الانترونات والمناطق البينية لا تختلف كثيرا خاصة بالنهاية 3' و 5' ، ولذا كان من الأفضل الجمع بين كل الإشارات المتاحة.

التماثل الجذري Homology

تستعمل طريقة التماثل او التشابه الجذري ايضا في تمييز جينات الخلايا حقيقية النواة اي اجراء عمليات الاصطفاف مع جينات أخرى معروفة او مع ESTs.

والواضح ان أفضل البرامج هي التي تدمج ما ذكر أعلاه من استعمال الإشارات ومحتويات الجين وكذلك التماثل الجذري للجينات في قواعد البيانات كما في حالة استعمال BLAST و FASTA. ولكن يجب الحذر فان جينومات حقيقية النواة تحوي على الجينات الكاذبة (التي

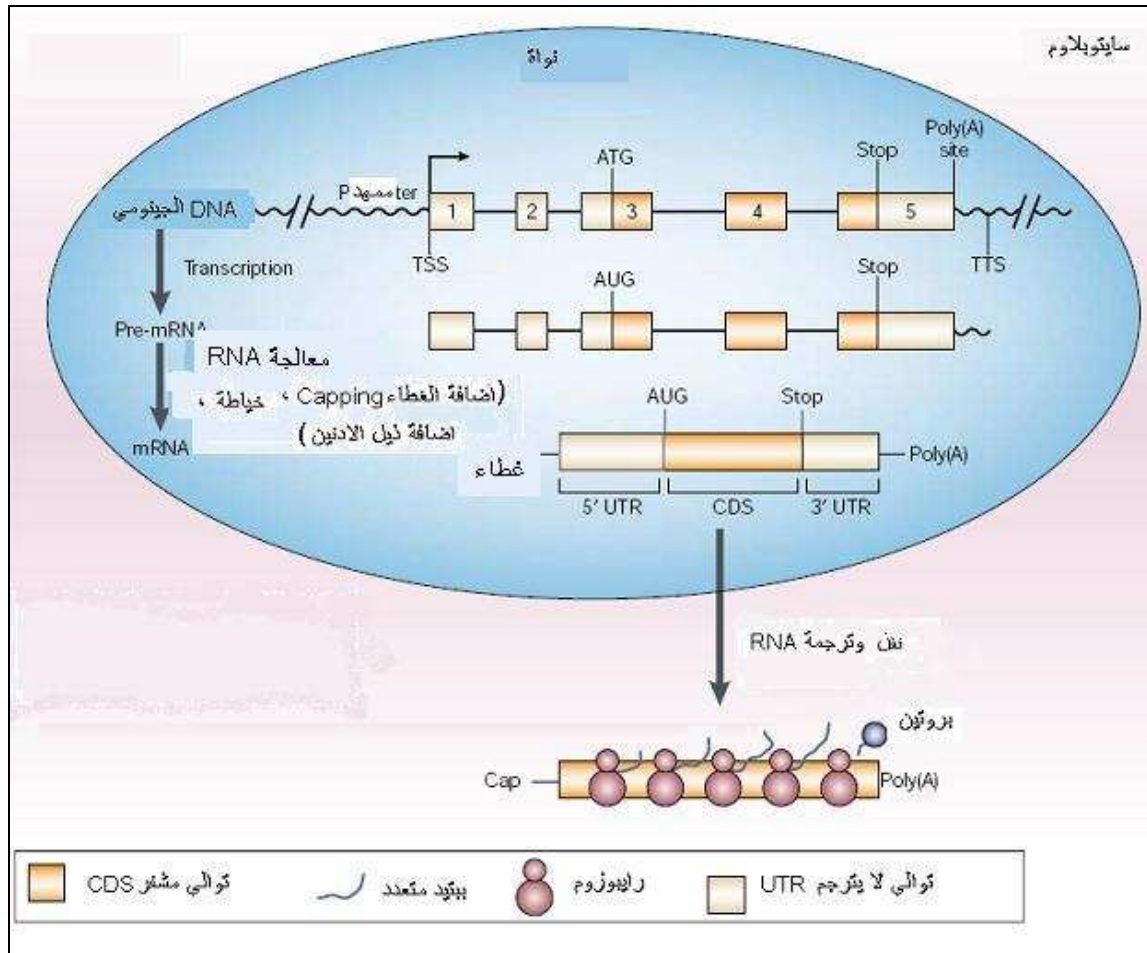
سيأتي ذكرها لاحقاً)، وكذلك تحوي على قطع من الجينات والمتمثلة بأكسون أو أكثر تنتج عن أحداث تأشب غير صحيحة، وهذه تحوي على بعض الإشارات وبعض محتويات الجين الحقيقية وتظهر تماثل جذري للجينات الفاعلة ولكنها لا يتم التعبير عنها لذلك تخطأ البرامج في التعرف عليها. كما ان إجراء عمليات الاصطفاف لإيجاد التشابه الذي قد لا يغطي كل الاكسونات المشفرة وانما يقتصر على الجزء الثابت منها. وبعض الأحيان تعبر عملية التشابه الى الانترونات او الى المناطق غير المشفرة UTRs وعناصر الممهد خاصة في الجينومات القريبة من بعضها جدا من الناحية التطورية، ولذا تصبح عملية التمييز بين المناطق المشفرة وغير المشفرة غير واضحة المعالم. وفي جميع الأحوال فان أهمية وقوة الطرق المعتمدة على التشابه تعتمد على ما متوفر ومتراكم من المعلومات البايولوجية الموجودة في قواعد البيانات والتي يمكن ان تعاق برداءة محتويات هذه القواعد .

تحديد الاكسونات والانترونات

ان طبيعة تركيب جينات حقيقيات النواة المكونة من اكسونات صغيرة وانترونات كبيرة يحتم ان تكون برامج ايجاد الجينات مقسمة الى عمليات ايجاد الاكسونات وايجاد الانترونات. واعتمادا على التوجهات المذكورة أعلاه من حيث استعمال الإشارات ومحتوى الجين تم بناء نموذج جين كامل يرافقه استعمال المعلومات المساعدة المشتقة خاصة من تواليات cDNA وعناصر التنظيم ومواقع الانفلاق وبه يمكن التعرف وتحديد الجينات في اغلب حقيقيات النواة .

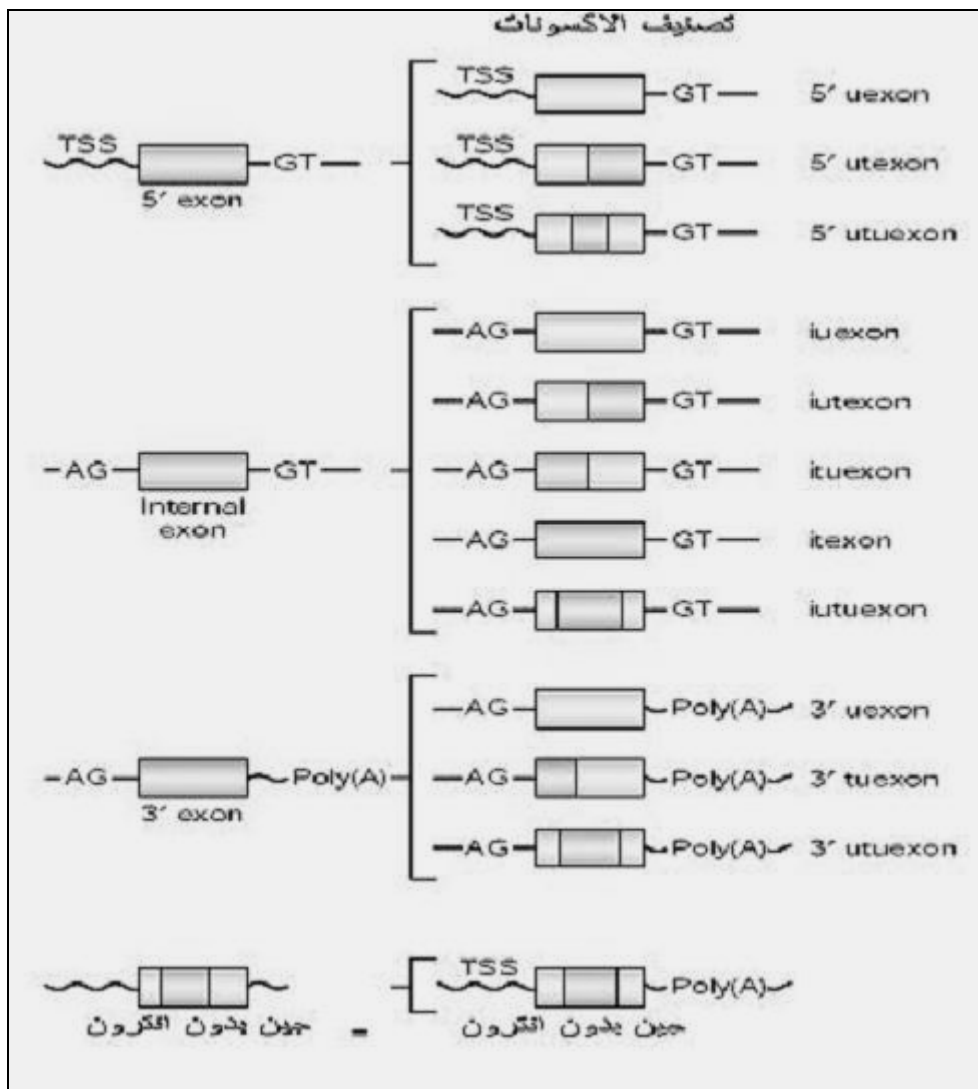
الاكسونات

من الضروري معرفة ان ليس كل الاكسونات سوف تشفر في النهاية الى بروتينات وانما فقط المناطق المشفرة CDS تقوم بهذه المهمة كما موضح في الشكل 71.5 :



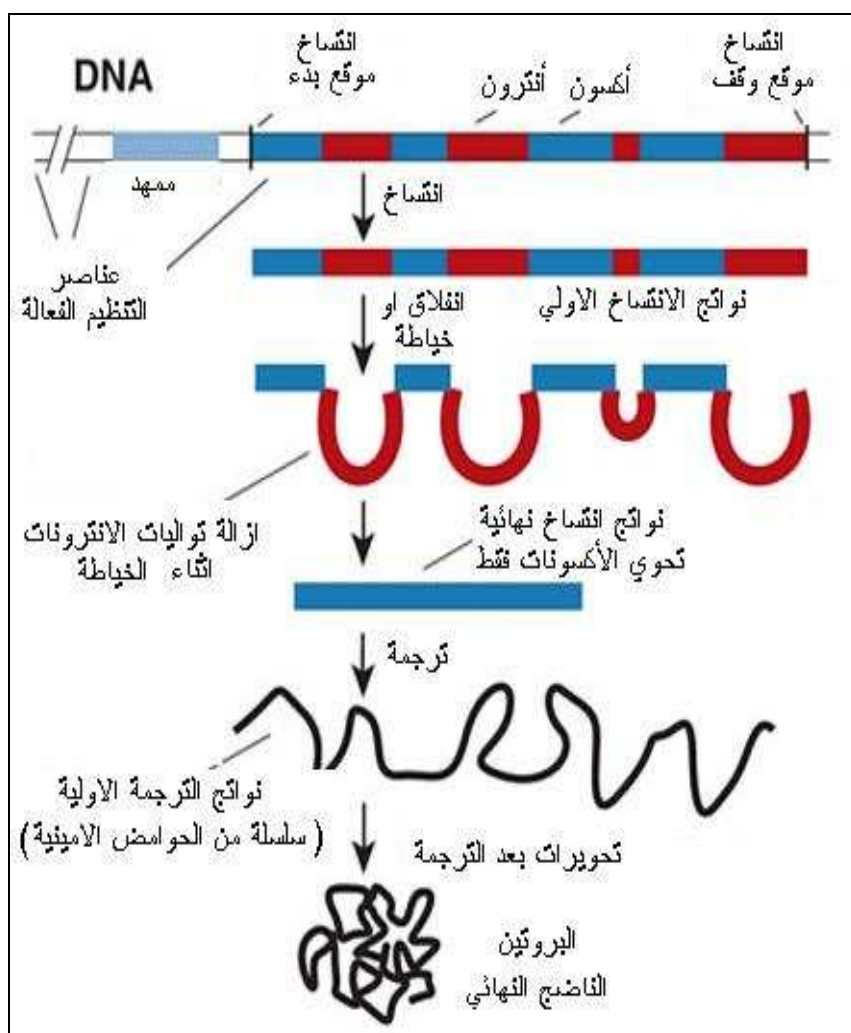
شكل 71.5 : مناطق الاكسون المسؤولة عن التشفير

وهذه الاكسونات تكون منتظمة التركيب ويمكن ان تصنف الى أربعة أصناف هي الاكسونات عند النهاية 5' والاكسونات الداخلية، والاكسونات عند النهاية 3' وأكسونات بدون انترونات وهذه يمكن ان تقسم الى أصناف ثانوية (12 صنف ثانوي) كما موضحة في الشكل 72.5 :



شكل 72.5 : أنصاف الاكسونات الثانوية

ونظرا لكون الفقرات تحوي جينات حاوية على عدة اكسونات فان الاكسونات الداخلية iexon التي تشكل أحد الأنصاف الثانوية هي التي يركز عليها في عمليات التنبؤ وحسب الجينات للابتعاد عن الالتباس الذي قد يحصل مع المناطق غير المشفرة UTRs، وتكون معدلات أطوالها 140 نيوكلوไทيد مقابل الانترونات التي تكون كبيرة ويمكن ان تصل الى 100 كيلوباقعة. ولإيجاد الاكسونات المشفرة يلاحظ ان جزيئات mRNA الأولية pre-mRNA التي تكون حاوية على الاكسونات والانترونات الموضحة في الشكل 73.5 :



شكل 73.5 : الاكسونات في جزيئات Pre-mRNA

تحتوي بعض الإشارات منها الموقع المعطي الذي يكون في الجهة 5' ويسمى 5' splice site (5'ss) الذي يمتاز بوجود AG/GURAGU والموقع المستلم 3' splice site (3'ss) الذي يمتاز بوجود التوالي YYYYYYYYYYNCAG/G فضلا عن وجود موقع CURAY غير الثابت [Y تمثل أي من البريميدينات ، N تمثل أي نيوكلوديد ، R تمثل البيورينات وهذه توجد في جسيمات الخيطة Spliceosome] كل هذه العناصر الوراثية تتجمع لتكون جسيم الخيطة مع مكونات RNA . والإشارات المذكورة تستعمل في تصميم بعض البرامج. والطريقة الثانية المستعملة في حذس الاكسونات هو استعمال مكونات الجين، وتعتمد هذه على مكونات mRNA الذي يبدأ فيها إطار القراءة المفتوح من أول منطقة مشفرة على mRNA الى شفرة الوقف في الإطار نفسه وهذه تعطي تنبؤا جيدا ولكن ليس مضبوطا بشكل متكامل. وقد استعملت نماذج ماركوف للتحري عن الاختلافات بين المناطق المشفرة ومناطق الظل المشفرة Shadow coding sequences (وهي المناطق على الشريط المقابل). ومن الملاحظ

ان الخطأ الذي يمكن ان يحصل في جزيئات mRNA هو احتمال وجود انثرون فيها مما يعطي إشارة وقف. وقد مرت برامج حدس الجينات في حقيقتات النواة بأربعة أجيال او مراحل :

- الجيل الأول اعتمد على تحديد المناطق المشفرة في DNA الجينومي وهذه لا تستطيع التنبؤ بمواقع الاكسونات المضبوطة.
- الجيل الثاني اعتمد الجمع بين إشارات الفلق والمناطق المشفرة للتعرف على الاكسونات وهذه البرامج لا تحاول التعرف على جميع الاكسونات في الجين الكامل.
- الجيل الثالث اعتمد مهام أكثر صعوبة لغرض التعرف على تراكيب الجينات الكاملة وأكثر هذه البرامج ضعيفة الأداء ، اذ تعتمد على إدخال التواليات الكاملة لجين واحد والذي يكون توجهها غير صحيح في العديد من الأحيان.
- الجيل الرابع وهي البرامج التي حاولت التغلب على المشاكل في برامج الجيل الثالث.

طرق الحاسوب المعتمدة

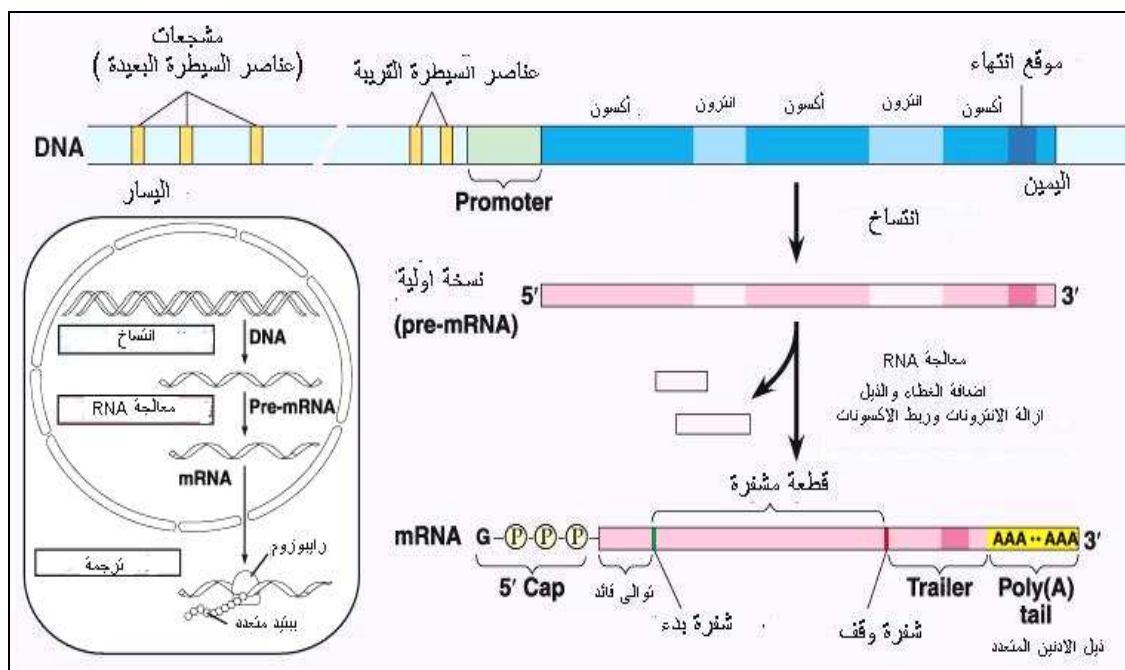
تستعمل الطرق الخارجية (البحث عن التشابه) او طرق البداية او اعتماد نسق التعبير ، ففي الحالة الأولى يتم البحث ليس عن تواليات الجين الكامل وانما على ESTs وكذلك تشابه البروتينات والتشابه يعتمد على عمليات الاصطاف الشامل والموقعي ، ولكن في هذه الحالة يكون الموقعي هو الأكثر استعمالا مثل استعمال BLAST ومجموعته التي تحدد التشابه مع جينات معروفة وبروتينات معروفة و ESTs معروفة. وهذا يعني ان الطرق تعتمد على ان الاكسونات هي المناطق الفعالة في الجين تكون ثابتة على المدى التطوري أكثر من المناطق غير الفعالة مثل الانترونات والمناطق البينية . وعند وجود تشابه بين ESTs او البروتينات فان معلومات التشابه تستعمل لاستنتاج تركيب الجين او وظيفة تلك المنطقة.

والتشابه المعتمد على ESTs له بعض المعوقات وذلك لان هذه المناطق تمثل جزءا صغيرا من تواليات الجين ويتم التعبير عنها تحت ظروف خاصة وتكون قليلة التمثيل في قواعد بيانات ESTs مما يعني صعوبة حدس التركيب الجيني كاملا في المنطقة تحت الدراسة.

وفي حالة استعمال طريقة نسق التعبير مثل استعمال mRNA والبروتينات فان المعلومات التي يتم الحصول عليها تكون قليلة وربما كانت عمليات التنبؤ كاذبة ولذلك فعند الحصول على نتائج تشير الى عدم وجود جين فهذا لا يعني انه لا يوجد جين وانما يمكن ان يعود السبب الى اعتماد مصفوفات تعبير ومصفوفات استبدال خاطئة وقد تكون المسافة التطورية بعيدة جدا. ولكن مع هذا فان التطابقات الجزيئية وان لم ترق الى تسجيل وجود الجين الا انها تعطي إشارة على وجوده، لذلك كان لابد من أقرانها بطرق التماثل الجذري للوصول الى نتائج أفضل.

الخيطة والخيطة البديلة Splicing and alternative splicing

من المعروف ان النسخ النهائية من جزيئات RNA لا تكمل الأشرطة المقابلة لها من DNA بشكل كامل وتجري عليها عمليات ثلاثة منها إضافة الغطاء عند النهاية 5' ومثيلة أول ثمالة من RNA وكذلك فصل الانترونات وربط الاكسونات بعد تكوين جسيم الخيطة من البروتينات وجزيئات RNA ، وعملية الفصل يمكن ان تتم بشكل متغير يؤدي الى إنتاج أكثر من نوع من جزيئات mRNA مما يؤدي الى تنوع كبير في وظائف جينات حقيقيات النواة، والتحويل الآخر هو إضافة ذيل الأدينين الذي يصل الى حوالي 250 ثمالة من الأدينين عند النهاية 3' وهذه الإضافة تكون تحت سيطرة قطيفة تقع الى يمين المنطقة المشفرة ومكونة من CAATAAA(T/C) والخطوات مختصرة في الشكل 74.5 :



شكل 74.5 : خيطة وإنضاج جزيئات mRNA

وأول خطوة في التعرف على جينات حقيقيات النواة هو التعرف على الاكسونات والانترونات ومواقع الانفلاق وهذه تكون من وجهة نظر الحاسبة صعبة ومسألة تحدي بسبب تركيب الجين المنشطر والانفلاق والخيطة البديلة وقلة الكثافة الجينية، والبحث عن الجين في هذه البيئة يكون بمثابة البحث عن إبرة في كوم التبن، والإبرة هذه مجزئة وموزعة في أماكن مختلفة لذلك كانت المهمة هو البحث عن أجزاء الإبرة وتجميعها بالترتيب الصحيح. وما يساعد في عملية التنبؤ والبحث عن هذه الجينات هو وجود مناطق ثابتة منها حدود الانترونات-الاكسونات وفق قاعدة GT-AG التي تكون بالنسبة للانترون عند النهاية 5' وتكون القطيفة عند موقع الارتباط هي **GTAAGT** وعند النهاية 3' هي **NCAG** ¹² (Py). ومن الجدير بالذكر ان بعض الاكسونات الباقية والتي تربط في عملية الخيطة وتكون غير مشفرة ولكنها ضمن برنامج

الحاسبة تعد على انها اكسونات عادية. فضلا عن ظهور حالة الخياطة البديلة لذلك فان الاعتماد على تواليات cDNA يمكن ان يؤدي الى فقدان او تحديد غير صحيح لبعض الاكسونات، ولذا كان هناك تهميش للاكسونات غير صحيح وهذه الأخطاء يمكن ان تقع ضمن إحدى المجموعات الآتية:

- الفشل في تحديد الاكسونات الحقيقية اي قلة حساسية الطريقة المستعملة.
- تعريف وحسب بعض الاكسونات الكاذبة وهذا يعني ضعف التخصص في الطريقة المستعملة.
- الخطأ في تحديد الحدود بين الاكسونات والانترونات.
- حصول دمج للاكسونات.

والحالة الأولى تحصل عندما تكون الاكسونات صغيرة جدا فمثلا بعض جينات الإنسان تحوي على اكسونات بطول 9 قاعدة فقط ، وكذلك تحصل للاكسونات الطرفية سواء في النهاية 3' او 5' من الجين التي تكون قريبة جدا من المناطق الطرفية غير القابلة للترجمة من جزيئة mRNA. وهذه قد لا تحدث عند وجود انترونات صغيرة كما في جينات بعض الخمائر، فالانترونات يمكن تمييزها بواسطة عوامل الخياطة التي ترتبط الى نهاياتها. اما الحالة الثالثة فهي تظهر بشكل جلي عند الخياطة البديلة والتي تشير الى وجود مواقع فصل وخياطة مخفية . ويمكن ان تظهر فيها ظاهرة تجاوز الاكسونات Exon skipping التي تم تحديدها بالتحليل الإحصائي لبيانات المصفوفات الدقيقة وتم التأكد منها بتجارب النسخ العكسي Reverse transcriptase PCR . ونظرا لأهمية الخياطة البديلة في تغيير عدد من المفاهيم فبدلا من البحث عن العدد الكلي لجينات الكائن لإيضاح الوظائف والتعقيد في الكائن أصبح الآن التأكيد حول إمكانية إعادة استعمال الجين ليقوم بأداء وظائف مختلفة وبطرق تنظيم جديدة. فالخياطة البديلة تكون طريقة عامة لتنظيم الجينات وقد امتدت دراسة هذه الظاهرة بتوفر الوسائل الجديدة مثل المصفوفات الدقيقة المستعملة للكشف عن الخياطة البديلة فضلا عن البيانات والرؤى الجديدة مثل استعمال بيانات الجينوم المتعددة لدراساتها وكذلك توفر إمكانيات علوم الحاسوب مثل التحليل المعتمد على تحليل الرسوم والمراجع التي كثرت في هذا المجال مما أدى الى زيادة الفعاليات في هذا المجال مثل إقامة ورشات عمل ومؤتمرات. وقد كان استعمال المصفوفات الدقيقة وتحليل ESTs القوة الدافعة لدفع الدراسات الجينومية وتسخيرها لدراسة الخياطة البديلة وكانت هي الرائدة في تطور هذا الفرع من العلوم، ولكن يجب الحذر من استعمال ESTs نظرا لوجود الأخطاء والانحيازات (المر ذكرها) في توضيح النتائج. وقد مكنت المصفوفات الدقيقة من إيجاد أكثر من 800 حالة من الخياطة البديلة لم تكشف مسبقا باستعمال ESTs وأمكن تحديد انحيازات ESTs مثل تلك التي تتحاز الى النهاية 3' من الجين. وبالنتيجة فان تحليل ESTs أشار الى وجود حالات كثيرة من الخياطة البديلة وأوضح ان 74% من جينات الإنسان تحصل فيها خياطة بديلة وبجمعها مع مصادر معلومات أخرى ربما كانت النسبة أعلى. وأظهرت دراسات ESTs ان الجينات التي لا تحصل فيها خياطة بديلة تحصل فيها ظاهرة تجاوز الاكسون وتشكل حوالي 20% من الجينات والمجال مفتوح في هذه المنطقة.

ونتايج الخياطة البديلة تشير ليس فقط الى اختلاف الكميات وانما الى اختلاف التواليات في النسخ الناتجة وهذا هو المعتمد في دراسات المعلوماتية الحيوية والتي تظهر الاختلاف في التعبير الجيني التي تمثل برسوم قياسية. ونقاس الاختلافات للخياطة البديلة في أنسجة مختلفة وتكون الاختلافات في مجالات مختلفة.

Alternative – 5' splicing

Alternative – 3' splicing

Alternative initiation

Alternative termination

Exon skipping

ويمكن عنقدة نمط هذه المعلومات بسهولة.

التنبؤ بالخياطة البديلة باستعمال دراسة الجينوم المقارنة

وجود تواليات جينومية كثيرة فتح المجال لاكتشاف الخياطة البديلة، وقد وجد ان استعمال ESTs والمعلومات حول الجينوم للأحياء المتقاربة مثل الإنسان والفأر يمكن ان توصل الى حدس الخياطة البديلة، ففي هذين الكائنين فان تواليات الجين (EST) وتركيبه تكون كافية لإجراء عمليات الاصطفاف، وهذه مكنت من التعرف على الاكسونات الأصلية في الفيران فضلا عن التعرف على ان نصف اكسونات هذا الكائن تعاني من الخياطة البديلة. واستعمال EST مكن من التعرف على ثلاث ميزات للاكسونات التي عانت من عمليات خياطة وهي :

- وجود تشابه كبير في التواليات التي وصل بعضها الى أكثر من 95% .
- ميل تواليات الاكسونات ان تكون القواعد فيها مكررات للرقم 3 (اي تكون هناك محافظة على قراءة الأطر للبروتينات).
- وجود ثبوت عالي للقواعد او التواليات التي تحيط بالانترونات وهذا يشير الى إمكانية تحديد اكسونات الخياطة البديلة ولكن بوجود خطأ يصل الى 25% وهذا ما أكدته دراسات RT-PCR.

ولذلك فان دراسة مقارنة الجينومات لإيجاد مواقع الخياطة البديلة ويمكن ان تستعمل مع الأحياء التي فيها ESTs غير معروفة.

تنظيم الخياطة البديلة

هناك عدة أوجه لتنظيم هذه العملية ، فهناك بعض المواقع القوية الموثوق بها Authentic sites تقع عند النهاية 5' فضلا عن وجود مواقع خفية عند الموقع 5'، وقد أشارت الدراسات الى وجود عوامل أخرى غير المواقع المذكورة لها علاقة بالخياطة البديلة. وهنا يبرز مفهوم وجود اكسونات أصلية وأكسونات بديلة ، والأخيرة لها مواقع انفلاق ضعيفة مقارنة بالاكسونات الأصلية ولكنها تحتاج قطيقات مشجعة متخصصة لعمليات الفلق، ولذلك كان من الضروري البحث عن قطيقات مشجعة وأخرى مسكتة (كعوامل منظمة) للكشف عن عملية تنظيم الخياطة البديلة. وقد استغل المكوثر السداسي الذي يوجد بكثرة في الاكسونات مقارنة بالانترونات وكذلك استغلال عملية المقارنة لتغايرات النيوكلويتيد المفرد SNPs بين الحيوانات لإيضاح وجود مشجعات للخياطة والتي تكون تحت ضغط انتخابي تطوري خاصة عندما تكون مجاورة لمواقع الفلق او الخياطة. وكذلك اكتشف وجود المكوثرات العشرية Decamers التي تكون في

المسكتات Splicing silencers وتميل الأخيرة للتجمع على شكل عناقيد تصل الى 7 عناقيد مكونة قطيفة تلعب دورا مهما في الخياطة البديلة ، وهذا يشير الى ان إحداث الفلق او الخياطة يمكن ان تكون لها نمطا" خاصة من قطيفات التنظيم وحدوث الطفرات في هذه القطيفات يمكن ان تكون وراء بعض الأمراض في الإنسان. وعلى العموم فان نسخ الخياطة البديلة يمكن ان تتفكك وبالتالي تخفض من التنظيم للبروتين الناتج.

وهناك عددا" من البرامج المستعملة في تحديد مواقع الخياطة البديلة وهي تقع ضمن أحد المجالين اما البحث عن الإشارات او استعمال الاصطفافات اي البحث عن محتويات التواليات. في الحالة الأولى تعدد البرامج على إيجاد الإشارات التي تشير الى المواقع الفعالة ثم البحث عن التطابقات المتوافقة في محتويات التوالي والسماح ببعض التباين وذلك باستعمال مصفوفات الترجيح الموضعية Position weight matrices (PWMs) التي تشير الى احتمالية ظهور قاعدة في كل موقع من الإشارة وهذه تحسب من الاصطفافات المتعددة لتواليات ذات علاقة وظيفية وفيها يتم تجميع الجين بالاعتماد على عدد من الإشارات المكتشفة، ودرجة المحتوى الناتجة تعرف على انها مجموع درجات المحتوى للقطع وبذا فهي لا تعتمد على الصفات العامة للقطع وأكثرها تعتمد على نماذج ماركوف بمراتبها المختلفة ويمكن تحسين أدائها باستعمال الشبكات العصبية ، ولزيادة التحسين تقرر الطرق بتشكيلات مختلفة من مصادر المعلومات لبناء البرامج فضلا عن إمكانية استعمال مصفوفات مختلفة وهذه تستعمل أصلا لتحديد مواقع الانفلاق اي تحديد مواقع الاكسونات او التنبؤ بالجينات ولكنها تستعمل ايضا في حدس الخياطة البديلة. وتكون نتيجة هذه البرامج قائمة من مواقع الانفلاق التي يمكن منها بناء تراكيب جينية مختلفة، ويزداد وقت تنفيذها بزيادة طول التوالي ويمكن تحسين البرامج في هذه الناحية (وقت التنفيذ) بالأخذ بنظر الاعتبار مؤشرات أخرى .

اما الطرق الأخرى وهي المعتمدة على الاكسونات ففيها يتم فصل المناطق المشفرة في الخطوة الأولى لإيجاد أعلى درجات من التطابق مع الجينات الأخرى وإنتاج قيمة تسمى درجة الجين وهي دالة جمع لكل الدرجات للقطع المجمعة ومؤشرات القوة في هذه الطرق تعتمد على ان بعض الصفات تكون متشابهة مثل طول القطع على سبيل المثال. وتعطي بعض البرامج نتائج مصورة، الاحداثي الصادي هو الاكسونات والاحداثي السيني هو درجة التطابق بين الاكسونات. والوقت يكون متناسبا مع مربع عدد القطع المتنبأ بها وقد عولجت مسألة الوقت بوسائل أخرى.

وفي معظم البرامج المستعملة للتوجيهين أعلاه تتم عملية التأكد من التغيرات في عمليات الفلق او غيرها من المؤشرات باستعمال البروتينات المترجمة عن تواليات الاستعلام او تواليات mRNA او ESTs الناتجة عنها ومقارنتها مع بيانات قواعد البيانات التي تكون فيها المعلومات مرتبة اعتمادا على نوع النسيج ويمكن عندها البحث فيها باستعمال كلمات دالة او مفتاحية ملائمة.

برامج التنبؤ بجينات حقيقيات النواة

بعد الاستعراض المذكور أعلاه يلاحظ ان هناك العديد من البرامج الموضوعية للتعرف على جينات حقيقيات النواة والتي يمكن وضعها ضمن ثلاثة مجاميع :

- البرامج المعتمدة على مكونات الجين أي طرق البداية .

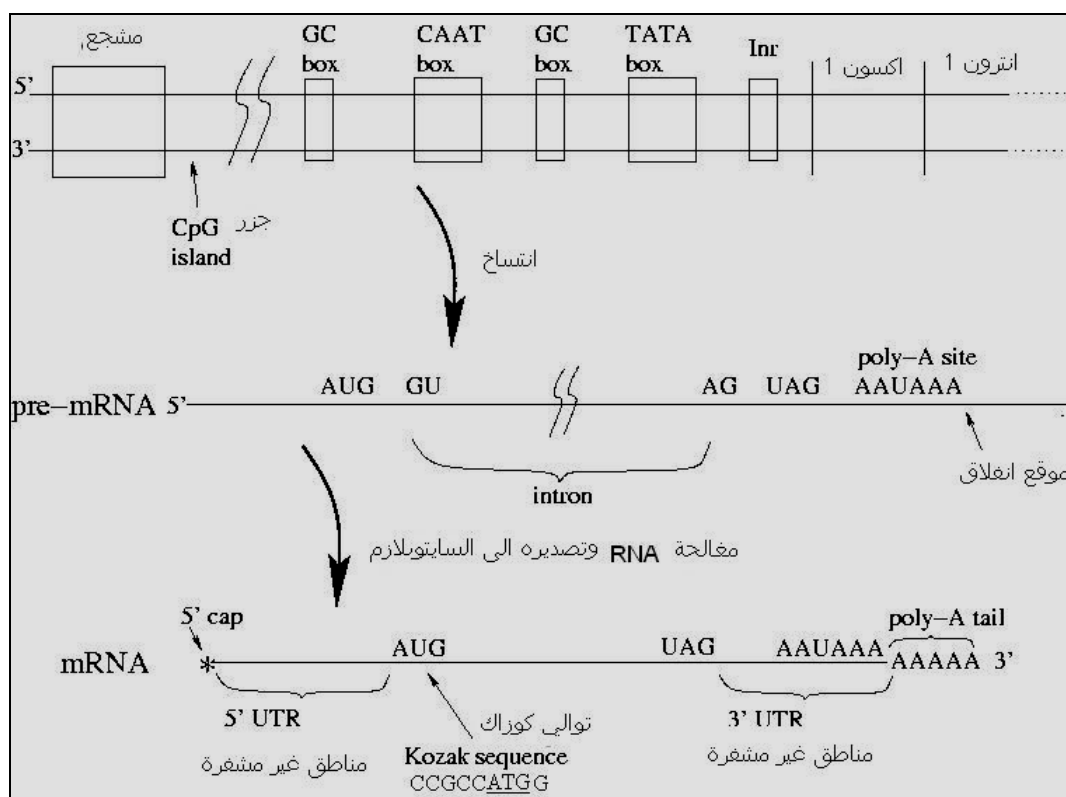
- البرامج المعتمدة على التماثل الجذري

- البرامج المعتمدة على توافق التواليات

وأكثر هذه البرامج تكون خاصة بنوع واحد من الكائنات الحية ويتم تدريبها على بيانات خاصة للحصول على مؤشرات إحصائية خاصة بالكائن. بعض البرامج تكشف عن كل الاكسونات المحتملة وكذلك الاكسونات الثانوية مشيرة الى المعلومات حول الخياطة البديلة ونواتج الانتساخ.

استعمال مكونات الجين في البرامج تهدف الى التمييز بين الاكسونات والتواليات غير المشفرة ، ثم ربط الاكسونات بالترتيب الصحيح والمشكلة في هذه البرامج هو التعرف الصحيح على الاكسونات ولأداء هذه المهمة فالبرنامج يعتمد الى استعمال الإشارات الخاصة بالجين وكذلك محتوى الجين، واستعمال محتوى الجين وإحصائياته Coding statistics التي تشمل توزيع النيوكليوتيدات واستعمال الشفرات المترادفة وتردد المكوثر السداسي والأخير يكون الأكثر أهمية لتمييز المناطق المشفرة وتستعمل معها الشبكات العصبية ونماذج ماركوف الخفية .

وربما كانت هذه أسهل من حيث الدراسة وذلك لوجود إشارات كثيرة يمكن استعمالها، فمعظم جينات الفقرات تستعمل شفرة البدء ATG للترجمة وتحتوي على توالي ثابت يحيط بهذه الشفرة وهو توالي كوزاك CCGCCATG Kozak sequence الموضحة في الشكل 75.5 :



شكل 75.5 : تشرح جين في خلايا حقيقية النواة وموقع توالي كوزاك

وتحتوي الجينات على كثافة عالية من ترددات النيوكليوتيدات الثنائية CG قرب موقع بدء الانتساخ ممثلة بالجزر CpG التي تساعد في التعرف على جينات حقيقيات النواة، وفضلا عن ذلك فان ذيول الأدنين تساعد في التعرف على نهاية التوالي المشفر. ويمكن استعمال بعض الأنماط الإحصائية المستعملة مع الخلايا بدائية النواة والتي ذكر بعضها آنفا حول محتوى النيوكليوتيدات وتوزيعها وتردد بعض المكوثرات ، ولكن تبقى عقبة هذه الطريقة ان المناطق المشفرة لا توجد فيها قطيقات ثابتة يمكن الركون اليها وانما يجب الاعتماد على تفاصيل الجين الدقيقة التي يكون من الصعب تحديدها.

وفي حالة البرامج المعتمدة على التماثل الجذري فانها تعتمد على حقيقة ان تركيب الاكسون وتواليه تكون ثابتة بين الأنواع ذات العلاقة، وتترجم الأطر المشفرة في توالي الاستعلام الى بروتينات وتستعمل للاصطفاف مع البروتينات القريبة منها الموجودة في قواعد البيانات الملائمة. وقد أكدت النتائج المختبرية صحة نتائج هذه الطريقة الى حد ما لذلك أصبحت هي المستعملة في إيجاد الجينات في DNA غير المعروف . ولكن المعوقات في هذا التوجه هو اعتماده على وجود المشابهات في قواعد البيانات وعند عدم وجود المشابه في القاعدة فان الطريقة لا يمكن استعمالها وبذا فان الجينات الجديدة في الأنواع المكتشفة حديثا لا يمكن ان تكتشف بدون وجود نظائرها في قواعد البيانات. وتوجد عدة برامج تستعمل هذا التوجه مع قواعد بيانات محددة.

والتوجه الثالث هو المعتمد على التوافق وهو التوجه الذي يحاول التقليل من اختلاف البرامج في مستوى حساسيتها وتخصصها وذلك بجمع نتائج البرامج المتعددة اعتمادا على التوافق، وهذه أدت الى نشوء وتطور برامج معتمدة على التوافق من قبل معظم البرامج وإزالة التنبؤ غير المتلائم مع الأغلبية، وهذا قد يزيد من التخصصية وبذا تقل النتائج الموجبة الكاذبة. وفي الوقت الحاضر توجد برامج مبنية على نتائج أكثر من برنامج الذي بدوره قد يكون اعتمد على أكثر من مؤشر للتنبؤ.

وعلى العموم هناك عدد من الخوارزميات التي يمكن ان تستعمل للتنبؤ بجينات حقيقيات النواة منها البرمجة الدائيميكية والتحليل الخطي للتمييز وطرق Linguistic methods ونماذج ماركوف والشبكة العصبية (المر ذكرها). وقد استعملت في بعض البرامج وتم تطوير بعضها لتكون خاصة بالكائنات الحية. وأكثر البرامج نجاحا هي المعتمدة على نماذج ماركوف الخفية التي تعتمد على تحديد الاحتمالية، نظرا لان طول الاكسونات والانترونات تحدد بعدد من العوامل ذات العلاقة بعمليات خياطة الجزيئات الأولية pre-mRNA والتي لا تظهر توزيع هندسي Geometric distribution ، لذلك احتاجت هذه الخوارزميات الى التحديد الدقيق لكل من الاكسونات والانترونات في الجينات الحقيقية ولهذا الغرض طورت نماذج ماركوف عامة . وأكثر محددات استعمال نماذج ماركوف هو قلة المعلومات حول تركيب الجين خاصة تواليات الجينات الجديدة، فضلا عن ان الجينات المعروفة في الوقت الحاضر لا تمثل كل الجينات، بالإضافة الى ان هذه الخوارزميات تحتاج الى عدد من التواليات للتدريب والفحص وتقييم الأداء.

ومن الطرق الأخرى هي الطرق اللغوية المذكورة أعلاه، وهذه تستعمل قياس المكوثرات السداسية المشفرة فبعض البرامج تفسر التواليات على انها تتابعات من الكلمات "Words" ،

والنيوكليوتيدات الموجودة لا علاقة لها مع بعضها ولكن يمكن ان تجمع على انها كلمة بطول K (K-tuples) اي ان المكوثر السداسي هو 6-tuples وهذا المكوثر وتردده يستعمل في التمييز بين المناطق المشفرة وغير المشفرة فمثلا قيمة S التي تشير الى تردد المكوثر يمكن حسابها كالآتي:

$$S \rightarrow \text{Hexamer word (CAGCAG)}$$

للكلمة هي : S وتكون قيمة الدرجة

$$S(w) = \text{Log (frequency of } w)$$

ونظرا لكون هذا المكوثر السداسي أكثر ترددا في الاكسونات لذلك تكون درجته في الاكسون أعلى من الممهديات التي تكون فيها تردد TATAAA أعلى.

والنوع الآخر من الخوارزميات المستعملة هي الشبكة العصبية الصناعية التي تمثل مجمع من النماذج الرياضية التي تحاكي بعض الصفات الملاحظة للجهاز العصبي البايولوجي وتقوم باستنتاج التشابه بعد التدريب. والذي يميزها هو ان التركيبية الجديدة لمعالجة المعلومات تتكون من عناصر متداخلة مع بعضها التي تكون مشابهة للوحدات العصبية وترتبط مع بعضها بارتباطات تشابه التشابكات Synapses في الجهاز العصبي، وعند تدريبها للمرة الأولى على اكسون معروف او توالي انترون تصبح قادرة على حدس الاكسون او الانترون في توالي الاستعلام أوتوماتيكيا. (يراجع الفصل الرابع – الشبكات العصبية)

تقييم برامج حدس جينات حقيقيات النواة

نظرا لتعقيد جينات حقيقيات النواة وصعوبة التنبؤ بها كان من الواجب تحديد كل من الحساسية والتخصصية للبرامج المستعملة سواء على مستوى النيوكليوتيدات او الاكسونات او الجينات الكاملة وإنتاج البروتينات . فالحساسية على مستوى الاكسون والجين تتناسب مع نسبة حدس الاكسون الصحيح او الجين من بين الاكسونات والجينات، اما التخصصية على مستوى الاكسون والجين فهي تمثل نسبة حدسها بشكل صحيح من بين كل عملية حدس او تنبؤ تجري. فبالنسبة للاكسونات يستعمل المعدل Average للحساسية والتخصصية بدلا من استعمال معامل الارتباط ومعامل الارتباط التقريبي (ACC) ، ويجب الإشارة الى نسبة الجينات والاكسونات المنسية وكذلك الجينات المتنبأ بها بشكل خاطئ والتي لا تتداخل مع الاكسونات والجينات الحقيقية وباستعمال هذه المؤشرات يصبح التقييم أكثر صرامة والجدول التالي (6.5) يوضح بعض هذه المؤشرات المسجلة للبرامج التي هي قيد الاستعمال.

جدول 6.5 : مقارنة دقة بعض برامج طريقة البداية في حدس الجينات على مستوى النيوكليوتيدات والاكسونات (المسجلة في بعض الدراسات).

البرنامج	مستوى النيوكليوتيدات				مستوى الاكسونات			
	الحساسية	التخصصية	معامل الارتباط	الحساسية	التخصصية	الحساسية+ التخصصية 2	الاكسونات المفقودة	الاكسونات المشخصة خطأ
FGENES	0.68	0.88	0.83	0.67	0.67	0.67	0.12	0.09
Gene Mark	0.87	0.89	0.83	0.53	0.54	0.54	0.13	0.11
Genie	0.91	0.9	0.88	0.71	0.7	0.71	0.19	0.11
GenScan	0.95	0.9	0.91	0.7	0.7	0.7	0.08	0.09
HMM gene	0.93	0.93	0.91	0.76	0.77	0.76	0.12	0.07

0.28	0.2	0.43	0.41	0.46	0.74	0.74	0.75	Morgan
0.23	0.32	0.59	0.59	0.58	0.66	0.73	0.7	MZEF

وعليه يمكن حساب الحساسية والتخصصية ومعدل الفقد (MR) Missing rate والمعدل للأخطاء Wrong rate (WR) وفق المعادلات الآتية:

$$S_n = CE/AE$$

$$S_p = CE/PE$$

$$MR = ME/AE$$

$$WR = WE/PE$$

AE الاكسونات الحقيقية

PE الاكسونات المحدوسة

CE الاكسونات الصحيحة

WE الاكسونات الخطأ

ME الاكسونات المفقودة

اما على مستوى البروتينات فيمكن قياس الدقة بمقارنة نواتج البروتينات المشفرة بالجينات الحقيقية في توالي الفحص بالبروتينات الناتجة من الجين المحدوس. والحالة الأخيرة غير مستعملة على نطاق واسع . فمثلا بالنسبة للاكسون حتى يتم التعرف عليه بشكل صحيح يجب ان تكون كل النيوكلويدات الداخلة في تركيبه معرفة بشكل صحيح، في حين ان الجين وعندما يراد التعرف عليه بشكل صحيح لابد ان تكون مكوناته من الاكسونات والنيوكلويدات معدة بشكل صحيح وخطأ واحد في نيوكلويت واحد يؤدي الى إبطال كل عملية التعرف على الجين، لذلك فان قيم الدقة المسجلة لمستويات الاكسونات والجينات هي اقل مما للنيوكلويدات.

وعند طرح اي برنامج جديد للتعرف على الجينات يتم ذكر مستوى الدقة ولكن مع هذا يجب التعامل معه بحذر شديد لان الدقة تحدد بشكل أساسي بمجموعة البيانات التي قد تكون وضعت بالشكل الأمثل لتلائم البرنامج، والبيانات المزودة عادة مع البرامج تكون بشكل أساسي من تواليات جينومية قصيرة مع تركيب جين بسيط ولذلك يلاحظ عند استعمال البرنامج للتعرف على جين غير معروف من خلايا حقيقيات النواة تقل الدقة . وبهذه الحال يمكن القول ان البرامج الموجودة — على الأقل في الوقت الحاضر — لا تزال بعيدة عما يرجى منها ولا يوجد برنامج يمكن ان يؤدي المهمة بدقة عالية، فكل برنامج يمكن ان يكون حساس لصفة او ميزة معينة والأغلبية العظمى منها تعاني من ظهور النتائج الموجبة الكاذبة وخاصة مع البرامج ضمن مجموعة طرق البداية . فبالنسبة لجينوم الإنسان فان أكثر البرامج استعمالا وشعبية لا يمكن ان تحدد أكثر من 40% بشكل عالي الدقة، ولكن باستعمال برامج عدة واستنتاج المتوافق منها يمكن ان يؤدي الى تحسين الأداء الى حد ما. ولذلك كانت البرامج الخاصة بجينات حقيقيات النواة بحاجة الى تطوير تقنيات حاسوب معقدة وبحاجة الى فهم أوضح وأعمق لعمليات الخياطة لغرض التوصل الى برامج ملائمة. ومقابل هذا فان البرامج التي تبحث في جينومات بدائيات النواة تكون الأمور أسهل نظرا لتواصل جيناتها والكثافة الجينية العالية ولذا فهناك العديد من البرامج المعتمدة على نماذج ماركوف وصلت الى دقة مقبولة جدا.

مشاكل ومثالب حدس الجينات

كان الهدف من بزوغ المعلوماتية الحيوية هو التعامل مع الكم الهائل من البيانات ولكن قد يكون هذا البزوغ لم ينضج بعد وهناك حاجات وجوانب فتحت للاجتهاد في هذا المجال، ويمكن ان يكون اغلبها في حدس والتعرف على الجينات وفي هذا الجزء سيتم تناول بعض هذه الجوانب، مع العلم ان هناك جوانب فرعية قد تم مناقشتها في مواقعها ولكن بقيت بعض الجوانب تحتاج الى تغطية.

تحديد دقة الجينات انخفضت بزيادة عدد تواليات DNA المسجلة وجزءاً منها قد يعود الى قلة الكثافة الجينية او وجود الانترونات او غيرها من المشاكل. ولذلك يلاحظ ان برامج التعرف على الجينات أصبحت بعيدة عن الكمال، فمثلاً في عام 2002 كانت البرامج الخاصة بالتعرف على النيوكليوتيدات تصل دقتها الى 80% اما الاكسونات فتصل الى 45% وعلى مستوى الجينوم الكلي 20% وهذا كان السبب في عدم إمكانية تحديد عدد الجينات في الجينوم البشري التي تعطى له قيم أكثر من موسعة تصل بين 30 ألف - 100 ألف. ويمكن ان تعزى الجوانب المحبطة في هذا المجال الى أكثر من جانب ومنها :

الجينات الكاذبة Pseudogenes

هي الجينات غير الفعالة المقابلة للجينات المعروفة وفقدت قابليتها للتشفير للبروتينات، والبعض منها يحوي على الممهّدات وعناصر الجين الأخرى، استنسخت بالنسخ العكسي من mRNA وعادت لتندمج في الجينوم ولذا تمتلك كل مقومات الجين من جزر CpG ومواقع الانفلاق ولكنها غير فعالة وذلك يعود الى عدة أسباب وراثية مثل:

- عدم وجود شفرات وقف
- حصول إزاحة في اطر القراءة
- نقص في مواقع بدء الانتساخ

وتسمى في بعض الأحيان بالـ DNA السقط (Junk DNA) ولكنها تحوي على معلومات بايولوجية مهمة لتأريخ التطور لأنها تشارك الجينات الفعالة بالحدود وانفصلت عن الجينات الفعالة قبل ملايين السنين، ويكون من الصعب تحديدها في الجينوم.

صفات الجينات الكاذبة

هناك صفات خاصة بهذه الجينات منها :

- تشارك الجينات الفعالة في التماثل الجذري حيث تظهر تشابهاً في التواليات وقد يصل التشابه الى 40% ، ويصل بعض الأحيان قريبا من 100% مما يشير الى ان الجين الكاذب والجين الفعال المقابل له قد انشقا او انفرجا من توالي أصل مشترك لذلك فهي متماثلة ومن المستبعد ان تكون هذه التواليات قد نشأت بشكل مستقل.
- الوظيفة والتي تمثل مرور التوالي بعدد من الخطوات لإنتاج البروتين وإذا فشل التوالي في أحد هذه الخطوات (المعروفة في التعبير الجيني) فهو يعد غير فعالاً.
- وعند استعمال الطرق السريعة للتعرف على الجينات الكاذبة يكون البحث مركزاً لتشخيص ما أقعده عن العمل وهي شفرة البدء وإزاحة الأطر التي تكون عادة شائعة في منع عملية الترجمة الى بروتينات فعالة. وفي هذا المجال يلاحظ ان الكشف عن الجينات الكاذبة الخاصة بـ

rRNA سهلة الكشف اذ ان هذه الجينات توجد بعدد محدد من النسخ ويتم الكشف عنها بإيجاد تشابه التواليات ومواقعها في المناطق الخاصة وبعيدا عن متطلبات الترجمة .

أنواع وأصول الجينات الكاذبة

هناك ثلاث أنواع رئيسة من الجينات الكاذبة كل منها له أصل مميز وصفات محددة منها:

الجينات الكاذبة المعالجة وتسمى ايضا Retrotransposed pseudogenes ، وهذه شائعة اعتمادا على حقيقة ان القفز الارتدادي عملية شائعة في جينومات حقيقيات النواة الراقية ويكون تأثيره كبير في إعادة ترتيب ومكونات الجينوم، فمثلا في جينوم الإنسان هناك 30-40% مكون من عناصر مكررة مثل SINES و LINEs. وعند حصول عملية القفز الارتدادي فان جزءا من mRNA لنسخ الجينات يعاد انتساخه عكسيا بصورة تلقائية ويعود للاندماج مع الكروموسومات، وعند إحام الجينات الكاذبة ثانياة فهذا يعني انها تحوي على ذيل الأدينين ولكن بدون انترونات وهذه هي الصفة المميزة للـ cDNA ولان هذه الجينات آتية من نسخ mRNA فهذا يعني خلوها من الممهدات الطبيعية اي يوجد فيها بتر في النهاية 5' مقارنة بالتوالي الأبوي ولذلك يقال عنها انها ميتة قبل ولادتها، وتصبح بمثابة جينات كاذبة بمجرد انتهاء عملية القفز الارتدادي، ولكن البعض منها قد يشارك في اكسونات الجينات الموجودة من خلال الخياطة البديلة.

الجينات الكاذبة غير المعالجة وتسمى ايضا بالجينات الكاذبة المتضاعفة. فمن المعروف ان تضاعف الجينات هي إحدى الوسائل المهمة في تطور الجينات اي إيجاد حالة Paralogy ، وفي هذه الحالة يمكن ان تحصل للجين الفعال الناتج من التضاعف طفرة تجعله غير فعال وبذلك تكون الجينات الكاذبة في هذا الصنف تمتلك تركيب الجين الفعال من انترونات وأكسونات وممهدات وفقدان الوظيفة في هذه الحالة يكون له تأثير قليل في المضيف نظرا لوجود النسخة الأصلية.

الجينات الكاذبة المعطلة Disabled pseudogenes او ما تسمى بالجينات الكاذبة الوحودية Unitary pseudogenes. فالمعروف ان هناك عددا من الطفرات التي تؤدي الى تعطيل الجين وخاصة عندما يثبت في المجتمع ويكون ذلك بعدة طرق منها:

- الانجراف الوراثي Genetic drift
- اختناق المجتمع Population bottleneck او الاختناق الوراثي (ويحصل في الأحياء وحيدات الخلية بشكل خاص)
- الانتخاب الطبيعي

ان وجود الجينات الكاذبة يعرقل عمل البايولوجيين خاصة عند إجراء عمليات تضخيم الجينات خارج الأنظمة الحية (PCR) اذ يمكن ان يتضخم جين كاذب يشارك الجين المطلوب بالتواليات نفسها وهذا ما يطلق عليه انحراف عمليات التضخم ، وكذلك الحال عند تهميش التواليات والإعلان عن انها من تواليات الجينوم.

وفي مجال حدس الجينات فان الجينات الكاذبة المعالجة المذكورة أعلاه تخرب عمل برامج التعرف على الجينات وذلك بالتشخيص الخاطئ للجينات الحقيقية او الاكسونات، وان كانت في

أحيان خاصة تساعد في التعرف على الجين. وقد وجد بالتجربة ان الجينات الأبوية التي تعطي جينات كاذبة معالجة تفقد وظيفتها أسرع من تلك الجينات التي تعطي جينات كاذبة غير معالجة.

الجينات الكاذبة الفعالة

من التعريف المذكور في مستهل الموضوع ان الجينات الكاذبة فاقدة للوظيفة ولكن وجد في بعض الأنسجة المتخصصة ان هذه الجينات يمكن ان تكون فعالة في بعض الأحيان اما باستعمال مهادتها الخاصة او استعمال مهادت الجينات القريبة منها، وقد وجد ان هذه الجينات يمكن ان تعمل في مجال تنظيم التعبير الجيني كما هو الحال في اشتقاق بعض siRNA من بعض الجينات الكاذبة اي انها تأخذ دورا في تنظيم الجينات المشفرة للبروتينات وكذلك وجد ان لها دور في بعض الأمراض ومنها السرطان.

قواعد البيانات

تشكل قواعد البيانات المستعملة جزءا مهما من إخفاق عمليات حدس وتعريف الجينات. ففي جميع الطرق والحالات تكون الحاجة ماسة الى قواعد بيانات نظيفة ومشذبة ويقصد بها أنها تحوي على ضروريات معقولة وتهميشات ذات علاقة وتعطي روابط ملائمة لمصادر معلومات إضافية ذات علاقة (وقد مر مناقشة هذا الموضوع ضمن قواعد البيانات) وبدون هذه المواصفات فان هناك الكثير من الأخطاء التي تحدث والتي تتضخم باستمرار أثناء الاستعمال. وتختلف الأخطاء وفقا للمستوى المدروس سواء كان توالي نيوكليوتيدات او جينات او غيرها. فالملاحظ ان هناك أخطاء في تهमيش الجينوم البشري الذي اعتمد على قواعد البيانات الخاصة بـ cDNA. كما ان الاعتماد على قواعد البيانات للـ ESTs تكون بعض الأحيان مضللة لانها تنتج من الجينات التي يعبر عنها بمستويات عالية، وفي هذا المجال يمكن ان يكون الرجوع الى قاعدة البيانات Refseq ولكن هذه الأخرى لا تحوي على كل الجينات او التواليات للأحياء جميعا. ولكن يمكن الركون الى البرامج التي تتخصص بكائنات او أنواع معينة حيث تكون فيها التواليات المقارنة مختصرة وربما تكون مشذبة ومرتبطة الى حد ما. ويجب تذكر ان أوثق البرامج يمكن ان تفشل في إعطاء نتائج صحيحة عند استعمال قواعد بيانات لا تحوي على ما يشابه توالي الاستعلام وتكون النتائج مخيبة للآمال بشكل كبير عند استعمال قواعد بيانات تحوي الملايين او أكثر من التواليات، فضلا عن ان الركون الى قاعدة واحدة يمكن ان يؤدي الى أخطاء كثيرة.

الحاجة الى توحيد المفردات والنتائج

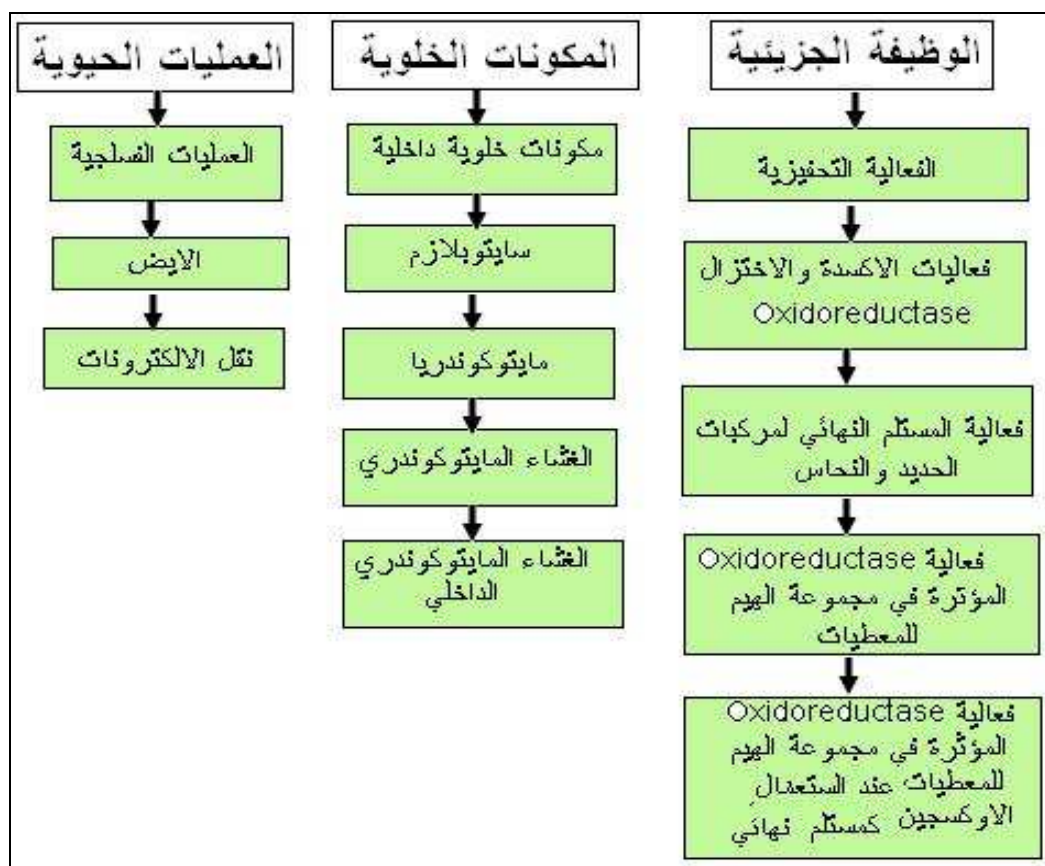
العاملون مع أحياء مختلفة يعطون أسماء مختلفة للجين او البروتين نفسه ومن جهة ثانية يكون المصطلح نفسه المستعمل مع أحياء مختلفة يعني جينات وبروتينات مختلفة ولذا كانت الحاجة ماسة الى توحيد التسميات وهذا ما أدى الى ظهور ما يسمى توحيد تسميات الجين Gene ontology (GO) مثلا. والتوحيد يستعمل عدد محدود من المفردات لوصف وتسمية الجينات ووصف وظائفها الجزيئية والعمليات الحيوية ومكونات الخلية.

وهناك سيطرة للمفردات الموضوعية بحيث ان وظيفة البروتينات ترتبط بوظيفته في الخلية وذلك بوصف هرمي الشكل يعتمد على زيادة التخصصية، فأعلى الهرم يعطي الصورة العامة حول أصناف الوظائف في حين ان المستويات الأسفل تعطي تخصص أكثر للوظائف وبذا تكون

وظائف البروتينات محددة وقياسية وغير غامضة. والوصف في هذا الهرم يعطي ثلاثة نواحي من المعلومات:

- المعلومات الحيوية
- المكون الخلوي
- الوظيفة الجزيئية

ولكل منها مفردات غير متداخلة ، ان تقيس ووضع قياس للأسماء والفعاليات والمسارات الايضية ذات العلاقة كلها هذه تعطي تناسقا" في الوصف العام لوظائف البروتين ويسهل تجميع البروتينات في مجاميع ذات علاقة وظيفية وإعطاء القائمين عليها الإمكانات الكاملة لتهميش كل ناتج عند مستويات مختلفة من الهرم. ولذا فعند البحث في قواعد البيانات باستعمال GO يمكن ان تجلب كل البروتينات ذات العلاقة الوظيفية كما في استعمال القاموس او الموسوعات. وقد أمكن إيجاد وتطوير عددا" من قواعد بيانات GO databases لعدد من الأحياء والنهج يظهر في المثال التالي (76.5) لإنزيم Cytochrome C oxidase.



شكل 76.5 : قواعد بيانات GO

ومن الجوانب الأخرى التي تحتاج الى عمليات توحيد والاتفاق عليها وضع صيغ عامة للملفات (GFF) بعد الاتفاق مع مطوري البرامج وتوحيد المعطيات للبرامج العاملة في حدس الجينات وكذلك المفردات ، وفي الوقت الحاضر هناك عدد قليل من البرامج وتخص بشكل أساسي الجينات المشفرة وأخرى خاصة بجزيئات RNA غير المشفرة ncRNA التي تعطي نتائج بصيغ GFF التي يمكن ان تستعمل كمدخلات وكذلك مخرجات . ان الحصول على معطيات موحدة ستسمح بتطوير وسائل مشتركة لتستعمل في التحليل بعد عملية التعريف بالجين وتقييمها وكذلك وضع المصورات لتمثيلها والربط بين عمليات الحدس والتعريف.

المعلومات البايولوجية والتدخلات مع معوقات التنبؤ

مهما يكن التوجه الذي يستعمله البرنامج فان التعرف على الجينات يجب ان يكون معتمدا على المعلومات البايولوجية خاصة تلك التي تكون على المستوى الجزيئي ولكن مع هذا يمكن ان تظهر بعض حالات الشذوذ. وهناك قضايا يجب ان تؤخذ بنظر الاعتبار منها:

توجد جينات طويلة جدا قد لا تستوعبها البرامج فمثلا الجين Dystrophin gene في الإنسان يحوي على 79 اكسون تشغل منطقة بطول 2.3 ميكاقاعدة، والانترونات فيه طويلة قد تصل الى 100 كيلوقاعدة وهذا الجين 99% منه هو بمثابة انترونات .

كما ان وجود المناطق غير المشفرة UTRs واختلافاتها كل هذه تؤثر عند البحث عن تواليات معينة، فمثلا عند مقارنة 77 من الجينات المتناظرة Orthologous بين الإنسان والفأر وجد ان 56% من 3'UTRs قابلة للاصطفاف بين النوعين . فضلا عن ذلك فان الاكسونات التي هي الهدف في اغلب عمليات البحث تكون في اغلب الأحيان قصيرة ففي رشاد الصخر هناك اكسونات مكونة من 3 قواعد فقط ، وبعض الأحيان قد تكون من قاعدة واحدة ومثل هذه تفقد عند استعمال الطرق المعتمدة على قياسات محتوى الجين خاصة عندما تكون مجاورة لانترونات كبيرة وذلك لان المؤشرات الإحصائية من الصعوبة ان تظهر للتواليات القصيرة، وتتعدد الحالة أكثر ان مكونات الاكسونات هي مضاعفات 3 وبشكل امثل تكون 3، 6، 9 وفقدان هذه التواليات (يعني فقد 3 قواعد منها مثلا) سوف لا يؤثر في تجميع الاكسونات لانها سوف لا تغير إطار القراءة العام مقارنة بفقد واحدة او اثنتين .

ومن المشاكل الأخرى تداخل الجينات Gene overlapping وهي حالة نادرة في جينات حقيقيات النواة ولكنها توجد في بعض الحيوانات والنباتات وتشمل هذه الحالات تداخل المناطق 3'UTRs . وفي هذا المضمار يذكر ان الجينات المتعددة الرسائل كان يعتقد انها محصورة في الخلايا بدائية النواة ولكنها وجدت في حقيقيات النواة كما في snoRNA genes في النباتات وبعض الجينات المشفرة للبروتينات في الديدان وبعض اللبائن ولكن بشكل نادر جدا. الحالة الأخرى حالة إزاحة الإطار وهذا موجود في بعض التواليات المخزونة في قواعد البيانات والتي تحدث أثناء عمليات تحديد التواليات او عند عمليات الطباعة وهذا يؤدي الى تخريب الأطر الطبيعية نتيجة لحذف او إقحام قاعدة واحدة وهذه تعرقل عمليات إيجاد الجينات وحجب الحلول الصحيحة. وان كانت بعض البرامج تعالج هذه المشكلة خاصة في cDNA و ESTs.

وحالة أخرى يمكن ان تلاحظ وهي وجود انترون او أكثر في المناطق غير المشفرة 3'UTRs و 5'UTRs في mRNA وهذه قد تكون أتت من عمليات الخياطة البديلة والبرامج المعتمدة على cDNA و ESTs يمكن ان تميز هذه الانترونات.

ومن المعوقات الأخرى هو وجود مواقع فلق اي فصل الانترونات عن الاكسونات غير صحيحة او غير قانونية ولا توجد برامج خاصة بها ولكن توجد برامج خاصة بالانترونات غير القانونية والتي تستعمل الموقع GT-AG كأساس ويمكن ان تستعمل الواهبات GC بدلا من GT وعندها يمكن الاستعانة بـ ESTs لتصميم البرامج.

والحالة الأخرى وجود البدائل البايولوجية ومنها وجود مناطق بدء انتساخ بديلة او وجود أكثر من مهاد لإعطاء بروتين واحد، ووجود أكثر من مهاد مركزي وذلك لان عدة توليفات من التواليات يمكن ان تؤدي الغرض.

وفي هذا المجال تبرز مسألة الخياطة البديلة التي تم مناقشتها سابقا وتستعمل معها الطرق المعتمدة على ESTs والدراسات الحديثة أشارت الى ان 35-60% من جينات الإنسان لها مواقع خياطة بديلة، ولذلك تحاول بعض البرامج إيجاد الاكسونات الثانوية او تركيب الجينوم الثانوي اي دون التركيب الأمثل. وفي هذه الحالة يمكن استعمال ذيول الأدنين وان كانت توجد أنواع منها بديلة. او استعمال شفرات البدء AUG في mRNA ولكن يجب تذكر ان هذه لا تتماشى مع كل الأحياء فشفرة البدء مثلا في رشاد الصخر هي ACG وكذلك تختلف في بعض بروتينات الإنسان.

وفي اغلب الأحيان لا توجد برامج تتعامل مع الحالات البديلة، وذلك لحقائق موجودة منها عدم وجود جينات قانونية (واضحة وعامة المعالم) ووجود النماذج البديلة للجينات. وبما ان الإعلان وتهميش المعلومات التركيبية للجينوم سبترتب عليها العمل التجريبي (في بعض الأحيان) لذلك كان من الواجب تقييم عمليات الحدس بإعطاء قيم ثقة Confidence value لكل جين، بدءا من الجينات التي تم الحصول على تراكيبيها بشكل كامل من طرق واضحة وغير غامضة الى الجينات ذات القيمة الحدسية الواطئة والمعتمدة أساسا على التوجهات الداخلية أي طرق البداية Ab initio البديلة .

الحلول الآتية

توجد العديد من البرامج وأعدادها تزداد يوما بعد يوم على شبكات الانترنت البعض منها جديد والآخر محدث من برامج سابقة والبعض الآخر طوع للاستخدام مع أحياء خاصة وكثرت إصدارات البرامج وهذا ما يشوش المستخدم الذي يبحث عن برنامج له مصداقية كبيرة لحل المشاكل المذكورة آنفا وغيرها من الأسئلة، والإجابة في هذه الحالة هي "كلا" وقد تبقى الحال كما هي في المستقبل القريب او البعيد نظرا لعدم الواقعية في تصور ان عمليات بايولوجية معقدة مثل الانتساخ والترجمة يمكن ان توضح بالنظر فقط الى تواليات DNA لذا وجب إجراء عمليات التصفية على المعلومات المستوردة من قواعد البيانات وبنوك البيانات ليتسنى استعمال التواليات الموثقة تجريبيا كمراجع. ولذلك كان لابد من جمع المعلومات والشواهد مثل موقع بدء الترجمة، شفرات الوقف، ومواقع الانفلاق وغيرها من الإشارات وكذلك مسألة مقاييس المحتوى قبل وضع نموذج للجين ولكن في هذه الحالة سيزداد عدد الجينات بشكل تزايد. ولكن العدد يمكن ان يقلص بعد وضع بعض الضوابط منها:

- انه لا يوجد تداخل بين الاكسونات
- ان تكون الاكسونات المشفرة متوافقة في نواتجها
- ان يكون دمج اثنين من الاكسونات المتتالية لا يؤدي الى خلق إطار وقف عند منطقة الارتباط

ومن جهة أخرى يجب الأخذ بنظر الاعتبار ان مقارنة المناطق الداخلية للجينات تعطي فكرة عن العوائل متعددة الجينات والتي تمثل ظاهريا نسبة مئوية عالية من الجينات، فهذه المناطق تمثل 80% في نبات رشاد الصخر. وعلى الطرف الثاني يجب الأخذ بنظر الاعتبار المناطق البينية والتي تساعد في تشخيص الجينات المتناظرة حتى عند عدم وجود اي معلومات مسبقة عنها.

تحديد جينات RNA غير المشفرة

معظم البرامج تكون مصممة للتعرف على الجينات المشفرة اي التي تترجم الى بروتينات كما مر سابقا، ولكن هناك عددا من الجينات التي تعمل وتؤدي وظائفها على مستوى RNA اي بعد الانتساخ ومنها جينات rRNA و tRNA فضلا عن العدد الهائل من ncRNA العاملة في مجال التنظيم.

والطريقة المثالية للتعرف على هذه الجينات هو البحث عن التماثل الجذري في قواعد البيانات. ولكن هناك ملاحظة مهمة وهي ان ثبوت التوالي في هذه الحالة ليس مهما جدا وانما المهم هو ثبوت التركيب الثانوي كما هو الحال مع tRNA وعدد من ncRNA التي تعتمد وظيفتها على التركيب الثانوي. لذلك كانت البرامج التي تحدد المناطق المسؤولة عن تكوين التراكيب الثانوية هي الأكثر أهمية والأكثر فائدة، ولأجل هذا وضعت برامج ونماذج إحصائية خاصة للتعرف على نماذج ncRNA التي تحوي على تراكيب ثانوية مختلفة وكذلك البرامج الخاصة بـ tRNA.

ومن الجدير بالذكر فان التعرف على جينات العديد من ncRNA تكون غير معروفة وهذا يشكل مشكلة كبيرة اذ لا توجد خصائص معروفة يمكن البحث عنها فمثلا لا توجد اطر قراءة مفتوحة ORFs ولا انحياز في الشفرات لاستغلال محتوى التوالي في البحث. وفي الجينومات البكتيرية وكذلك حقيقيات النواة البسيطة مثل خميرة الخبز فان وجود مناطق كبيرة بدون اطر قراءة مفتوحة ORFs تشفر لبروتينات فانه يمكن ان يشير الى وجود جينات خاصة بـ ncRNA وذلك يحتاج الى إجراء تحليلات مفصلة لإثبات ان هذه المناطق تنتج نسخ غير قابلة للترجمة ومثل هذه الطرق يمكن ان تستعمل مع جينومات حقيقيات النواة ولكن تكون الطرق صعبة. وفي هذا المجال تتوفر بعض البرامج الملائمة.

تغايرات النيوكلو تيد المفرد SNPs

اغلب البرامج في هذا المجال تركز على كشف التغايرات التي تؤدي الى تغيير كبير في البروتينات الناتجة. وقد مر في فصل سابق التحدث عن تغايرات النيوكلو تيد المفرد ولا بأس من ذكر بعض المعلومات في هذا الموضع. ففي المجاميع البشرية هناك تغايرات اغلبها من النوع الحميد اي لا يؤثر في البروتينات الناتجة عنها كما أشارت العديد من الدراسات التي أجريت على الإنسان والأحياء القريبة منه، ولكن هناك 67,000-200,000 من التغايرات غير

المتبادلة nsSNPs اي التي تؤدي الى تغيرات في توالي الحوامض الامينية وبالتالي تغيير البروتينات و حدوث الأمراض في اغلب الأحيان البعض منها يقع في جينات إصلاح DNA او في المناطق المسؤولة عن الاستماتة او إيقاف دورة الخلية او إزالة السمية، وكل شخص يكون متباين الزيجة لحوالي 24,000-40,000 من nsSNPs . ويكون من الصعوبة تحديد تأثير كل واحدة من هذه التغيرات في وظيفة البروتينات، ونظرا لأهمية تغيير الحوامض الامينية طورت طرق حاسوب تستعمل في الكشف فيما اذا كان استبدال الحامض الأميني Amino acid substitution (AAS) تؤثر في وظيفة البروتين وتبحث فيما اذا كان زيادة او نقصان في وظيفة البروتين والتأثير في التركيب وكذلك ملاحظة اي الدومينات ستؤثر في الإنزيمات متعددة الفعاليات ، والمعروف ان هذه التغيرات تؤثر بشكل كبير في ثبوت البروتين. ومن الملاحظات العامة ان الطفرات او الاستبدالات المسببة للأمراض التي تشكل حوالي نصف الأمراض الوراثية التي تنتج من استبدالات الحوامض الامينية وطريقة دراسة هذه الاستبدالات يمكن ان تميز بنجاح بين الأمراض المنديلية وغيرها من الأمراض التي تحدث في المواضيع التي تكون ثابتة أثناء عمليات التطور مقارنة بتلك المواضيع التي لا تكون ثابتة وهذا يشير الى ان معظم البرامج تعتمد على تشابه التواليات ، وقد لوحظ ان استبدالات الحوامض الامينية المسببة للأمراض لها مظهر تركيبى مشترك يميزها من الاستبدالات المتعادلة او غير الضررة ومن هذا يمكن استعمال التركيب لأغراض الحدى والتنبؤ، والتي يمكن ان تنتبأ بقسوة تغيير النمط المظهري ، والثابت منها في المجتمعات يكون عن طريق الانتخاب يكون نتيجة ان الشخص متباين الزيجة الحامل لها يكون أكثر لياقة من الشخص متجانس الزيجة، فمثلا الطفرات او الاستبدالات الحاصلة في β -globin يكون الأشخاص الحاملين لطفرة استبدال هم أكثر لياقة ولا يصابون بالمalaria مقارنة بالأشخاص الطبيعيين متجانسي الزيجة.

ان وجود عدد كبير من طرق تحديد استبدالات الحوامض الامينية يؤشر الى أهميتها ولذلك تم الاهتمام في محاولات زيادة دقة الحدى وأنشأ لها مزودات خدمة خاصة بها وهي في حالة تطوير نحو الأفضل ، وفضلا عن ذلك أنشأت مزودات خدمة إضافية MetaServers لجمع المعلومات من مزودات الخدمة الأخرى، وقد تم أتمتة بعض طرق الحدى ووسعت ليتمكن استخدامها على مستوى الجينوم الكامل. وعمليات التنبؤ بها والدراسات التي أجريت حولها أشارت الى ان الأنواع المدمرة لها تردد الليلات واطى مما يشير الى حصول عمليات تصفية وانتخاب ، وبهذه الطرق أمكن التعرف على تغيرات نادرة وأخرى للأمراض العامة والشائعة وبذا أمكن تحديد العديد من الجينات المسببة للأمراض.

واسست لها قواعد بيانات خاصة مثل تلك الموجودة في NCBI اذ تحوي قواعد nsSNPs على أعداد كبيرة منها ، ووجد ان 25-30% منها تؤدي الى اضطراب وظائف البروتينات عند دراستها بأكثر من طريقة للتنبؤ بها وهي اقل توزيعا من تلك التي تعد متعادلة من ناحية تغيير الوظيفة.

الطرق الأساسية المستعملة

تعتمد الطرق الخاصة بـ SNPs على عمليات تشابه التواليات وكذلك المعلومات المتوفرة عن التركيب في عمليات التنبؤ والكشف، وفيها تستعمل ما يمكن من مواصفات البروتينات مثل حالة الانتروبي وقابلية ذوبان البروتين Solvent accessibility اذ تتوزع تواليات البروتينات بعد

حصول التغيير فيها على مناطق مختلفة من البروتين اي على السطح او الأخرى المطمورة الى الداخل. وعلى العموم فان الطرق المعتمدة على التركيب تعطي مقادير من التنبؤ اقل من تلك المعتمدة على تشابه التوالي، وتحتاج الى مصفوفات استبدال خاصة وليست تلك المستعملة في تحليل البروتينات العادي مثل BLOSUM62.

إيجاد التراكيب الخاصة في الجينومات البكتيرية

قد مرت بعض جوانب تعامل المعلوماتية الحيوية مع الخلايا بدائية النواة وفي مقدمتها البكتيرية ولكن قد تبقى بعض الجوانب الخاصة بالخلايا بعيدا عن المقارنات واستخدام جينومات هذه الأحياء سلما للوصول الى جينات الخلايا حقيقية النواة، وسيتم في هذا الجزء معالجة بعض الجوانب الإضافية والتركيز على بعض البكتريات التي تعد نماذجاً دراسية.

الصفات العامة للجينومات البكتيرية

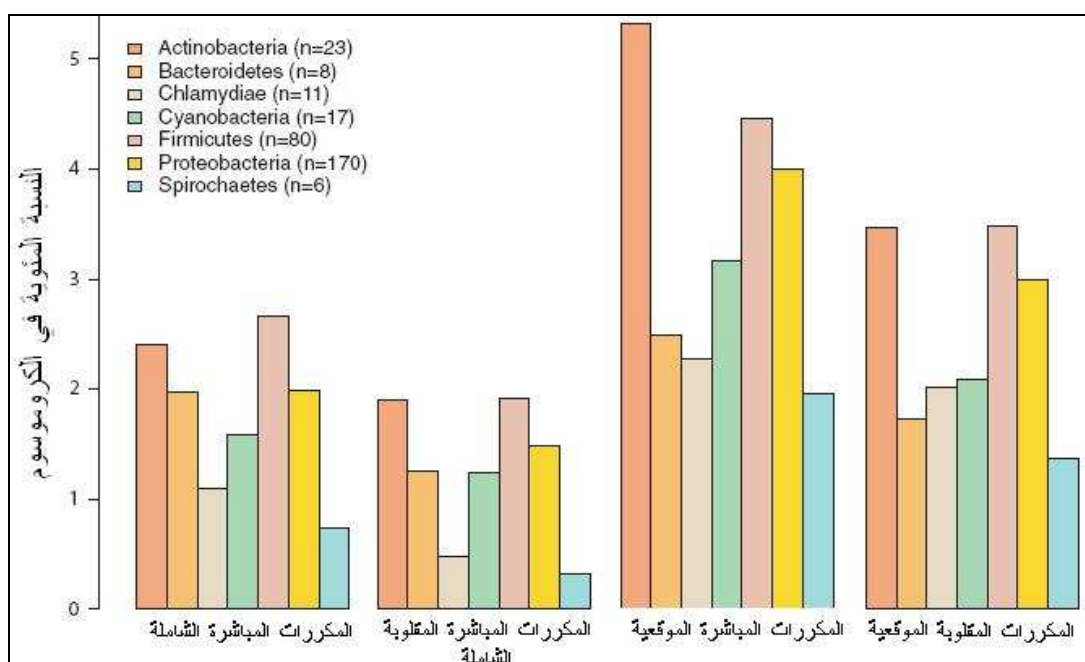
حجوم البكتيريا في العموم قد تكون ليست ذات علاقة مع محتواها من القواعد النتروجينية ولا بالبيئة التي تعيش فيها الأحياء وتكون جينومات البكتيريا والاركية صغيرة نسبيا وتتراوح بين 0.5-1 ميكافاعدة، فضلا عن كون الكثافة الجينية عالية، فأكثر من 85-95% من الجينوم تحوي على تواليات مشفرة ولكن بعض الأحيان تصل الى 50% وهذا يعني ان الكثافة التشفيرية أعلى من حقيقيات النواة. ومن المتوقع ان تكون كثافة التشفير متشابهة للبكتيريا المتشابهة اي بين جينومات أنواع الجنس الواحد. ومقارنة كثافة التشفير قد لا تكون كذلك ولكن هذه قد لا تنتج عن تغاير بايولوجي حقيقي ويمكن ان تنجم عن التذليل والتهميش من قبل المجاميع التي تقوم بعملية تحديد التواليات وإضافة المعلومات، لذا فالنوعية يمكن ان تختلف بشكل كبير مؤدية الى قلة الثقة. فمثلا البكتيريا *Leptospira interrogans* المسببة لمرض Weil الحاسوبية على ضربين تم تحديد توالياتها وهي تحوي على جينومات متشابهة في الحجم الكلي، واثبت مجموعة من العاملين حجم الجينوم لأحد الضروب 4,627,366 قاعدة وكثافة تشفير 74% وعدد من الجينات 3658 من الجينات المشفرة بينما الضرب الثاني فكان الجينوم 4,691,184 قاعدة وبمجموع 4725 جين مشفر اي بفرق حوالي 1000 جين وهذا يفسر ان هناك فرق في الكثافة التشفيرية يصل الى 4%. لذلك كان لابد من تحديد توزيع الجينات المشفرة حسب أطول في الجينوم اذ ان المجموعة الثانية سجلت اطر قراءة مفتوحة قصيرة. وهنا يدخل نوع قاعدة البيانات المستعملة فبعض القواعد لا تحوي على مثل هذه الأطر القصيرة كما ان بعض قواعد البيانات لا تدخل البروتينات الأقل من 100 حامض أميني، وبذا فان الأطر القصيرة يمكن ان تعكس تذييل وتهميش رديء ومثل هذه الأطر القصيرة قد تحدث بالصدفة ولا تعمل كجينات في الخلية، وهي أكثر مما في حالة الأطر الطويلة. لذلك تستعمل معظم البرامج قيمة حد فاصل Cutoff point محدد وتعد الأطر الأقل منه على انها مشكوك فيها، والأطول تحتاج الى المزيد من المعلومات لاعتبارها او لترشيحها كجين وليس نتيجة كاذبة. وبطبيعة الحال هناك عددا من الجينات الصغيرة يتم التعبير عنها في الخلية وهي حقيقية موجودة، لذلك كانت هناك محاولات لتحديد الجينات الصحيحة واستبعاد الكاذبة.

وركزت بعض الجوانب على إيجاد الجينوم الأدنى الذي يمثل الحد الأدنى من الجينات اللازمة لاستمرار حياة البكتيريا حرة المعيشة، فأيجاد الجينوم الأدنى يساعد في فهم الجينات العاملة في مسارات الايض الرئيسة والتي تكون حرجة ومهمة لاستمرار حياة الخلايا وبعض البكتيريا لها

القابلية العالية على حذف الجينات غير الأساسية وهذا يعكس إمكانية اختزال الجينوم Genome reduction وهذا ما يفسر حدوث عمليات التآكل التي تؤدي الى فقدان الجينات وبالتالي – ربما – قلة الكثافة التشفيرية فمثلا في البكتريا المسببة للجذام *Mycobacterium leprae* الذي جينومها 3.3 ميكاقاعدة والبكتريا *Sodalis glossinidius* التي تتطفل في ذباب النوم الذي يكون جينومها بحدود 4.1 ميكاقاعدة فان الكثافة التشفيرية اقل من 50%. وهذا يعني ان التحليل يهدف الى التعرف على الجينات المتناظرة المشتركة بين عدد من الجينومات المتباعدة. والبرامج في هذا المجال تقارن عدد من الجينومات مع استعمال احد الجينومات كمرجع والأخرى بمثابة تواليات استعمال وبمقارنة أزواج من التواليات تجمع المشتركة التي تكون مجموعة مركزية من الجينات.

المكررات في جينومات البكتريا

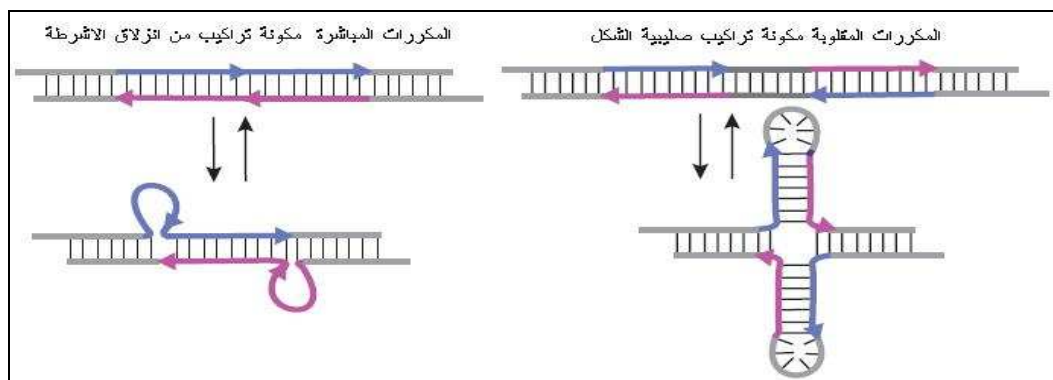
النقطة التي تؤخذ بنظر الاعتبار في جينومات البكتريا انها تحوي القليل من المكررات مقارنة بجينومات حقيقيات النواة واذا وجدت فهي تكون نسخ متعددة لجين معين كما هو الحال مع اوبرونات rRNA، اما النوع الآخر من المكررات فهي قد تكون جينات متناظرة نتجت من تضاعف الجينات بدافع من قوى التطور، فضلا عن ذلك توجد هناك مكررات في المناطق غير المشفرة مثل CRISPRs (Clustered regulatory interspersed short palindromic repeats) وهذه توفر الحماية ضد غزو العاثيات كما ان هناك مكررات من حوالي 8 قواعد نتروجينية تعمل على ثبات الجينوم مثل crossover hotspot instigator (Chi sites). وهناك مكررات أخرى يمكن ملاحظتها في الشكل التالي (77.5) لعدد من البكتريات.



شكل 77.5 : أنواع المكررات في بعض الأجناس البكتيرية

ومن الشكل يلاحظ ان *Actinobacteria* و *Firmicutes* و *Proteobacteria* تحوي النسب الأعلى من المكررات الشاملة والتي تكون غير شائعة في الحلزونية *Spirochaetes* ولا تكون متعلقة بوجود AT على عكس المكررات الموضعية التي لها علاقة وثيقة بتكرار AT وتلعب دورا مهما في تحديد تركيب وثبوت DNA. ومن الجدير بالذكر ان التواليات الشائعة توجد في عدد كبير من البكتريات.

ان بعض التراكييب الخاصة بـ DNA تتكون عند وجود المكررات وخاصة الموضعية المذكورة آنفا، كما هو الحال في تكوين التراكييب الصليبية الشكل وقد تكون متناظرة او تتخذ أشكال أخرى كما موضح في الشكل 78.5 :



شكل 78.5 : التراكييب الناتجة في جينومات بعض البكتريا نظرا لوجود المكررات

والتي تكون تراكييب تشبه ماشة الشعر في جزيئات tRNA وبعض جزيئات rRNA. ومثل هذه التراكييب تساعد في عمليات الانزلاق الموضحة في الشكل ويمكن ان تساعد في إقحام العناصر المتحركة في DNA فضلا عن مشاركتها في عمليات استئصال او مضاعفة الجين، وهذه التواليات مسئولة عن تغاير الأطوار Phase variation التي تلاحظ في بعض البكتريات مثل *Neisseria* وغيرها من الأجناس.

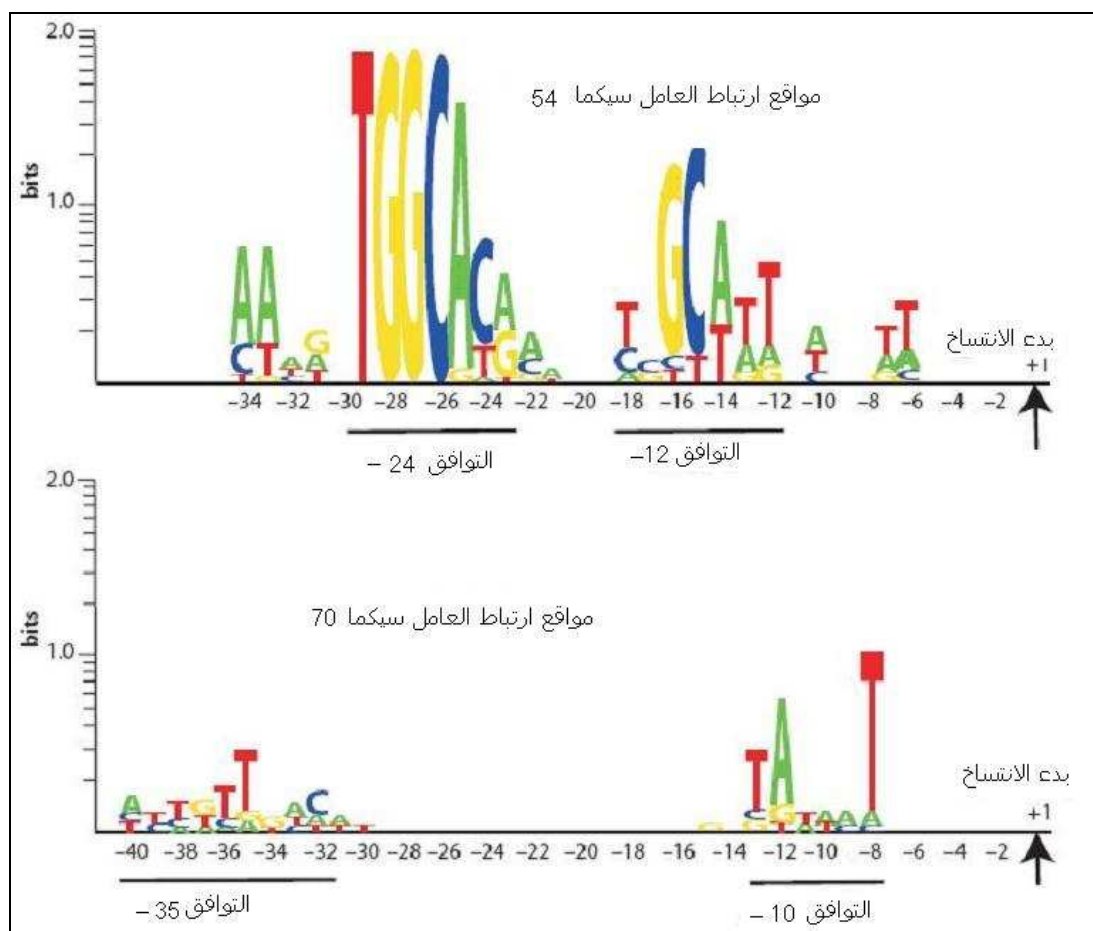
إيجاد جينات البكتريا

لتحديد جينات البكتريا يمكن النظر الى ان كل جين في الخلايا بدائية النواة المتمثلة بالبكتريا يتكون من امتداد مفرد متجاور لأطر القراءة المفتوحة المشفرة لبروتين واحد او RNA بدون وجود قطع في داخل الجين. وتبدأ معظم الجينات بشفرة ATG وفي أحيان قليلة تستعمل GTG او TTG كبديل لشفرة البدء. وبسبب وجود ATG و GTG و TTG في الإطار فان وجود هذه الشفرات لا يعني بالضرورة ان تعطي دليل واضح على بدء الترجمة ، وانما قد تكون ضمن توالي محدد كما في توالي كوزاك في الخلايا حقيقية النواة .

الممهدات في البكتريا

المناطق التي يبدأ عندها الانتساخ بإنزيم RNAP المتعدد الوحدات وما يساعد في ذلك عامل سيكما مثل سيكما 70 وهو بروتين يرتبط بـ RNAP ويساعده في تمييز الجين ويقع الممهد الى يسار الجينات المسئول عنها واختيار نوعية العامل سيكما يقع ضمن المستوى الرابع من

مستويات التنظيم (المارة الذكر سابقا) ويمكن حدس مواقع ارتباط سيكما بدرجات مختلفة من الدقة، وعند ثبوت موقع الارتباط يكون التنبؤ به سهلا كما في حالة العامل سيكما 54 (ويكون الموقع أكثر ثباتا من مواقع ارتباط العامل سيكما 70 كما موضح في الشكل التالي) وتستعمل معها طرق الحدس ببرنامج الشعار Logo كما موضح في الشكل 79.5 :



شكل 79.5 : تحديد مواقع ارتباط العوامل سيكما في ممهدات البكتريا (*E. coli*) باستعمال برنامج الشعار Logo

وموقع الارتباط تمثل القطع -35 ، -10 أي يسار موقع بدء الانتساخ فالعامل سيكما 70 كما موضح في الشكل يرتبط الى التوالي (-35) TTGACA والتوالي -10 TATAAT. وفضلا عن ذلك فان هناك مجموعة من البروتينات التي ترتبط الى DNA وهي عوامل الانتساخ التي تسهل عملية الانتساخ ويطلق عليها عموما عناصر التنظيم التي ترتبط الى تواليات DNA القريبة من الممهد او على بعد مئات من القواعد من الممهد وعند ارتباطه البعيد يمكن ان تؤثر في الممهد نظرا لمرونة تركيب DNA للنفوس وإظهار التأثير وذلك لجلب عوامل الانتساخ ليصبح بتلامس مع معقد RNAP كما مر في الشكل (67.5) .

والعوامل التي تؤثر عن بعد تقع ضمن المستويات 5 و 7 من مستويات التنظيم (المارة الذكر/الفصل الثالث) والتي لا يمكن ان تتم بدون وجود مواقع موضعية Cis-acting elements. ويجب الأخذ بنظر الاعتبار ان عمليات التنبؤ بالممهدات وتحديدتها يجب ان تعتمد على التجارب المختبرية التي تجري عادة خارج الأنظمة الحية او باستعمال عدد قليل من السلالات المنتخبة وهذه تكون قد تطبعت للنمو في الأوساط المختبرية والتي لا يمكن ان تمثل المجاميع البكتيرية الموجودة في العالم الحقيقي او الطبيعة لذلك يمكن الحصول على نتائج متضاربة بين تجارب المختبر والحدس الحاسوبي كما يحصل في تحليل ممهدات جينات الرايبوزومات.

النقل الأفقي للجينات وتأثيره في التنبؤ بالجينات

يعرف النقل الأفقي على انه تبادل المعلومات الوراثية بين الأنواع والذي يحدث بشكل رئيس في البكتريا بعدد من الطرق ، وعمليات النقل قد تكون قديمة او جديدة. فالحديث منها يمكن الكشف عنها بالتحري عن آثار من عملية النقل وذلك بالكشف عن مناطق من تواليات الجينوم غير الطبيعية مقارنة بالمناطق المجاورة. والصفات غير الطبيعية قد تشمل المكونات من النيوكليوتيدات او استعمال الشفرات الخاص او مكونات الحوامض الامينية (بعد ترجمة القطع). او تردد نوع من النيوكليوتيدات، وإحصائيات النيوكليوتيدات غير الطبيعية مقابل مناطق محددة من الجينوم يمكن ان تساعد في تشخيص الجينات الغريبة في الجينوم. ومن أهمها استعمال الانحياز

$$GC\ skew = (G-C)/(G+C)$$

اذ يكون الانحياز نحو G في تواليات DNA القادم وتستعمل بشكل عام دليلا على العناصر الوراثية المكتسبة حديثا وعند تقرير حالة النقل الأفقي يكون للعاملين الرأي في تهميش الجينوم لتحديد فيما اذا كانت هذه الملاحظات تدل على النقل الأفقي للجينات وذلك لان مثل هذا الاستنتاج يحتاج الى عدد من الشواهد.

اما التوجه الثاني في تحديد النقل الأفقي للجينات فهو الدراسات المقارنة اي بتحليل العلاقات التطورية بين الجينومات وهذه يمكن ان تكشف عن عمليات النقل الحديثة والقديمة، فعملية التجمع غير الطبيعية في شجرة العلاقات التطورية تفسر في العادة على انها أحداث نقل جيني أفقي.

جينات RNA في البكتريا

تعد جينات RNA الرايبوزومي (rRNA) هي الأهم في البكتريا لانها تمثل لب ومركز العلاقات التصنيفية في البكتريا. والجينات المشفرة للـ 16S و 23S و 5S توجد في اغلب الأحيان في اوبرون واحد. والبكتريا يمكن ان تحوي على واحد او أكثر من هذه الاوبرونات التي تنتج نسخ متشابهة كثيرة. وتستعمل جينات 16S بشكل خاص في عمليات التصنيف والتي تكون مشابهة لأشجار التصنيف الأخرى. وتحديد تواليات rRNA يجب ان يحظى بتغطية خاصة وشاملة لأهميته في دراسة العلاقات التطورية ، وذلك لان معظم عمليات التذييل والتغطية تعطي اهتماما اكبر للجينات المشفرة للبروتينات ولذلك طورت برامج خاصة بها.

برامج التنبؤ

تمتاز البرامج الخاصة بالبكتيريا لإيجاد الجين في البحث عن الاوبرونات باعتبار ان أكثر جينات هذه الأحياء تتجمع في مجاميع هي الاوبرونات وتعمل تحت ممهد واحد يقع يسار أول جين، لذلك كان إيجاد الاوبرونات في البكتيريا مهما في إيجاد المهدات وعناصر التنظيم الأخرى وهذه كلها متعلقة بالجين الأول وباقي الجينات لا تسبقها مثل هذه التراكيب. ولذا فان طرق إيجاد الاوبرون تعتمد على إيجاد ترتيب الجينات وكذلك المسافة بين اثنين من الجينات والعلاقة الثابتة بين الجينات التي تستنتج من عمليات تحليل الجينوم المقارن. وقد استهدفت البرامج :

1- وضع خرائط الجينومات وقد كان أول جينوم اكتمل عام 1995 للبكتيريا

Haemophilus influenzae, الحرة المعيشة ولم تكن لهذه البكتيريا خرائط وراثية

او فيزيائية. وعند تحديد توالياتها قسم الجينوم الى قطع صغيرة ثم تم تحديد توالياتها

كل على حدة ثم استعملت برامج الحاسوب لتجميع هذه التواليات بترتيب صحيح وذلك

بالبحث عن التداخلات Overlaps .

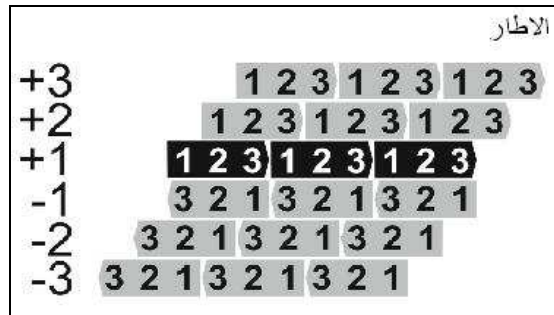
اما الأحياء المدروسة جيدا مثل *E. coli* و *Bacillus subtilis* والتي لها خرائط وراثية وفيزيائية والتي تعد مراجعا لغيرها لم يكن من الضروري استعمال طريقة التشظية وتقسيم الجينوم الى قطع صغيرة حيث يمكن الاستعانة بالخرائط الموجودة ولكن بعد إيجاد الجينات لابد من إيجاد اطر القراءة المفتوحة وتهميشها. اما بالنسبة للبكتيريا *H. influenzae* فقد كانت الحاجة ماسة لتهميش كل التواليات من جديد مقارنة بالبكتيريا *E. coli* و *B. subtilis* التي كانت كميات كبيرة من المعلومات متوفرة.

2- الكشف عن اطر القراءة المفتوحة اي تحديد تركيب الاوبرونات ونظرا لغياب

الانترونات فان جينات البكتيريا تمتلك اطر قراءة مفتوحة طويلة تبدأ بشفرة البدء

المعروفة وتنتهي بشفرات الوقف، ولذلك فان أفضل طريقة لتحديد الأطر الطويلة هو

استعمال طريقة ترجمة الشفرة السداسي لتوالي الاستعلام كما موضح في الشكل 80.5



شكل 80.5 : طريقة ترجمة المكوثر السداسي Hexamer

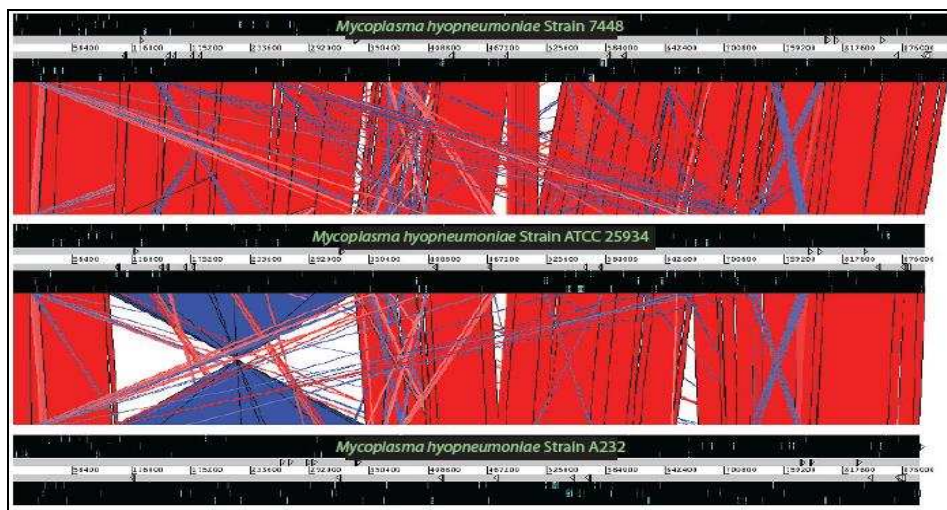
اذ تترجم التواليات بكل الاحتمالات ثلاثة للأمام وثلاثة للخلف وان إطار القراءة الطويل لا يمكن ان يحدث بالصدفة ولذلك عمليا تكون هي جينات . وبعض البرامج تفقد الجينات الصغيرة التي تشفر لـ 100 حامض أميني (اي 300 نيوكلوئيد) . وفي جميع الأحوال وبعد إيجاد اطر القراءة المضمون فيها فان التواليات يتم ترجمتها الى بروتينات وتستعمل التواليات البروتينات في

البحث مقابل قواعد بيانات البروتين وهنا يكون الكشف عن التشابه أقوى دليل على وجود إطار يشفر للبروتين. فضلا عن ذلك فان جينات الظل Shadow genes وهي الأطر المفتوحة القراءة المتداخلة الواقعة على الأشرطة المقابلة في DNA للأشرطة المسوحة يكون من الصعب تحديدها لذلك تستعمل معها نماذج ماركوف الخفية HMM ، وتكون عمليات البحث عن الجينات في البكتريا مشابهة للبحث عن الجينات في المايكوبكتيريا والبلاستيدات الخضراء. وعند البحث عن الجينات يجب الحذر من التغيرات المحتملة التي تؤدي الى زيادة نسبة التنبؤ او فقدانها ومنها ان الجينات قد لا تبدأ بشفرات البدء القياسية ATG فالبعض يستعمل GTG او TTA ، TTA ، ATA ، CTG وللتذكر ان شفرة الميثايونين عندما تكون شفرة بدء فهي تعطي N-formylmethionine اما في المواقع الأخرى فهي تعطي الميثايونين العادي. لذلك فعند وجود شفرة بدء لغير الميثايونين عند النهاية 5' يمكن ان يؤدي الى إيجاد بروتين مبتور. والتعرف على شفرة البدء هناك بعض الخصائص المرتبطة مع عملية الترجمة التي يمكن استعمالها ومنها وجود موقع ارتباط الرايبوزوم rbs والذي يسمى Shine-Dalgarno sequence (SD sequence) يكون غنيا بالبيورينات المكملة لقواعد 16S rRNA على الرايبوزوم وهي تقع مباشرة بعد (Downstream) موقع بداية الانتساخ وأمام او الى يسار شفرة البدء.

وشفرة الوقف TGA هي مثال آخر للمراوغة وعدم ضبط إيجاد الجينات فمثلا في البروتينات الحاوية على السلينيوم فان هذه الشفرة TGA تقع في مكان خاص من الإطار وتحدد اندماج حامض أميني غير طبيعي هو السستئين السيليني وعند عدم تميز هذه الشفرة يكون تحديد الجين مبتورا قرب النهاية 3'.

ونقطة أخرى جديرة بالاهتمام وهو إدخال شفرات الوقف في عمليات التعريف فالأوبرونات في البكتريا تحوي على تواليات لارتباط العامل رو (ρ) وله تركيب ثانوي مميز يكثر فيه الثايمين T التي تكون في نهاية الأوبرون يستعمل في التنبؤ بالجين كإحدى الإشارات المهمة وتسجل درجات جيدة خاصة بالنسبة لتركيب ماشة الشعر والطاقة التي تكون مؤشرات لعملية التنبؤ.

والبرامج الحديثة للتعرف على الجينات البكتيرية تستخدم الألوان والصور لتوضيح صورة الجينوم الذي يبلغ الملايين من القواعد النروجينية وكان من اللازم توضيحه كوحدة واحدة لأنه في اغلب الأحيان يكون هو الوحدة الوراثية الرئيسة في الخلية ولا جدوى لتجزئته كما هو حال الكروموسومات في الخلايا حقيقية النواة. وبعض البرامج لها إمكانية تكبير الصور الملونة والتحكم بها لإظهار منطقة ما كما في الشكل 81:5 :



شكل 81.5 : إحدى طرق إظهار الجينات في بدائية النواة باستعمال احد برامج Artemis

الشكل يوضح اصطفاف تواليات ثلاث جينومات لسلالات البكتريا المسببة لذات الرئة في الخنازير *Mycoplasma hyopneumoniae* ، والملاحظ من الشكل ان المناطق الثابتة تظهر باللون الأحمر. اما شكل الساعة الرملية (المثلثين المتقاطعة) فيشير الى المناطق التي حصل فيها انقلاب اما وجود حالات الإقحام والحذف فتظهر بشكل إسفين او مثلث ضمن المناطق المشفرة. ومثل هذه البرامج تصلح لمقارنة 2 من التواليات وقد تستعمل لصف تواليات متعددة على ان لا تزيد عن خمسة تواليات وذلك لان التمثيل اللوني يصبح صعبا وغير واضح ومثل هذه الرسومات تحتاج الى عمليات تحليل دقيقة لانها لا تميز التغيرات ولا تظهر عدد الفجوات لذلك لا تمكن من اختيار الاصطفاف الأمثل.

اما البلازميدات فلها برامجها الخاصة التي تحدد مواقع القطع بإنزيمات القطع ويوجد العديد منها.

الوراثة اللاجينية Epigenetics

الوراثة الخارجية او اللاجينية تعد الآن في طليعة العلوم البايولوجية الحديثة والتي ساعدت في إعادة النظر في العديد من المفاهيم الوراثية السابقة وقد مر إيضاح الموضوع في الفصل الثالث ، ويحاول الباحثون في الوقت الحاضر دراسة دور DNA و RNA والبروتينات والبيئة في عملية الوراثة وذلك باستعمال الوراثة اللاجينية في فهم التنظيم الوراثي وتمايز الخلايا وعلم الأجنة وحالة الشيخوخة والهرم والأمراض وفي مقدمتها السرطانات.

وتركز البرامج في هذا المضمار على إيجاد جزيئات RNA غير المشفرة ودورها في السيطرة على تنظيم الجينات والتأثير في تركيب الكروماتين والإنزيمات اللازمة في عمليات تحويل جزيئات RNA بعد إنتاجها. وبرامج أخرى تبحث في دور RNA في تعطيل الكروموسوم الجنسي X في إناث اللبائن للسيطرة على الجرعة الجينية المشفرة من قبل هذا الكروموسوم. ويكون ذلك بالاعتماد على قواعد بيانات متخصصة مثل قواعد بيانات خاصة بـ miRNA

التي تستعمل لأغراض مختلفة مثل عمليات طوي هذه الجزيئات والكشف عن فعاليتها وكذلك الكشف عن علاقتها بالجينات كما توجد منها قواعد بيانات متخصصة بالأحياء مثل الإنسان. وتتناول المعلوماتية الحيوية مسألة الوراثة اللاجينية من جوانب أخرى منها البحث عن مثيلة DNA وجزر CpG المتعلقة بها. فهناك العديد من البرامج التي تتناول النواحي الآتية:

- برامج لإيجاد CpG الموجودة في الممهدات.
- برامج لإيجاد المناطق الغنية بـ CpG
- برامج تبحث عن CpG باستعمال مؤشرات خاصة
- برامج لرسم العلاقات للـ CpG

فضلا عن البرامج التي تبحث في مثيلة DNA في حالة السرطانات وتحت الظروف البيئية المختلفة وكذلك برامج لتصميم البوادرى المستعملة في تحليل مثيلة DNA . وكل هذه البرامج تعتمد على قواعد بيانات خاصة منها قواعد بيانات خاصة بمثيلة DNA للأغراض المذكورة.

الوراثة اللاجينية الحاسوبية Computational epigenetics

وهي التسمية الأخرى لمعظم الأغراض المذكورة أعلاه وتعني استعمال طرق المعلوماتية الحيوية لتكميل البحث التجريبي، ونظرا للانتساع الحاصل في الوراثة اللاجينية كان هناك حاجة ماسة لتطوير وتطبيق المعلوماتية الحيوية لحل الأسئلة في الوراثة اللاجينية وإيجاد النماذج النظرية لذلك وهناك العديد من الطرق المتخصصة للتعرف على الوراثة اللاجينية من صفات تواليات الجينوم وهي تخدم أغراض مزدوجة أو أكثر من مزدوجة ومنها:

- عمليات الحدى الصحيحة في مجالات الوراثة اللاجينية يمكن ان تغني عن العمل المختبري الى حد ما خاصة المتعلقة بالاكشافات الحديثة لآليات عمل الوراثة اللاجينية لأنواع الأحياء المختلفة.

ولعل أهم تطبيقات الوراثة اللاجينية كانت في مجال السرطان ، فالاختلال في النواحي اللاجينية فتح مجالات لغرض تجديد وتحسين التشخيص والعلاج لهذا المرض العضال. ففي هذا المجال أمكن الحصول على قائمة بالمناطق الجينومية التي يوجد فيها اختلافات في الوراثة اللاجينية بين الخلايا الورمية والخلايا العادية وحتى بين أنواع المرض الفرعية المختلفة، لذا أسس ما يسمى بالوراثة اللاجينية للسرطان Cancer epigenetics.

وبرزت في مجال الوراثة اللاجينية حقول عدة منها:

- الوراثة اللاجينية وعمليات التنظيم والتي مر ذكر معظمها فيما سبق
- الوراثة اللاجينية للمجتمعات Population epigenetics والتي تهدف الى استخلاص الآليات الملائمة من دمج بيانات الوراثة اللاجينية مع نسق التعبير الجيني وخرائط Haplotype (مجموعة من الجينات التي تستورث سوية) من نماذج كبيرة من المجتمعات متباينة الأصول.
- الوراثة اللاجينية التطورية Evolutionary epigenetic وفيها يتم التعرف على الجينوم اللاجيني للأحياء وإدماجها في دراسات مقارنة للأحياء مختلفة الأصول للتعرف على الحالة التطورية للكائنات.

• ومن ثمرات مثل هذه الدراسات برز مشروع تحديد الجينوم اللاجيني للإنسان Human epigenome project.

• الوراثة اللاجينية الطبية وتشمل البحث عن آليات الوراثة الخارجية أو اللاجينية التي تلعب دورا مهما في أمراض غير السرطان حيث ان هناك شواهد قوية لمشاركة هذا النوع من الوراثة في الاضطرابات العقلية وأمراض المناعة الذاتية وغيرها من الأمراض المعقدة مثل الزهايمر.

كل هذه التطورات رافقها إيجاد نماذج نظرية لفحص الآليات وفهم النواحي الكمية وذلك بإنشاء ما يسمى ببرامج محاكاة للوراثة اللاجينية النظرية أو على الحاسوب *In silico simulation* epigenetics ، والتي بطبيعة الحال تحتاج الى تطوير مزودات خدمة خاصة من النواحي الإحصائية التي تمكن الباحثين من إجراء تحليلات معقدة في الوراثة اللاجينية بطريقة سهلة . وهذا يحتاج الى قواعد بيانات متخصصة إضافة الى التي ذكرت أعلاه منها قواعد تخص خطوط الخلايا الطبيعية والسرطانية، وبيانات حول جينات إنزيمات مثيلة DNA . وهناك قواعد بيانات خاصة بالهستونات المعروفة وتواليها وكذلك قواعد للبروتينات اللاهستونية ذات العلاقة بالكروماتين أو غيرها مثل RNAi-associated proteins . فضلا عن وجود قواعد تتخصص بالعوامل المحورة للكروماتين التي توجد توالياتها في جينومات حقيقيات النواة. وتوجد قواعد بيانات لمعظم ما ذكر والبعض متخصصة جدا مثل تلك الخاصة بالكروموسوم البشري رقم 21 و 22 وأخرى خاصة بالخلايا الجنسية . فضلا عن وجود قواعد تخص مثيلة DNA في التوائم المتباينة أو أحادية الزيجة .

تقنيات تضخيم DNA والمعلوماتية الحيوية

تحتاج العديد من العمليات في مجال البيولوجي الجزيئي الى تضخيم تواليات معينة أو جينات معينة لاستعمالها في أغراض خاصة. وعمليات التضخيم والكثرة قد تكون داخل الأنظمة الحيوية كما في تكثير النواقل الحاملة لجين معين داخل مضاييف بكتيرية ملائمة وأكثرها استعمالا سلالات خاصة من *E. coli* ولكن هذه تواجه العديد من المشاكل لذلك وجدت طرق لتكثير وتضخيم قطع DNA خارج الأنظمة الحية وأكثرها أو أهمها استعمالا هي تقنية تفاعلات الكثرة Polymerase chain reactions (PCR) التي تساعد في تضخيم قطع DNA بشكل تزايدى وهو تفاعل إنزيمي يتبع أساسيات رياضية بسيطة معروفة مسبقا، والأخيرة هي الأوثق علاقة بالمعلوماتية الحيوية من حيث البرمجة والوسائل اللازمة للتحضير لإجراء التجارب المختبرية. لذلك سيتم التأكيد في هذا الجزء من الكتاب على تفاعلات الكثرة ومتطلباتها عند وضع البرامج، واهم المتطلبات هو تصميم البوادي Primers لبدء تفاعلات التضخيم، مع التطرق الى بعض الأنواع الرئيسية نظرا لعلاقتها الوثيقة بعمليات البرمجة.

تفاعل الكثرة PCR تقنية تقوم بتضخيم قطع DNA الواقعة بين تواليات معروفة خارج الأنظمة الحية. وظهرت التقنية عام 1985 وبعدها ازداد استعمالها وتحويرها لتخدم أغراض مختلفة، وقد مر ذكر بعض جوانبها مسبقا ونظرا لأهمية التقنية ظهرت لها موسوعات خاصة PCR encyclopedia تضم كل ما تحتاجه التقنية من قواعد بيانات للبحث وكتب خاصة بالأنواع المختلفة للتقنية، أودعت البيانات من قبل العاملين في المجال والموسوعة تكون مباشرة الاستعمال على صفحات الانترنت. وقد خصصت مواقع عديدة للتعامل مع التقنية من حيث إيجاد البوادي

اللازمة لتضخيم قطع DNA مع تقديم التفسيرات اللازمة. وقد حورت الطريقة بطرق شتى لتستخدم أغراض مختلفة ولكن الأساس فيها واحد.

الأغراض من استعمال تقنية تفاعل الكوثرية

هناك العديد من الأغراض التي تساعد فيها التقنية وازدادت بعد ظهورها وتنوعت بشكل كبير جدا ومنها:

- تضخيم تواليات محددة من الجينومات مثل الجينوم البشري
- التمييز بين أنواع الكائنات الحية والأنواع الثانوية والسلالات العائدة للنوع الواحد
- تحديد تواليات DNA .
- الكشف عن الطفرات وعلاقتها بالأمراض وكذلك تتبع بعض المعالجات الدوائية مثل مراقبة علاج السرطانات .
- الكشف عن الإصابات البكتيرية والفيروسية وغيرها من الممرضات خاصة البطيئة النمو مثل بكتريا السل .
- دراسة التطور على المستوى الجزيئي .
- دراسة مثيلة DNA التي أصبحت مهمة جدا في تشخيص السرطانات ودراسة الوراثة اللاجينية. ويتضمن ذلك تحديد تواليات السايتوزين غير الحاوية على المثل
- تحديد حدود الاكسونات والانترونات وأنواع الخياطة البديلة
- تحديد Alu sequences المنتشرة في الجينوم البشري لتحديد البصمة الوراثية Finger print
- تحديد وتمييز المكررات وتوابع DNA الصغيرة والمكررات المتغيرة VNTRs والأخرى التي تتكدس عند أطراف الكروموسوم على عكس الأولى التي تكون منتشرة على الكروموسوم

مقومات تفاعل PCR

هناك أكثر من مقوم لإجراء التفاعل منها نوعية خليط التفاعل وما يحويه من محاليل دائرية وتراكيز أيونات موجبة ، وأوليات لبناء DNA مثل النيوكلويتيدات ، وإنزيمات الكوثرية DNAPs مثل (Taq) والبوادي التي تعد الأساس في نجاح عملية التفاعل ، لذلك سيتم التركيز على هذه النقطة لعلاقتها الوثيقة بالمعلوماتية الحيوية .

خليط تفاعل PCR

- يشمل عدة مواد كما ذكر أعلاه ومنها مؤثرة جدا" مثل
- **تركيز البادئ** : وهذا يمكن ان يحسب من القياسات عند طول موجي 260 نانومتر للقواعد الأربعة والتركيز العام الأمثل يقع ضمن 1 مايكرومول ، والأفضل ان يكون بين 0.1 – 0.5 مايكرومول ، ويجب ان يكون بكميات كافية ليشجع تفاعل التحام البادئ بأشرطة القالب بدلا" من التحام الأخيرة مع بعضها نتيجة المنافسة .
- **تركيز القالب** : ويكون بمستوى نانوغرام أي بمستويات واطئة ليتم تضخيمها
- **الدوراء** : محاليل توفر الظروف الملائمة لفعالية DNAP وثبوته وتحوي على أيونات ثنائية التكافؤ مثل Mg^{++} ، Mn^{++} والأكثر استعمالا" هو Mg^{++} ، اما Mn^{++} فيمكن ان يستعمل في حالة تطفير DNA الذي تتوسطه عمليات استعمال PCR لان التراكيز العالية من Mn^{++} تزيد من معدل الخطأ أثناء عملية تخليق DNA . وكذلك وجود أيونات أحادية التكافؤ مثل K^{+} .

الإنزيمات المستعملة : وتتمثل بأنزيم كوثرة DNAP والمستعمل بكثرة هو المشتق من البكتريا المحبة للحرارة *Thermus aquaticus* (Taq) على ان يكون بمواصفات خاصة لتعديل سرعة التفاعل . وفي بداية استعمال تقنية تفاعلات الكوثرة استعملت وحدة Klenow subunit من DNAP للـ *E. coli* الذي يستطيع ان يضخم تواليات قصيرة اقل من 100 قاعدة ولكن كونها حساسة للحرارة استبدلت بإنزيمات مشتقة من أحياء محبة للحرارة كما ذكر آنفاً . وهناك العديد من الإنزيمات التي يمكن ان تستعمل في هذه التفاعلات قد تكون طبيعية او مشتقة من أحياء محبة للحرارة او تكون مهندسة وراثيا لأداء هذه الأغراض وتوجد شركات متخصصة لتسويقها وفق الطلب وتختلف المستحضرات المختلفة للإنزيم الواحد وفقاً لتشكيلة المستحضر Formula وظروف التفاعل وكذلك بالنسبة لتعريف الوحدة . وبالنسبة لإنزيم Taq فالتركيز الموصى به 1-2.5 وحدة / 50-100 مللتر من خليط التفاعل عندما تكون الظروف الأخرى في الحدود المثلى . وفي العموم تستمر عملية التخليق بشكل تزايدى الى حد معين ثم تتراجع لعدة أسباب فمثلاً بعد 30-25 دورة حرارية يكون الإنزيم قد خلق حوالي 10⁶ نسخة من قطعة DNA وبعد ذلك تقل كفاءة التفاعل نظراً لتأثر الإنزيم بالحرارة ، فضلاً عن كونه يصبح غير كافياً لأحداث التفاعلات التزايدية .

تركيز أوليات DNA

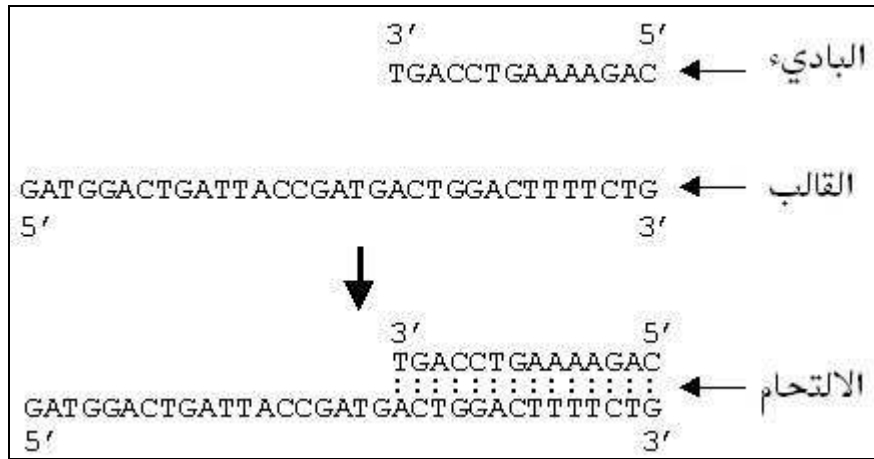
وتتمثل بالنيوكليوتيدات A , C , G , T وهذه يجب ان تكون بتركيز كافية وكذلك يجب ان تحسب التراكيز الملائمة من الايونات الموجبة وفق الخليط ووفق ظروف تطبيق عملية التضخيم .

تفاعلات PCR

اغلب الأحيان لا يحتاج تفاعل PCR الى نماذج نقية من DNA على شرط عدم احتواء النموذج على مثبطات إنزيم الكوثرة . ولكن في حالة تضخيم القطع الطويلة جداً من الجينومات تحتاج الطريقة الى استعمال مصادر DNA نقية وفي العموم فأن لكل نموذج طرق خاصة لاستخلاص DNA منه التي أثبتت فاعليتها بالخطأ والتجربة . وتشير التجارب الى ان تواليات بحدود 100-600 قاعدة يمكن ان تتضخم بسهولة والعموم هو بين 150 – 1000 قاعدة التي يمكن ان تكون من نماذج الأنسجة المثبتة او DNA الجينومي او البلازميدي وهذا الحجم يكون مثالياً لعمليات دراسة تهجين المجسات Probe hybridization .

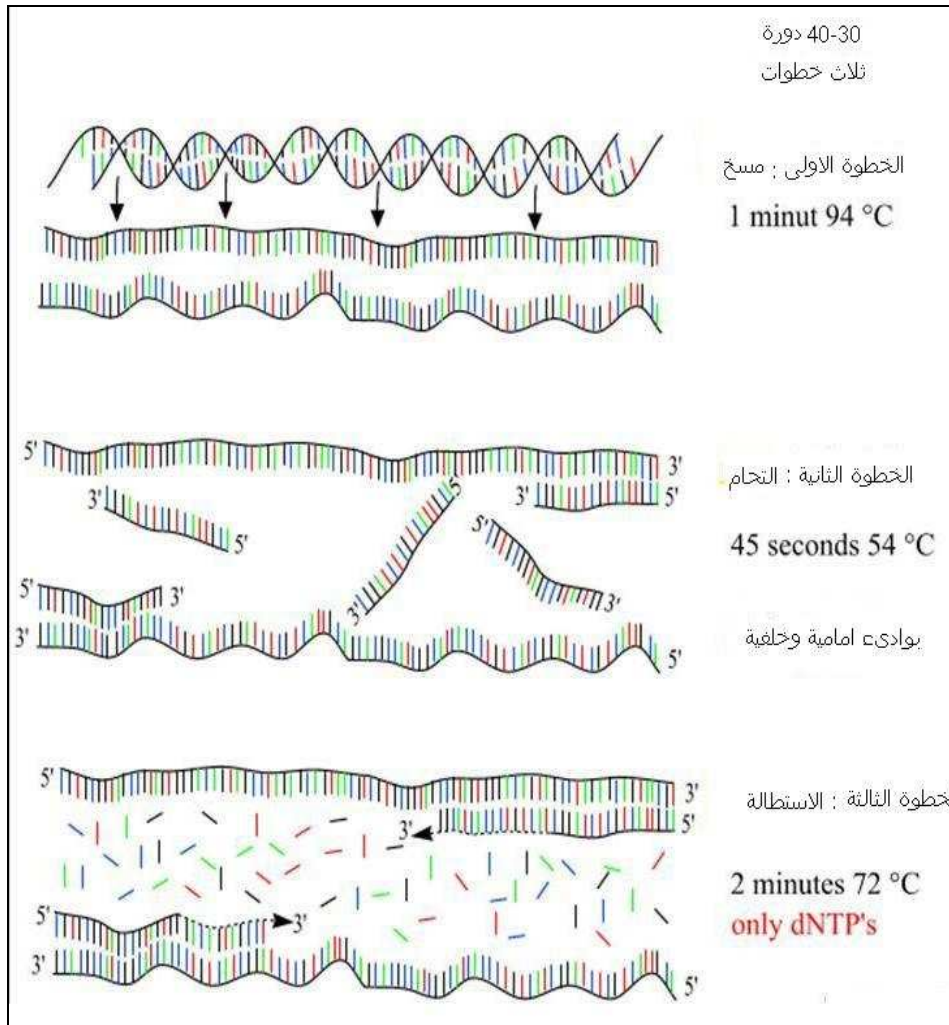
ويمكن تقسيم عملية التضخيم الى خطوات

- 1- صهر أشرطة DNA المستهدف
 - 2- التحام البوادي في المناطق المحددة لها على الأشرطة والذي يفترض ان يكون منشطاً
- كما موضح في الشكل 82.5 :



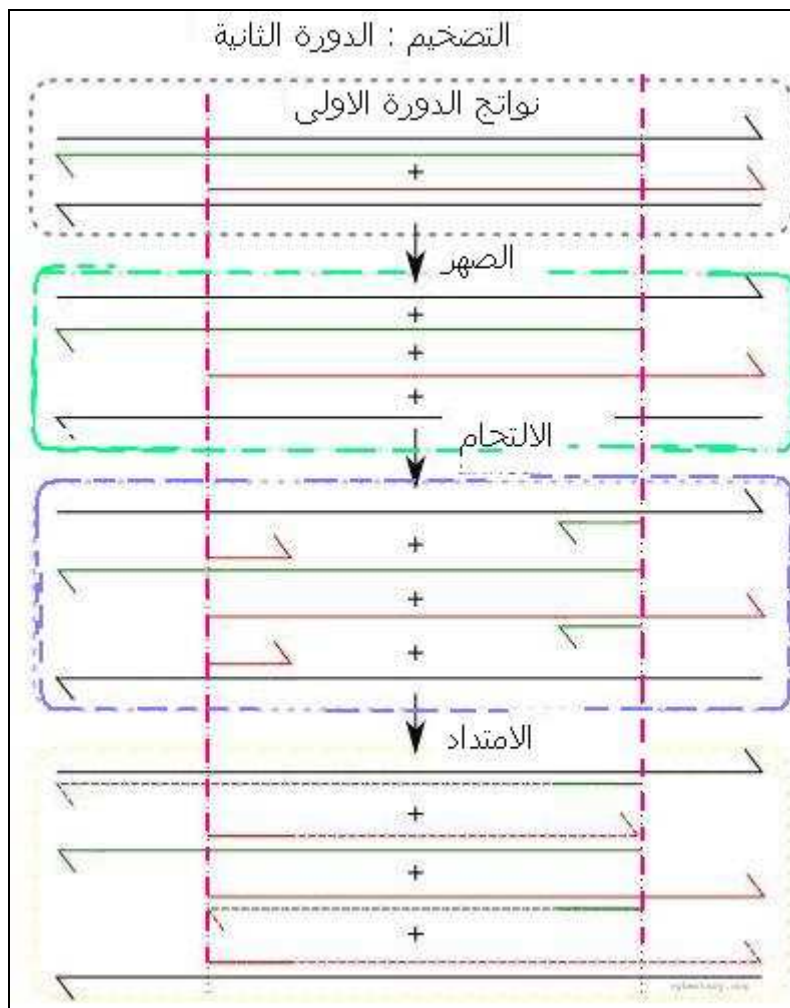
شكل 82.5: الالتحام البادئ مع DNA المستهدف للتضخيم

- 3- مباشرة DNAP بفاعليته بإضافة النيوكليوتيدات الى النهاية 3' أي ان الاتجاه يكون من 3' الى 5' كما موضح في الشكل 83.5



شكل 83.5 : تفاعل PCR

والملاحظ انه بعد انفتاح أشرطة DNA المستهدف ترتبط اليها البوايد ولا تعود الأشرطة للالتحام مع بعضها ثانية لان تركيز البوايد يكون أعلى . وفي الشكل أعلاه تكون نتيجة الدورة الأولى من عملية التضخيم 4 أشرطة ، 2 أصلية و 2 جديدة والملاحظ من الشكل التالي (84.5) ان المنطقة المراد تضخيمها محددة ولكن الإنزيم لا يعرف أين يقف ويعبر الخط الأحمر ويصل الى أي مكان يستطيعه كما موضح في الشكل 84.5 :



شكل 84.5 : دورات PCR المتتالية

وبدرجة الحرارة في خطوة الاستطالة تعتمد على إنزيم الكوثر المستعملة فمثلاً Taq polymerase تكون فعاليته المثلى بدرجة حرارة 75 - 80 °م وفي العادة تستعمل درجة حرارة 72 °م وفي هذه الخطوة يتخلق شريط مكمل للشريط المفرد بإضافة dNTPs التي تكون مكاملة بالاتجاه 3' الى 5'. والوقت اللازم لهذه الخطوة يعتمد على كل من DNAP المستعمل وطول قطعة DNA المراد تضخيمها ، والإنزيم يكوثر الآلاف القواعد \ دقيقة ، وعند توفر الظروف الملائمة يتم كوثر DNA بشكل تزايد.

والخطوة الأخيرة من الاستطالة فتتم عند درجة حرارة 70-74 °م لمدة 5 - 15 دقيقة بعد آخر دورة من دورات PCR لضمان ان كل الأشرطة المفردة قد تم تفاعلها ، والتوقف الأخير Final hold تجري بدرجة حرارة 4 - 15 م . وفي العادة فان الاستطالة غير المتخصصة لا تستمر كثيرا". الدورة الثانية من التضخيم تنتقل من 4 أشرطة الى 8 أشرطة أي يكون التزايد بواقع 2^x حيث يمثل X عدد الدورات والأشرطة تكون :

- الأشرطة الأساسية وتحتوي على امتدادات من الجهتين وتكون أطول من المنطقة المستهدفة

- نواتج PCR وتبدأ عند البادئ ولكنها مستمرة ابعد من ارتباط البادئ الثاني وهذه تنتج من الدورة الأولى
- في الدورة الثانية تظهر أشرطة تكون من بداية الى نهاية القطعة المستهدفة أي تكون قد قصرت عند النهايتين وبذا تمثل القطع المرغوبة
- لذلك يكون طول القطع المتضخمة Amplicons يحدد بالغرض من التجربة بالنسبة للعمليات التي يرام تحديد الكميات فيها Quantitative PCR (qPCR) وتكون بحدود 100 قاعدة اما في العمليات العادية والقياسية فهي يمكن ان تصل الى 500 قاعدة . واذا كان موقع ارتباط البادئ معروف مسبقا" فإن كمية النواتج يمكن ان تقاس بالمعادلة :
طول الناتج = موقع ارتباط الشريط النقيض Antisense - موقع ارتباط الشريط الحساس Sense + 1
ومرحلة الكثرة تقسم الى ثلاثة مراحل
- مرحلة التضخيم التزايدى وفيه يتم مضاعفة الجزيئات في كل دورة على افتراض ان كفاءة التفاعل تصل الى 100 %.
- مرحلة التباطؤ وتحدث نتيجة فقدان فعالية DNAP ونضوب مواد التفاعل مثل dNTPs
- مرحلة الاستقرار او الهضبة Plateau ولا تحصل فيها زيادة في النواتج نتيجة لاستنزاف مواد التفاعل وانخفاض فعالية الإنزيمات
- وللوصول الى امثل الحالات وتخطي فشل عملية PCR الحساسة للتلوث يجب فصل جزيئات DNA الملوثة والاهتمام باختيار البوادئ حتى لا تحصل عمليات كثررة مزيفة ، وكذلك الاهتمام باختيار إنزيمات الكثرة . وبالنسبة لأفضلية القطع المراد تضخيمها من ألجين مثلاً" فيفضل تضخيم القطع القريبة من الطرف 3' للجين لانها موثقة أكثر .
- وفي العموم تستمر عملية التضخيم بشكل تزايدى الى ان يصل الى حد الهضبة او الاستقرار او طور الخط الأفقى أي عدم حصول تفاعل تزايدى ويعود ذلك لعدة أسباب ذكر بعضها في مواقع أخرى .
- وتجرى العمليات في جهاز الدورات الحرارية Thermal cycler الذي يسخن ويبرد اعتماداً على عكس التيار الكهربائي في الأغلبية العظمى من الأجهزة المستعملة ، ثم تجرى عملية الترحيل الكهربائي لنواتج التفاعل لفصل الجزيئات وفق الحجم الذي يحدد باستعمال سلم التعبير (Ladder) المكونة من خليط في جزيئات DNA المعروفة الحجم او الوزن الجزيئي (واسمات جزيئية) ويكون الترحيل في هلام الاكروز بتركيز ملائم وظروف ملائمة من حيث الجهد الكهربائي وغيره ويتم إظهارها بعد معاملتها بمركب بروميد الاثيديوم Ethidium bromide او باستعمال طريقة ELISA .

البوادئ وتصميمها

البوادئ هي اثنين من قطع DNA الصغيرة المفردة الأشرطة التي تستطيع ان ترتبط وتتجهن على جانبي ألجين او التوالي المراد تضخيمه . ويعد البادئ من العوامل المؤثرة جداً" في التفاعل ، اذ ان الاختلال في تواليه عند التصميم يمكن ان يؤدي الى إجهاض التفاعل ، ويؤثر البادئ في تحديد طول النواتج ودرجة انصهارها وكمياتها فضلاً" عن إمكانية ظهور نواتج تضخيم غير متخصصة مثل تكوين مزدوجات البادئ التي تصبح بتراكيز عالية وتقوم بالمنافسة على عملية البدء وإحباط عملية تكوين النواتج فهي تقنية حساسة جداً" للتلوث الذي يمكن ان يأتي من النماذج ويعطي نتائج موجبة كاذبة لذلك كانت هناك قيود على طرق استخلاص DNA للاستعمال في هذه

التقنية . ومن المؤشرات التي تؤخذ بنظر الاعتبار عند تصميم البودائى او تصميم برامج الحاسوب الخاصة بتصميمها وقابلية توالي قصير من القواعد ان يكون بادئا "كفوءا" تعتمد على عدة عوامل منها

- حركات الارتباط وفك الارتباط بين مزدوجات القالب DNA template والبودائى عند درجات حرارة الالتحام والاستطالة Extension

- ثبوت المزدوجات عند وجود نقطة او نقاط من عدم التوائم او التطابق Mismatches وكل هذه المواصفات تحدد بمؤشرات عدة منها

0 الطول : يمثل طول البودائى احد المؤشرات المهمة في تصميم البودائى نظرا " لارتباطه الوثيق مع مؤشرات أخرى مثل حرارة الانصهار T_m وحرارة الالتحام T_a . والحقيقة العامة ان طول البودائى يتناسب عكسيا" مع عملية التضخيم . ويمكن ان تصل أطوال البودائى من 12 - 500 قاعدة ولكن الأطوال المثالية المسجلة في الدراسات تتراوح بين 18 - 30 قاعدة وقد وجد بالتجربة ان التخصصية لا تزداد عند زيادة الطول لأكثر من 30 قاعدة فيما اذا كان الفرق بحدود 5 م بين حرارة الالتحام ودرجة حرارة التفكك بين مزدوج البودائى والقالب . والبودائى ليس بالضرورة ان يكون قصيرا" الا في حالات خاصة يمكن ان يستعمل بودائى مكون من 10 قواعد فقط .

وفي عموم الدراسات هناك شبه اتفاق على ان البودائى بطول أكثر من 16 قاعدة لا يلتحم بكفاءة وتخصصية عالية مع القطع غير المستهدفة (وهي الحالة المفضلة) اما البودائى القصيرة مثل اقل من 12 فقد وجد انها يمكن ان ترتبط الى مواقع كثيرة في الجينوم البشري المكون من 3×10^9 قاعدة في حين ان المكون من 20 قاعدة يرتبط باحتمالية 1×400 وهذه الحالة تفيد في حالات خاصة كما سيأتي ذكره لاحقا" . وبعض برامج الحاسوب المتخصصة بتصميم البودائى توفر إمكانية التحكم بطول البودائى من قبل المستخدم ولكن بحدود معينة .

ولكن مع هذا تكون الحاجة ماسة الى البودائى الطويلة لتمييز الجينات ذات التماثل الجذري في الأنواع المختلفة وكذلك عند توقع عدم تكامل كامل لكل التواليات في القالب . ويمكن ان تستعمل البودائى الطويلة عند الحاجة الى معلومات إضافية مثل مواقع قطيقات الارتباط او مواقع القطع بالإنزيمات القاطعة الداخلية او كماشة GC المرتبطة الى النهاية 5' .

0 درجات الحرارة : تشمل درجات حرارة الانصهار T_m وهي الدرجة الحرارية التي تكون عندها نصف الأشرطة المزدوجة بشكل أشرطة مفردة ، وتستعمل المقاييس المترية لتعكس ثبوت مزدوجات DNA فالمزدوجات الأكثر ثبوتا" هي التي لها درجات انصهار عالية ، وتكون الأشرطة الطويلة لها درجات انصهار عالية ، وكذلك تتأثر درجة الانصهار بمحتوى قطع DNA من G , C ، والمعادلة العامة المستعملة لقياس درجة الانصهار هي معادلة والاس Wallace formula التي تقيس درجة الحرارة السيلييزية (المئوية)

$$T_m = 4 (G + C) + 2 (A + T)$$

فتعطى درجتين للـ $A + T$ وأربع $G + C$ ، وهي معادلة غير دقيقة وتصلح للبودائى بطول 18 - 24 قاعدة .

وللتحديد المضبوط تستعمل نماذج الثرموديناميكية للجوار الأقرب Nearest- neighbor thermodynamic models ويتم التعامل مع هذه باستعمال الطاقة الحرة ΔG (Gibbs free energy) .

والطاقة الحرة ΔG هي مقياس لكمية الشغل الذي يمكن أدائه عند ضغط ثابت ، فهي تقيس تلقائية التفاعل والمعادلة التالية تمثل احد الصيغ المعتمدة على الطاقة والتي يمكن ان تحور بتغيير احد مفرداتها ولكنها في النهاية تعطي القيم ذاتها ولكنها تناسب طول معين

$$T_m = \Delta H / [\Delta S + R \ln (C/4)] - 273.15 + 16.6 \log 10 [J^+]$$

H هي Enthalpy كمية الطاقة الحرارية في المواد

ΔH التغير في Enthalpy

S Entropy الانتروبي

$[J^+]$ تركيز الأيونات أحادية التكافؤ

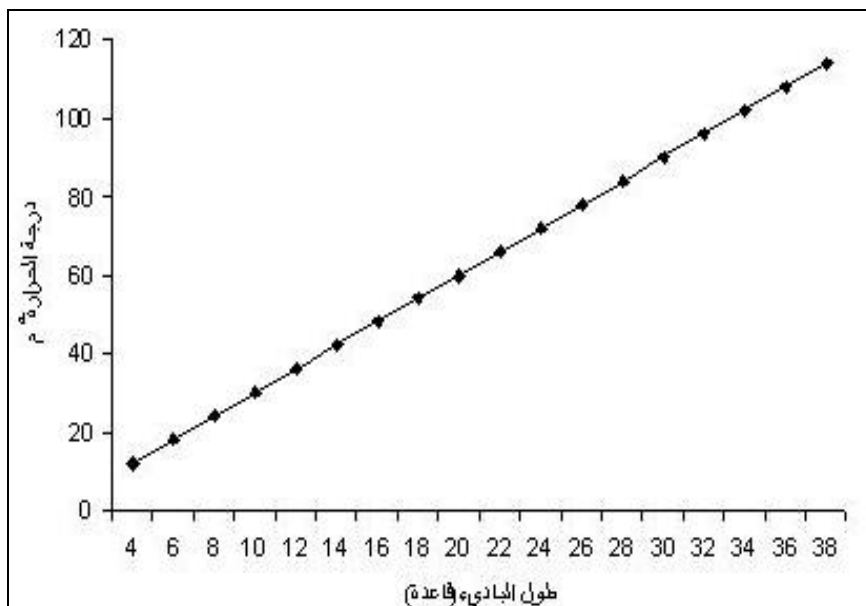
R ثابت الغاز المولاري

C تركيز البادئ

ودرجة حرارة الانصهار تقسم الى درجة حرارة انصهار البادئ (Primer T_m) وتمثل درجة حرارة الانصهار لحالة ازدواج البادئ – القالب التي يكون اقل ثبوتاً ، اما حرارة انصهار النواتج (Products T_m) فتمثل درجة انصهار نواتج التضخيم .

ويمكن ان توضع معادلة لحساب الطاقة وحرارة الانصهار بالبرنامج وهذا يجب ان يعلن مع مرفقات البرنامج . حسابات الطاقة تأخذ بنظر الاعتبار طاقة الرص Stacking energy للقواعد النتروجينية المتجاورة .

ويفضل ان تكون درجات الانصهار لكلا البادئين الأمامي Forward primer والعكسي Reverse primer متشابهة لانها ستزيد من التخصصية . وحرارة انصهار البادئ بين 52 - 58 °م وقد تصل الى 72 °م ، وتكون النتائج أفضل من البوادئ التي لها درجات انصهار أوطأ ، اما التي لها درجات انصهار أعلى من 65 °م فتستعمل بعض الأحيان لمنع حدوث الالتحام الثانوي. ولذا وجب قياس الحرارة عند الظروف المثلى فمثلاً " معادلة والاس المذكورة أعلاه مصممة ومقاسة عند رقم هيدروجيني (pH) قريب التعادل 7 وتركيز ملحي (من الأيونات الأحادية التكافؤ) بتركيز 1 مولر ويوضح الشكل (85.5) علاقة طول البادئ مع درجات الانصهار .



شكل 85.5 : علاقة طول البادئ بحرارة الانصهار

فالبوادئ التي تكون لها درجة انصهار T_m أعلى من درجات الالتحام يمكن ان تتجهن بشكل خاطئ اما الواطئة مقارنة بحرارة الالتحام فيمكن ان تفشل في الالتحام ولذلك وضعت المعادلات الخاصة بحسابها المذكورة أعلاه.

حرارة الالتحام T_a

ويفترض بها ان تكون اقل بحوالي 3-5° م من اقل درجة من درجات الانصهار وعند شدوذاها عن هذه القاعدة فانها تؤدي الى اضطراب عملية التضخيم ، فالواطئة تؤدي الى ارتباط غير متخصص وإعطاء نواتج PCR كثيرة ، اما العالية فيكون عندها التحام البادئ ضعيفا" ويؤدي الى قلة في النواتج ، وبطبيعة الحال فهي مثل حرارة الانصهار تعتمد على طول البادئ مباشرة وكذلك مكونات البادئ من القواعد النتروجينية . ويمكن قياس درجة الحرارة المثلى للالتحام وفق المعادلة الآتية :

$$T_a \text{ opt.} = 0.3 (T_m \text{ of primer}) + 0.7 (T_m \text{ of product}) - 25$$

ويمكن ان يتم الالتحام بدرجات 50 – 65° م والدرجة العظمى لا تتعدى 80° م والبادئ الأمامي والعكسي يلتحمون عند درجة الحرارة نفسها . وبعض برامج الحاسوب توفر القليل من المرونة للتحكم بدرجات الحرارة .

مكونات البادئ من القواعد النتروجينية

ان محتوى البادئ من القواعد النتروجينية يؤثر في كل من درجة حرارة الانصهار والالتحام ويفضل ان تكون البوادئ مكونة من أعداد متساوية من القواعد النتروجينية الأربعة. ولكن في بعض الأحيان لا ينطبق هذا على البوادئ بحيث لا تتطابق بشكل مطلق مع القالب كما في حالة مواقع القطع بالإنزيمات القاطعة وفي حالة البوادئ غير المنضبطة Degenerated primers . وزيادة محتوى البادئ من GC يشير الى ارتفاع درجة الانصهار لان الطاقة اللازمة لكسر الأواصر الثلاث بين G و C تكون اكبر من تلك الخاصة بـ A و T .

وتشير الدراسات الى ان أفضل محتوى للبداى من GC يكون بحدود 45-60 % والموصى به 45-55 % فعندما تكون النسبة اقل من 50% يكون من الضروري إطالة البداى الى أكثر من 18 قاعدة للحفاظ على درجة انصهار ملائمة او تضاف كماشة GC (GC clamp) الى النهاية 5' من البداى لرفع درجة الانصهار والتقليل من تكون الحزم الثانوية الكاذبة ويجب عدم إدخال قاعدة Inosine في توالي البداى لانه يمكن ان لا يعمل او يعطي نواتج قليلة . ويكون من الضروري تجنب وجود مكررات من اثنين من القواعد وأقصى ما يمكن السماح به هو أربعة مثل ATATATAT. اما تكرار قاعدة واحدة لمرات عدة فيجب تجنبها ايضا مثل AGGGGGATGGGG فهذا التوالي فيه 5 و4 مكررات للـ G وهنا يؤدي الى عدم كفاءة عملية البدء . وفي العموم يجب انتخاب التواليات بحيث تكون الوحيدة التي ترتبط الى DNA والابتعاد عن التهجين غير الصحيح . ويستعمل برنامج BLAST او غيره لإيجاد المناطق التي يتجهن البداى مع DNA مستهدف .

الطرف 3'

هذا الطرف في البواى يكون أساسا في التخصصية والسيطرة على عدم التوائم . ويكون وجود كل من G و C مرغوبا في هذا الطرف ولكن بمعدل ملائم ، اذ ان زيادتها تؤدي الى التحام البداى في مواقع مختلفة من قالب DNA . وعدم توائم البداى مع القالب في هذه المنطقة يمكن ان يؤدي الى عدم استطالة DNA . وتوجد بعض البرامج التي تسمح بالتأكد من ان 8 – 10 قواعد عند هذا الطرف هي فريدة . وفي العموم فأن وجود G و C في هذا الطرف مرغوبا كما ذكر آنفاً الا انه يجب ان لا تكون بأكثر من ثلاث مواضع متتالية يعني (CCC او GGG) ، وفي حقيقة الأمر فان كل من الطرفين لهما تأثير كبير في عملية التضخيم . فالطرف 3' اذا لم يكن ملائما فإنه يؤدي الى عدم البدء Mispriming.

وعند النظر الى تصميم البداى من أي البرامج يجب التركيز ان طرفه 3' ينتهي بـ G او C او GC او CG وذلك لمنع الارتخاء وزيادة كفاءة ربط البداى . والطاقة الحرة لهذا الطرف تقاس للثمالات الخمسة الأخيرة ، فالطرف 3' غير الثابت تكون قيمة ΔG له اقل سالبية وتؤدي اضطراب عمليات البدء .

مشاكل تصميم البواى

هناك العديد من المشاكل التي تواجه تصميم بواى كفوة ، فالبداى الجيد يجب ان يكون فريد التوالي ولكن بعض المشاكل تكون ناتجة عن طبيعة تركيب القالب التي يمكن ان تؤدي الى ظهور المسحات Smears في هلام الفصل بعد إجراء عملية التضخيم ، ففي العموم تعمل البواى بكفاءة أعلى مع تواليات القالب التي فيها تراكيب ثانوية قليلة لان الأخيرة تعيق التحام البداى وعمليات الاستطالة . ومشاكل أخرى تأتي من البداى المصمم وفي هذه الحالة فأن المشاكل الناتجة هي ما يتعلق :

أولاً : حالة التطابق ومنها

- تكوين ماشة الشعر وتنتج عن تكامل بعض التواليات ضمن توالي البداى نفسه .
- تكون مزدوجات البواى نتيجة للتكامل بين جزيئات البواى

فوجود مكررات من G (poly G) او مكررات C تؤدي الى التحام غير متخصص ، اما وجود مكررات A , T فيجب تجنبها أيضا لانها تؤدي الى تراخي المزدوجات المتكونة بين البداى والقالب وبالتالي تقلل من كفاءة التضخيم ، اما مكررات (T , C) و (A , G) فهي الأخرى يجب

تجنبها . ومثل هذه المشاكل يمكن التحكم بها بتحديد كماشة GC عند النهاية 3' للتقليل من تراخي الارتباط الذي يمكن ان يحصل وزيادة تخصص الارتباط ولكن بشرط ان لا تكون أكثر من ثلاثة ولعل مما ذكر أعلاه فإن تركيب ماشة الشعر وهو تركيب ثانوي يعد من أهم المشاكل التي تتكون نتيجة لتداخل تواليات البادئ نفسه وتكوين تراكيب ثانوية . ويمكن التعامل مع هذه المشكلة بقياس الطاقة الحرة ΔG التي تمثل الطاقة اللازمة لكسر التركيب الثانوي فالقيم العالية السالبة لـ ΔG تشير الى ثبوت التركيب وماشة الشعر غير مرغوب فيها ، ووجود ماشة الشعر في الطرف 3' يكون أكثر تأثيراً من وجوده في مناطق أخرى . ويقاس الفرق في الطاقة الحرة بالمعادلة :

$$\Delta G = \Delta H - T \Delta S$$

حيث يمثل ΔH Enthalpy و ΔS Entropy
وأقصى مدى يمكن تحمله عند الطرف 3' لتركيب ماشة الشعر من الطاقة الحرة ΔG هو (-2) كيلوسعرة \مول وللماشة الداخلية $\Delta G = -3$ كيلوسعرة \مول . والزيادة عن هذه الحدود ووجود ماشة الشعر يقلل من قابلية الارتباط الى المواقع المستهدفة . وتوجد برامج خاصة لتصميم البوادي التي تستطيع ان تكشف عما اذا كان البادئ يكون تركيب ماشة الشعر وكذلك تقيس ΔG الخاصة بها .

والمشكلة الأخرى لحالة التطابق هي تكوين مزدوجات البوادي وهذه تحدث عندما يكون البادئ الأمامي مكملًا للبادئ الخلفي لذلك وجب تجنب هذه الحالة . وفي حالة وجود التكامل فإن الطرف الأيسر (5') لأحد جزيئات البادئ يرتبط الى الطرف الأيمن (3') من جزيئة بادئ مماثلة أخرى ، ويمكن ان تنتج المزدوجات عند وجود تكامل القواعد في المناطق الداخلية للبوادي وخاصة ان تراكيز جزيئات البادئ تستعمل بمستويات أعلى من تراكيز قطع DNA القالب وترداد هذه المشكلة عندما تكون جزيئات البادئ أكثر ميلاً للتهجين مع بعضها أكثر من ميلها للتهجين والارتباط الى قالب DNA وبالتالي تؤدي الى تقليل النواتج ، وهناك مقاييس لمثل هذه الحالات ، فالبادئ الذي يشكل مزدوج ذاتي بطاقة ΔG مقاربة الى -5 كيلوسعرة \مول ، والمزدوجات الذاتية الداخلية بطاقة -6 كيلوسعرة \مول يمكن تحملها ومقبولة ويمكن العمل بها . وفضلاً عن تكوين ماشات الشعر نتيجة للتكامل الداخلي لتوالي البادئ يمكن ان تتكون الفقائيع Bulges او التراكيب الصليبية الشكل والعروات الداخلية .

ثانياً : المشاكل المتعلقة بعدم التوائم او التطابق ، ان مدى التطابق يؤثر في حرارة الصهر T_m لذلك تكون القيم المتقاربة لها (وبالتالي ملائمة البادئ الأيسر والأيمن) تؤدي الى زيادة الحاصل في عملية التضخيم ، والاختلاف بين درجتي انصهار البادئ الأيسر والأيمن الى أكثر من 5° فإنه يمكن ان يؤدي الى عدم حصول أي تضخيم .

وهناك جوانب جيدة وأخرى سيئة لحالة عدم التطابق للبوادي مع القالب . فعدم التطابق او التوائم حتى في موقع واحد عند الطرف 3' يؤثر في ثبوت البادئ وكفاءة تفاعل التضخيم وهو الأكثر تأثيراً من حالات عدم التطابق في مواقع أخرى . كما ان البادئ الذي يحوي على تواليات يمكن ان تتطابق مع مناطق مختلفة من تواليات القالب يعطي نواتج متعددة وهذا يعتمد على طبيعة القالب الذي سيمر ذكره لاحقاً .

ولكن بعض الأحيان تحتاج العمليات الى بوادي لا تنطبق بشكل مطلق مع القالب ، فمثلاً عند دراسة توالي معين في كائن ما والمطلوب معرفة قرينه من الأحياء القريبة منه ففي هذه الحالة تكون الحاجة الى بوادي قريبة ولكن ليست منضبطة بشكل مطلق . كما ان بعض البوادي غير

المتطابقة جدا" تعمل بكفاءة اكبر من المتطابقة في حالة استعمال حرارة التحام واطئة او غيرها من الظروف التي تطبق دون المستوى الأمثل ، ولكن في مثل هذه الحالات يجب إطالة البادئ قليلا" ، فمثلا" عند استعمال نظام تضخيم الطفرات Amplification refractory mutation systems (ARMS) يفضل إطالة البادئ الى 30 قاعدة والتي تؤدي الى زيادة كفاءة عمل النظام عندما يكون هناك عدم تطابق في 3-4 قواعد الى يسار الطرف 3' للبادئ والتي تزيد من التخصص لذلك يصار الى إنشاء حالة عدم التطابق المتعمد في البوادئ عندما يراد استعمال هذه الأنظمة .

المشاكل المتعلقة بتركيب القالب

عندما يراد تضخيم قطعة من DNA واستعمالها كقالب لعمليات PCR هناك بعض الملاحظات التي يجب أخذها بنظر الاعتبار . وللتذكير فأن الجينوم البكتيري يحوي في القليل على مليون من القواعد النتروجينية مقارنة بعدة بلايين من القواعد في جينومات اللبائن . لذلك يكون من السهولة إيجاد تواليات قليلة من القواعد النتروجينية التي يمكن ان تقابل مواقع محددة في الجينوم البكتيري مقارنة بزيادة من التواليات المقابلة التي يمكن ان تصمم لمقابلة تواليات في جينوم اللبائن . ولكن يجب التذكر ان تواليات الجينومات هي ليست عشوائية وان الجينومات الكبيرة (مثل جينومات اللبائن) تحوي على نسخ جينات مضاعفة وتحوي مكررات طويلة من التواليات وكذلك احتواء الجينومات على جزر ومستويات عالية من CpG وجينات كاذبة واذا كانت البوادئ المصممة تشبه احد هذه المكونات فأنها ستؤدي الى اضطراب او اختلال المقاييس الإحصائية وعليه يمكن ان تظهر مواقع ارتباط غير مقصودة وفي هذه الحالة تحتاج العملية الى شئ من المرونة وترتيب ملائم حتى يتم معرفة عمليات التضخيم غير المقصودة وهذا يظهر جليا" مع تحديد او التعامل مع الجينومات الكاملة التي سيرد ذكرها في موقع آخر .

ومن المعروف ان ثبوت مزدوجات DNA تعتمد على توالي النيوكليوتيدات فيها ، فانصهار منطقه محدده من مزدوج DNA يعتمد على ما يجاوره من النيوكليوتيدات ، لذلك فأن عمليات الانصهار للمزدوجات يمكن لن يتم التنبؤ بها من التوالي الأولى للنيوكليوتيدات ، اما في حالة المسخ وإنتاج الأشرطة المفردة فأن هذه الأشرطة قد تميل الى تكوين تراكيب ثانوية نتيجة لتكامل القواعد فيها وهو ما يحصل أيضا" للبادئ عند تكوين ماشات الشعر او التراكيب الصليبية الشكل وغيرها ، وهذا يعني ان الأشرطة المفردة تكون غير ثابتة ، وتكوين التراكيب الثانوية ومدى ثبوتها يعتمد على الطاقة الحرة ودرجة الانصهار ، وهذه الحالة تؤخذ بنظر الاعتبار وبحيطة شديدة عند إجراء عمليات PCR التي تهدف الى تحديد الكميات qPCR . لذا فأن من الضروري تصميم البوادئ لمناطق القالب التي لا تكون تراكيب ثانوية ثابتة أثناء تفاعلات PCR . وتوجد برامج خاصة لتحديد هذه المناطق مثل Mfold الذي يصمم البوادئ بعيدا" عن هذه المناطق .

أنواع البوادئ المصممة

يتبين من المناقشات السابقة واللاحقة ان تصميم البوادئ يختلف باختلاف جزيئة DNA المستهدفة وكذلك الغرض من التضخيم ويمكن تقسيم البوادئ الى عدة أنواع باعتماد مؤشرات مختلفة :

- البوادئ العادية والتي تستهدف تضخيم قطعة معينة من التوالي او جين معروف ولها برامجها الخاصة .
- البوادئ المستعملة لقطع DNA المحورة ، وهذه تكون خاصة لبعض عمليات تحديد تواليات DNA الحاوي على المثلل التي تتم بمعاملة DNA بمادة Bisulphite لتحديد نمط المثلية وهي

أول الأحداث الوراثية اللاجينية المكتشفة والتي لا تزال أكثر الأحداث تعرضاً من حيث الدراسة والبحث ، وقد مر في مواقع أخرى من الكتاب ان مثيلة DNA وخاصة CpG يمكن ان تسكت الجين ، فضلاً عن ان نمط المثيلة يختلف في الأنسجة المختلفة فضلاً عن اختلافه مع المرحلة العمرية او تخصص الخلايا . وعند معاملة DNA بأن مركب الكبريت Bisulphite يحول السايكوتوزين الى يوراسيل ولكنه لا يؤثر في السايكوتوزين الحاوي على المثل ، ولذلك فإن المعاملة تغير من توالي DNA وهذا يؤثر في تقليل التمييز بين تغيرات النيوكلووتيد المفرد SNPs للسايكوتوزين والثايمين الناتجة من تأثير مركب الكبريت.

لذلك وعند تحديد تواليات DNA تستعمل معها بواقي عامة وتضخيم المناطق الحاوية على المثل والأخرى الخالية منه ويمكن استعمال التضخيم المعنقد Nested PCR وتعتمد الطريقة على قياس نسبة C الى T في مناطق محددة كمياً واستنتاج كمية هذه القواعد المدمجة أثناء عملية استطالة التوالي . والمعرق لهذه الطريقة هي الكلفة المرتفعة لذلك لا تساعد في طريقة المسح السريع ، وتستعمل في تحليل الاليلات الأبوية والأموية وكذلك تحديد طمغة الجينوم Genomic imprinting وتحليلها.

ويمكن استعمال طرق مختلفة لتحديد التواليات في حالة DNA الحاوي على المثل مثل Pyrosequencing ، فضلاً عن وجود طرق تستعمل DNA المعامل بالمركب الكبريتي وخاصة تلك العائدة لتغيرات النيوكلووتيد المفرد SNPs. وطريقة تضخيم DNA الحاوي على المثل تكون مهمة جداً في تحديد الخرائط اللاجينية خاصة بعد البدء بمشروع الخاص بالإنسان Human epigenome project (HEP) . وذلك لان الجينوم الخارجي او اللاجيني يكون اقل ثباتاً من الجينوم . وعليه فإن تحديد الخارطة اللاجينية او الخارطة الخارجية تكون اعقد من تحديد تواليات الجينوم . فالفرد يحوي جينوم واحد ولكن الجينوم اللاجيني يختلف وفق العمر وكذلك في الأنسجة المختلفة ، ووفق الظروف البيئية وأثناء حدوث الاضطرابات مثل الحالات المرضية ولذلك تفقد خرائط الجينوم اللاجيني في تفسير ظواهر كثيرة مثل الهرم او الشيخوخة والأمراض المستعصية ، وربما فسرت الفشل في الحصول على حيوانات سليمة وبكل فعاليتها ومدى الحياة الطبيعي بعد إجراء الكلونة وعمليات الهندسة الوراثية عليها .

لذلك كانت البواقي للـ DNA الحاوي على المثل مهمة جداً خاصة فيما اذا عرف ان انخفاض المثيلة العامة Global hypomethylation تؤدي الى قلة ثبوت الجينوم ، كما ان المثيلة العالية الموضعية Local hypermethylation لجينات إحباط الورم والممهدات تؤدي الى فقدان وظيفتها ، فضلاً عن ان بعض أنماط المثيلة تكون خاصة بأنواع السرطانات ولها قيمة في تحديد المعالجات المختلفة .

بواقي الاكسونات

تحتاج بعض الدراسات والبحوث تضخيم بعض الاكسونات ويتوفر لهذه الحالة برامج تحدد بواقي خاصة بمناطق ارتباط الاكسونات Exon-Exon junction (E-Ejn) . وهذه البواقي تسري عليها كل المحددات والمواصفات الخاصة بالبواقي الأخرى ويمكن ان تستعمل معها البرامج الخاصة بتصميم البواقي العادية . ولكن يجب تحديد موقع البادئ . فالمناطق في mRNA تحوي على الطرف 5' لأحد الاكسونات و3' للاكسون المجاور ، فتحدد E-Ejn في النسخ يكون ضرورياً لتسجيل نسق التعبير للخياطة البديلة.

وبوادي مناطق الارتباط (E-Ejn) تكون ضرورية، والتحديد يكون صعباً ويحتاج الى وقت ولكن برامج التصميم العادي يمكن ان تحد من هذه المشاكل الى حد ما. وبرامج تصميم البوادي للاكسونات تستعين ببرامج أخرى مثل BLAST وغيرها ، وتختلف البرامج في قواعد البيانات العامة التي تستعملها والأفضل في هذه الحالة اختيار قواعد البيانات الأكثر تخصصاً والموثقة بشكل جيد .

وبعد الحصول على البوادي من البرامج لابد ان تعرض لعملية تحكيم بالرجوع الى مؤشرات مرجعية خاصة بالبوادي مثل حرارة الانصهار ومحتوى GC % وغيرها، وتحدد درجة جودة البادي بوضع الجزاءات (Penalties) بين البوادي المرجعية والبوادي المصممة وتقدر الدرجة بعد طرح الجزاءات لتحديد صلاحية البادي .

تصميم البوادي غير المضبوطة Degenerate primers

البوادي في هذه الحالة هي أيضاً من تواليات قصيرة ولكن تصمم ليس اعتماداً على تواليات DNA وإنما باستعمال تواليات بروتينات معروفة التوالي من الحوامض الامينية لذلك فهي بوادي متشابهة ولكن غير متماثلة . اذ يمكن عكس عملية ترجمة تواليات الحوامض الامينية لتحديد تواليات القواعد النروجينية المحتملة التي يمكن ان تشفر للحوامض الامينية والتي قد يشفر لها بأكثر من شفرة وراثية ، لذلك يمكن ان ينتج عدداً من القواعد التي لا يمكن ان توائم تواليات DNA وعند القيام بتصميم مثل هذه البوادي يكون هناك أكثر من بادي يمكن ان يجرب لملاحظة التوائم . ويمكن التقليل من عدد البوادي الناتجة من إدخال اسم الكائن الحي في البرنامج مع جداول تفضيل الشفرات واستعمال شفرات ورموز IUPAC . وفي معظم الأحوال يتم التصميم من عمليات الاصطاف مع جينات موجودة في قاعدة بنك الجينات ولكن التشويش يأتي من استعمال تسميات IUPAC ، وعندها يتم تصميم خليط من البوادي الملائمة لكل الاحتمالات.

ان استعمال البوادي غير المتماثلة يؤدي الى تقليص تخصص عملية التضخيم والأخيرة يمكن ان يقلل منها باستعمال طريقة Touchdown PCR او ما يسمى بـ Step- down PCR وهي إحدى التحويرات التي أجريت على عمليات PCR لتقليل الخلفية غير المتخصصة ويكون بتقليل حرارة الالتحام عند استمرار عمليات PCR ، فدرجة حرارة الالتحام تكون بحدود 2-3 °م اقل من درجة انصهار البوادي T_m ، فالدرجات الحرارة العالية تعطي تخصصية عالية لارتباط البوادي والدرجات الحرارية الواطئة تعطي عمليات تضخيم أكفاً من النواتج المتخصصة الناتجة من الدورات الأولية أي ان العملية تجرى بدرجات حرارة 2-3 °م أعلى من درجة الانصهار للحصول على التخصص ثم في الدورات المتأخرة تستعمل حرارة 2-3 °م اقل من درجة الانصهار.

وتستعمل البوادي غير المتماثلة في مجال بيئة الميكروبات ، اذ تسمح بتضخيم الجينات من الأحياء غير القابلة للزرع او لتضخيم الجينات من الأحياء التي لا تتوفر معلومات عن جينومها .

تصميم البوادي المشتركة

منذ ظهور عملية تضخيم الحوامض النووية (DNA) كانت اغلب العمليات تهدف الى تضخيم قطعة معينة او جين معين ، ولكن طرأت على هذه التقنية تطويرات وتحويرات كثيرة لتواكب الحاجة من البحث . ورافقت عمليات تطويع PCR ظهور التقنيات التي تستعمل المصفوفات الدقيقة للـ DNA التي أصبحت من التقنيات الأكثر استعمالاً في البايولوجي الجزيئي لتحديد حالة الانتساخ في مجتمعات الخلايا .

ومن المتوقع ان مجموعة من الجينات التي تظهر نمط تعبير متشابه في كائن واحد او عدة أنواع فأن احتمالية علاقة هذه الجينات من النواحي الوظيفية واردة جدا" وهي أكثر بكثير من احتمالية الصدفة ، فمثلا" الجينات المتناظرة التي تظهر نمط تعبير متشابه يكون من المحتمل انها تحوي على تواليات تنظيم ثابتة .

ومن مجمل هذه المناقشة يلاحظ ترشيح مصفوفات DNA كتقنيات واعدة لدراسة العلاقات بين الأحياء بغض النظر عن العلاقات التطورية بينها . ولكن المشكلة فيها ارتفاع الكلفة ، لذلك كانت المحاولات الى لإيجاد بوادي تضخيم كل او معظم الجينات في الجينوم . ومن هنا نشأت فكرة تصميم بوادي مشتركة لتضخيم جينات من جينومات مختلفة واعدت البرامج لهذا الغرض ، فمثل هذه البرامج تحدد التواليات ذات التشابه العالي ومنها تشتق البوادي المناسب الذي يكون مشتركا" بين اثنين او أكثر من الجينومات . او تصميم بوادي التي يكون بينها تشابه تواليات ليس بالكبير وتكون عائدة الى أكثر من جين في الجينوم نفسه .

وصفة هذه البوادي انها تكون قصيرة ويمكن ان تجد مكملاتها في جينات متعددة بحيث ترتبط قريبة من بعضها وبالترتيب الملائم ولكن ما يحصل عليه الحاسوب قد لا يطابق الواقع التجريبي ، ومع هذا فالمحاولات قائمة ، فقد أمكن تصميم بوادي لمناطق قصيرة عامة بين اثنين من الجينومات وليس بالضرورة ان تكون هذه المناطق في الجينات ذات التماثل الجذري . ولذلك فأن مثل هذه البرامج تتهاون بدرجة التطابق التي يمكن ان تنخفض الى 67% كما في البرنامج Dualprimer على مدى 30 نيوكلو تيد في الجينوم نفسه . وفي هذه الحالة يكون طول البوادي في الحد الأدنى مثل يعد 18 قاعدة الطول الأعلى وهو يمثل الطول الأدنى في منطقة التشابه التي يتم البحث عنها ، وفي مثل هذا الطول الأدنى تسمح البرامج بوجود حالة عدم تطابق في موقعين ليس أكثر .

وفي هذا المجال يمكن إضافة البوادي العامة Universal primers ليوازي الزيادة في تحديد التواليات وإيداعها في بنوك المعلومات اذ تضاف آلاف القواعد يوميا" ، والعديد من الجينات المهمة يتم تحديد توالياتها بشكل سريع كل هذه تحتاج الى دراسات ومعلومات حول هذه الجينات في الأنواع الأخرى التي تستعمل كموديلات دراسية ، فالبوادي العامة تفيد في الدراسات السريعة للجينات الجديدة ومثل هذه البوادي تصمم من تواليات الحوامض الامينية لمناطق البروتينات التي فيها اقل ما يمكن من الاختلاف في الشفرات الوراثية وهذه البوادي تصمم باستعمال الحروف ضمن تسميات IUPCA او غيرها من شفرات المنظمات الفاعلة بدلا من تخصيص القواعد بشكل دقيق جدا واستعملت مع جينات خلايا الحقيقية النواة مثل الطفيليات وبعض الحيوانات .

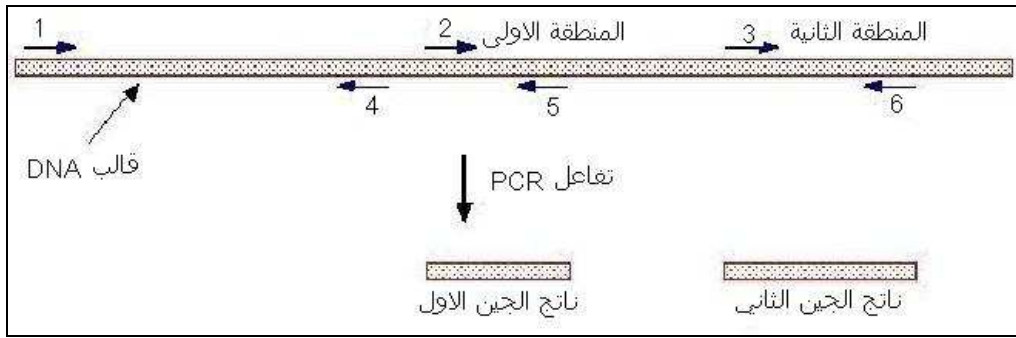
أنواع PCR واستعمالاتها

تعرضت التقنية الأصلية للعديد من التحويرات التي طالت مكوناتها الرئيسية من الإنزيمات والمحاليل الدارة ونوعية البوادي المستعملة والأغراض التي تستعمل لها . فالبوادي المصممة يمكن ان تعلم بالمواد المشعة او الصبغات المتفلورة او إضافة البايوتين او غيرها لإغراض محددة . ويمكن لنواتج PCR بعد إجراء التطفير او التحوير المرغوب فيه استعمالها وادخالها ضمن مجال التطفير الموجه Site-directed mutagenesis . كما يمكن استعمال النواتج بعد تحوير البوادي لعمليات الانتساخ والترجمة خارج الأنظمة الحية وفيما يلي ذكر لأهم أنواع PCR المستعملة على نطاق واسع ولكن في حقيقة الأمر توجد أنواع وتحويرات كثيرة لن يتم التطرق اليها وانما ما يذكر فيما يلي هو ماله علاقة عامة بالمعلوماتية الحيوية وبرامجها .

التضخيم العشوائي Random amplification polymorphic DNA (RAPD)

يسمى أيضا "التضخيم الصناعي"، ويستعمل لتضخيم مناطق عشوائية غير معروفة، على عكس أنواع التضخيم العادي التي تستهدف تضخيم مناطق محددة ومعروفة بعد اختيارها وتصميم البوادي الخاصة بها.

ولكن في حالة التضخيم العشوائي يتم تصميم بوادي عشوائية بطول حوالي 10 قواعد ويمكن ان يتم إيجاد هذه البوادي العشوائية بواسطة برامج حاسوب خاصة. ثم تستعمل لإجراء عملية التضخيم والكشف عنها باستعمال الترحيل الهلامي. وتظهر نواتج التضخيم في حالة ارتباط البوادي الى مناطق بطول معين من الجهة اليسرى واليمنى لقطعة ما، اما التي ترتبط الى مناطق بعيدة فلا تعطي نواتج كما موضح في الشكل 86.5



شكل 86.5 : التصميم العشوائي للبوادي

فالملاحظ من الشكل ان المنطقة الأولى الواقعة بين البادي رقم 2 و 5 يتم تضخيمها وكذلك المنطقة المنطق الثانية الواقعة بين البوادي 3 و 6، اما المنطقة الواقعة بين البادئين 1 و 4 فلا يتم تضخيمها لانها بعيدة عن بعضها مما لا يسمح بتفاعلات التضخيم، كما ان المناطق ذات الترتيب غير الصحيح لا يتم تضخيمها أيضا.

وعملياً "فإن البوادي القصيرة ترتبط الى القواعد تحت درجات صرامة واطئة في الدورات الأولى ويسمح بعدم التطابق بين البادي والقالب في عملية البدء، ثم تطبق ظروف صارمة بعد ظهور النواتج الأولية.

وتستعمل طريقة التضخيم العشوائي لإيجاد الاختلافات والمقارنات بين الجينومات التي تستعمل لها بوادي بطول 8 – 12 مقابل قالب DNA كبير أملاً في إيجاد المناطق التي سيتم تضخيمها، وتحليل نمط الحزم على الهلام يمكن عمل نسق للمناطق، وعند استعمال هذه التقنية لا تكون هناك حاجة لمعلومات حول DNA المستهدف لان البادي سيرتبط في أي مكان من التوالي يجده مكملًا له. والطريقة غير ملائمة لإنشاء بنوك معلومات للـ DNA وتستعمل في توصيف وتتبع العلاقات التطورية بين الأنواع المختلفة من النباتات او الحيوانات وبعض الجينومات البكتيرية. وقوة الإيضاح لهذه التقنية تكون اقل من حالة استعمالها مع قطع DNA مستهدف وكذلك عند مقارنة قطع DNA الخاصة من الأنواع مثل مكررات التواليات المترادفة Short tandem repeats (STRs).

ومساوئ استعمال التضخيم العشوائي تكمن في ان النواتج تكون متغلبة أي انه لا يمكن تحديد هل ان القطعة الناتجة هي متباينة أي وجود نسخة واحدة او من وجود نسختين، ولذلك فإن استعمالها

المختبري مزعج وتحتاج الى طرق خاصة لتعطي نتائج متطابقة لان عدم التوافق بين القالب والبادئ يمكن ان يؤدي الى عدم إعطاء نواتج او إنتاج كميات قليلة منها وبذا يصعب تفسير نتائج التضخيم العشوائي .

التضخيم العكسي Inverse PCR

تستعمل في حالات خاصة مثل تضخيم التواليات المحيطة بالعناصر القافزة وتحديد المقحّمات الجينومية Genomic inserts . ففي حالة عمليات التضخيم العادي فإن التواليات عند الطرف 3' و 5' للـ DNA الهدف تكون معروفة ، وباستعمال IPCR يمكن ان تسمح بالتضخيم عند معرفة توالي طرف واحد فقط مثل تلك الموجودة في حالة الإقحام وتقوم التقنية بتضخيم DNA مثل باقي التقنيات العاملة في PCR والبوادي تكون معكوسة بالنسبة للترتيب العادي ، وتستعمل فيها إنزيمات القطع الداخلية التي تؤدي الى معرفة التواليات على جانبي القطع غير المعروفة . والطريقة المستعملة تشمل :

- قطع DNA المستهدف الى قطع اصغر بواسطة إنزيمات القطع الداخلية
- إجراء عملية اللحم الذاتي تحت تراكيز قليلة ليتكون العمود الفقري من الفوسفات وإعطاء تركيب للـ DNA
- قطع DNA المستهدف (الحلقي) بإنزيمات قطع معروفة مؤدياً الى تكوين قطع مفتوحة ومعروفة النهايات والنواتج يمكن ان تستعمل في PCR

تضخيم النسخ العكسي Reverse transcriptase PCR

ويسمى أيضاً "تضخيم RNA ، او rtPCR وتشمل العملية خطوتين أساسية في الخطوة الأولى وتسمى تفاعل الشريط الأول لإنتاج cDNA من mRNA باستعمال إنزيم النسخ العكسي وفي الخطوة التي تليها ترتبط البوادي الى polyA sequence 3' عند النهاية غير القابلة للترجمة 3'UTR من جزيئة mRNA والتي تكون موجودة في اغلب هذه الجزيئات . وهذا يعني انه في الخطوة الأولى يصمم بادئ ليقابل مكوثرات الأدنين ، وهذه الخطوة تأخذ بعض الوقت عند درجات معتدلة مثلاً 37° م . بعد هذه الخطوة يضاف إنزيم RNase H لهضم RNA من الأسرطة المهجنة الناتجة (RNA – cDNA) .

وبعد ذلك تبدأ خطوة تفاعل الشريط الثاني اذ يتم تضخيم cDNA ببوادي خاصة للتواليات المطلوبة وبذا يتم التضخيم والكشف عن تواليات mRNA التي كانت بنسخ قليلة العدد او نادرة . وتكون الطريقة حساسة لتقدير RNA من خلية مفردة ولذلك تعد من أكثر الطرق حساسية للكشف عن mRNA المتوفرة في الوقت الحاضر مقارنة بطرق تحديد كميات ونوعيات mRNA الأخرى مثل تحليل وصمة نورذن او طريقة RNase protection assay .

وتستمر عملية التضخيم بشكل تزايد بمعدل ثابت الى ان يتوقف التفاعل التزايدي ويصل الى حد الهضبة أي عدم حصول زيادة في النواتج ، وهذه الحالة تختلف بشكل كبير حتى بين مكررات النموذج الواحد ويبدو ان السبب يعود الى زيادة تركيز النواتج التي تتنافس للارتباط بالبادئ نظراً لان إضافة المزيد من الإنزيم او التلاعب بالعوامل الأخرى يكون له تأثير قليل ، ويحصل توقف الزيادة بعد عدة دورات من التضخيم ، وبطبيعة الحال ولغرض دراسة عملية التضخيم ودقتها يكون من الضروري جمع البيانات الكمية عن كل نموذج هو في حالة تفاعل تزايد من أطوار التضخيم التي تكون خلالها العملية منتجة .

التضخيم غير المتناظر Asymmetric PCR

تقنية PCR تستعمل في تحديد التواليات والتهجين مع المجسات عندما يراد تضخيم احد الشريطين المتكاملة والطريق تشبه عملية التضخيم العادي ولكن يضاف تركيز عالي لبداي الشريط المطلوب الذي يكون بمثابة قالب للدورات اللاحقة ، ونظرا " لبطئ العملية فهي تحتاج الى دورات حرارية أكثر ، ويمكن ان تستعمل البودائ المعلمة بالكبريت والفسفور لإظهار النتائج ، وتستعمل في تضخيم النماذج القديمة التي يستعمل DNA المايكروكوندريا والكلوروبلاست التي توجد منها مئات آلاف النسخ في الخلية مقارنة بنسخة واحدة او اثنين للجينات الموجودة في النواة .

التضخيم الموضعي In situ PCR

تقنية تضخيم قطع DNA في الخلايا والأنسجة المثبتة بعد التلاعب بالحوازر النضوحية للخلايا للسماح لمكونات خليط تفاعل PCR بالدخول الى الخلايا المثبتة على المواد الساندة مثل الشرائح الزجاجية . ويتم تحديد النواتج بطرق مباشرة او غير مباشرة مع الأخذ بنظر الاعتبار الاحتياجات اللازمة لمثل هذا التحويل .

وقد تدمج التقنية مع مؤشرات الكيمياء الخلوية والكيمياء المناعية . وقد كانت الحاجة ماسة لمثل هذه التقنية للكشف عن الأمراض الناتجة عن عطب في الجينات التي تكون بطيئة التطور ، ففيها يمكن تضخيم نسخة واحدة من الجين المعطوب وتشخيصه بعد التضخيم ، وبذا فأنها تساعد في تحديد الواسمات الخلوية والجينية وتحديد مواقعها في الخلايا المصابة من الأنسجة او نماذج الدم ويمكن ان يحصل تداخل لهذه التقنية مع تقنية تضخيم النسخ العكسي rtPCR للكشف عن RNA الذي قد يوجد بمستويات عالية من حيث عدد النسخ .

والتقنية المقاربة لها هي تضخيم الخلية المفردة الذي له تطبيقات كثيرة أهمها التشخيص قبل الولادة الذي يعتمد الى تحليل نماذج الدم من النساء الحوامل لتحديد الاضطرابات الوراثية المنديلية التي لها مسبب واحد او ناشئة عن اضطراب جين واحد ، وكذلك تستعمل في دراسة مشاكل التطور الحاصلة في الجوانب العصبية والمناعية فضلا عن استعمال التقنية في المجالات البحثية .

تضخيم العقدة Nested PCR

تقنية تستعمل في العادة لتقليل النواتج غير المرغوب فيها وتؤدي الى زيادة حساسية العملية بشكل كبير ، وتستعمل عندما تكون كمية DNA المستهدف قليلة جدا" او ان DNA المستهدف غير نقي ، وبهذه الطريقة تقلل الخلفية غير الضرورية وتزداد كفاءة الكشف عن الأهداف . ويستعمل مع النسخ القليلة من DNA (كما ذكر توا) التي تكون اقل من 100 نسخة وكذلك تستعمل مع التضخيم الكمي Quantitative PCR .

في هذه التقنية يتم تضخيم الجزيئات بالطرق العادية وفي الخطوة الثانية تستعمل بودائ خاصة بنواتج تفاعلات الخطوة الأولى . ولذا تصمم مجموعتين من البودائ ، المجموعة الأولى تكون بودائ خارجية والمجموعة الثانية هي البودائ الداخلية لنواتج الخطوة الأولى ، ويمكن إجراء العمليتين او الخطوتين دون فصل او تخفيف بينهما . وفي هذه الحالة يجب الاهتمام الشديد بتصميم مجاميع البودائ حيث يتم الابتعاد تماما" عن حالة حصول مزدوجات البودائ الخارجية والداخلية . ويمكن ان تستعمل الطريقة وتطبق مع ما يسمى بالوراثة الكيماوية Vectorette PCR او تضخيم الفقاعة المستعملة في تضخيم قطعة ما ، التي تستعمل فيها نيوكليوتيدات قليلة تكون غير متطابقة جزئيا" (من البادئ) تسمى Vectorette لأحد جوانب البادئ والآخر يزود ببادئ معروف التوالي

ونظرا " لعدم التطابق الجزئي للبادئ المخلوق كيمائيا" تظهر فقاعة ولهذا سمي بهذا الاسم وبعد الخطوة الأولى يمكن تطبيق تقنية تضخيم العنقدة المذكورة

التضخيم الكمي الحقيقي RT PCR

معظم التقنيات المذكورة أعلاه تؤدي الى تقدير نوعي للمواد المستهدفة ولكن تقنية Real time PCR تهدف الى تحديد الكميات فضلا " عن النوعيات ، ففي هذه التقنية يتم تحديد كميات نواتج التضخيم لكل نموذج ولكل دورة ، اذ يمكن باستعمال أجهزة خاصة قياس عدد النسخ في كل حالة ، ونتائج الأجهزة الخاصة تعتمد على انبعاثات أشارات خاصة من الصبغات المتفلورة . وتختلف الصبغات في آليات إصدارها للإشارات فالبعض تزداد الإشارة الصادرة عنها عند ارتباطها بجزيئات DNA المزدوجة وهذه قد تكون مضللة لأنها يمكن ان ترتبط الى مزدوجات البادئ مع القالب لذلك يجب العناية باختيار الصبغات المستعملة .

وتعتمد على الوقت المحدد أثناء عمليات الإعادة والتدوير Cycling وتحدد أول نواتج عملية التضخيم دون الاعتماد على النواتج النهائية بعد انتهاء الدورات الحرارية وتسجل بالاعتماد على ما يشع من الصبغات المتفلورة أثناء تفاعلات PCR التي تكون دليل على كمية التضخيم التي تحدث أثناء دورات PCR وبذا يمكن ملاحظة التطور الحاصل وإظهاره ، وتكون أفضل الحالات أثناء عمليات التضخيم التزايدية .

وتستعمل طريقة تحديد الكميات مع تقنيات التضخيم العكسي وعندها يمكن تحديد الكميات بأكثر من طريقة :

- طريقة المنحنيات القياسية وفيها يتم تحضير منحنى قياسي من RNA بتركيز محددة ويستعمل كمرجع لتحديد تراكيز RNA غير المعروفة من النماذج . وهذه المنحنيات يمكن ان تختلف فيما بينها ، وبعضها يحتاج الى تحضير cDNA والمنحنيات القياسية يمكن ان تحضر أيضا " من DNA النقي مثل استعمال أشرطة DNA البلازميدية المزدوجة النقية وتحديد تراكيزها بالقياس عند طول موجي 260 نانومتر والنتائج يحول الى عدد نسخ اعتمادا " على الوزن الجزيئي للنموذج المستعمل . والأفضل استعمال cDNA للتخلص من التغيرات الحاصل في اختلاف التعبير ويمكن تصحيح النتائج بالرجوع الى جينات الإدامة .

- الطرق الثانية هي طرق المقارنة للنماذج مع نماذج غير معاملة من RNA الأنسجة الطبيعية وتصحح أيضا " بالنسبة لجينات الإدامة .

وتكون الأجهزة الخاصة بالتقنية خاصة وحساسة وتحوي على وحدات ضوئية لقياس تهيج الصبغات المستعملة وجمع المنبعث منها وتحديدده ، وتزود في العادة مع برامج حاسوب للقيام بمعظم هذه المهامات . ولهذه التقنية أنواع أخرى .

وقد تكون الفكرة مشابهة qPCR الذي يهدف الى تحديد الكميات التي يجري تضخيمها اذ ان تقنيات PCR العامة تهدف الى الحصول على نتائج نوعية ولكن في بعض الحالات تستعمل qPCR لتحديد الكميات ومعرفة نسق التعبير عن المعلومات الوراثية . وتعتمد التقنية على إدامة حالة التضخيم التزايدية الذي يكون ملائما " لتحديد الكميات ضمن 20 دورة الأولى

ولكن في حالة البدء بكميات صغيرة فإن عدد الدورات هذا لا يكفي لإعطاء نتائج واضحة لذلك تحتاج الطريقة الى وسائل حساسة جدا " للكشف عنها ويمكن باستعمال التقنية تحديد كمية DNA او RNA بعد إجراء التحويلات اللازمة .

التضخيم المتعدد (MP PCR) Multiplex PCR

تقنية تستعمل لتضخيم أكثر من قطعة DNA من الشريط القالب في نموذج واحد وفيها يستعمل زوج من البوادئ لكل هدف . والأفضل تصميم بادئ يضخم أكثر من هدف أو أكثر من منطقة . وعلى العموم فالبوادئ المستعملة في خليط التفاعل المستعمل تكون ذات درجات التهام ونسب GC % متشابهة لارتباط الحرارة الأولى ودرجة الانصهار بالمحتوى من GC . كما انه يجب الأخذ بنظر الاعتبار طول النواتج المضخمة حتى يمكن التمييز بينها عند فصلها على الهلام . وتعاني التقنية من مشاكل عدة منها :

- البدء الخاطئ نتيجة لعدم الارتباط المتخصص الى DNA المستهدف
- تكوين مزدوجات للبوادئ المستعملة
- عدم قابلية فصل قطع DNA المضخم النقي على الهلام اذا كانت حركتها (وبالتالي حجمها) متشابهة.

وتوجد برامج خاصة لتصميم البوادئ لمثل هذه التقنيات . اذ تقوم بتصميم مجموعة من البوادئ Primer set (PS) وهذه تستعمل في التطبيقات السريرية والبيئة الميكروبية والكشف عن الاضطرابات الوراثية . ومثل هذه المجاميع من البوادئ تتصف بالمواصفات العامة للبوادئ الأخرى مثل طول توالي البادئ ونواتج التضخيم التي تكون بحدود 100-500 قاعدة . وحرارة انصهار البادئ الأيسر والأيمن بالحدود المعروفة الواقعة بين 58-65 °م والاختلاف بين البادئين يكون بحدود 3 °م ، والطاقة الحرة لخمس ثملات في الطرف 3' تكون بحدود 9 كيلوسعرة \مول . وفضلا" عن ذلك هناك بعض المؤشرات الأخرى التي يجب ان تؤخذ بنظر الاعتبار في MP PCR وهي وان ذكرت أعلاه الا انه يجب التشديد عليها وهي :

- عدم الازدواج بين البوادئ المستخدمة
- تشابه حرارة الانصهار لكل بادئ
- تخصص البادئ يجب ان يكون عاليا" لتجنب فشل عملية البدء
- تقيد وتحديد حركة المتضخمات (نواتج PCR) في الهلام لغرض فصل وتنقية النواتج بسهولة.

تضخيم المسافات الطويلة Long distance PCR

عند تضخيم قطع DNA الطويلة التي تصل الى 20-50 كيلو قاعدة تحتاج الى نوعية جيدة من نماذج DNA النقية والسليمة لغرض إجراءها. وعند تصميم البوادئ لمثل هذه القطع يتم التلاعب بدرجات انصهار البادئين ، فضلا" عن الحاجة الى التلاعب بإنزيمات الكوثر DNAPs واستعمال تشكيلات منها وبنسب مختلفة .

برامج تصميم البوادئ

يوجد العديد من البرامج الخاصة بتصميم البوادئ لإجراء عملية تضخيم قطع DNA لأداء دراسات مختلفة في البيولوجي الجزيئي، ويعد هذا التوجه من التوجهات المهمة في المعلوماتية الحيوية . وكثرة البرامج تشير الى أهمية الموضوع وكذلك تشير الى ان التصميم الصحيح يعد الأساس في نجاح تقنيات PCR. ولكن يجب التذكر ان كثرة البرامج وكثرة استخدام الحاسوب لا يمكن ان تحل محل العمل التجريبي وان كانت تطبيقات الحاسوب ساعدت وتساعد في تطور علم الحياة الجزيئي وهو ما أعطى بعدا" جيدا" للمعلوماتية الحيوية ونظرا" لان عملية تصميم البوادئ تكون عملية مرهقة فاحتاجت الى تقنيات الحاسوب ولكن من جهة ثانية فأن الزيادة المفرطة في

البرامج وتنوعها احتاج الى عمليات تقييم لأدائها وتحديد صلاحيتها قبل وبعد استعمالها في التجارب.

طبيعة برامج تصميم البوادي

ان عملية تصميم البادئ من وجه نظر الحاسوب تكون سهلة ، اذ تقوم البرامج بالمسح والبحث في توالي طويل لإيجاد تواليات ثانوية تلاؤم مؤشرات خاصة . وتستعمل البرامج اللغة الكيماوية العامة او المفردات اللغوية القياسية ماعدا بعض الاستثناءات والمقصود هنا استخدام الأحرف الأبجدية A , C , T , G او N او غيرها في كتابة البرامج . كما ان اغلب برامج الحاسوب تستعمل مؤشرات وخوارزميات حسابية عامة ويتم بواسطة هذه الحسابات تكوين تشكيلات عشوائية تعتمد على التجربة والخطأ" ثم تمحص هذه التشكيلات لإيجاد الأفضل من البوادي الأيسر والأيمن لقطعة DNA المزمع تضخيمها . وهناك بعض النقاط الممكن ذكرها بشكل مختصر حول برامج تصميم البوادي .

● العديد من البرامج توجد على شبكة الانترنت العامة او مزودات الخدمة الخاصة ببعض الجامعات والتي يمكن الدخول اليها واستعمالها وتصميم البوادي سواء كان باستعمال تواليات الحوامض النووية او البروتينات .

● البرامج الجيدة هي التي تستفيد من المعلومات المتوفرة فبعض البرامج تدمج نظرية الحركية الحرارية Thermodynamic theory في التصميم وهذه تعد نقطة مهمة في حسابات درجات الحرارة والطاقة الحرة المار ذكرها .

● بعض البرامج يمكن ان تكون مجانية وأخرى مقابل ثمن ، كما ان بعض البرامج تحدد نظام الحاسوب الذي تعمل عليه مثل Linux او Windows او Macintosh ، وفضلاً عن تحديد جيل معين من الحاسبات وتحديد مواصفات الحاسوب المستخدم مثلاً " RAM وUPC وغيرها من المواصفات

● البرامج يمكن ان تستعمل عند الارتباط مع شبكة الانترنت Online وهي الغالبية العظمى والقليل منها يعمل بشكل مستقل Off line software او Standalone software . كما ان البرامج يمكن ان تكون ضمن الخدمات الذي يقدمها مزود الخدمة او الموقع او تكون ضمن حزمة من البرامج التي تستعمل لأغراض أخرى .

وتختلف إصدارات البرامج وفي العادة تكون الإصدارات الحديثة حاوية على تحسينات للإصدارات القديمة ، تحتفظ بعض المواقع بالإصدارات القديمة التي تبقى فعالة وقابلة للاستخدام وفي أحيان أخرى تختفي بعض الإصدارات القديمة من الواجهة .

● تختلف صيغ الملفات التي تقبلها البرامج ، فالبعض يقبل أي صيغة ويقوم البرنامج بتشذيب المدخل مثل إهمال الحروف الصغيرة او غيرها من الحالات التي تحتاج الى إزالة . اما البرامج أخرى فتحدد صيغة الملف المدخل وأكثر البرامج تقبل صيغة FASTA format ، ولكن هناك برامج مصممة بحيث تحتاج الى المزيد من المعلومات والتهميش وهذه تفيد معها صيغة بنك المعلومات GenBank format لأنها الصيغة الأكثر توفيراً للمعلومات (كما ورد سابقاً) ، في حين ان بعض البرامج تشترط ان يكون توالي الاستعلام موجود ضمن ملفات خاصة مثل Notepad ، وعند وجود محددات في شروط قبول الصيغ او غيرها من المتطلبات تزود البرامج بروابط تساعد المستخدم للوصول الى المتطلبات

• بعض البرامج تعطي تصميم بواى قياسية وأخرى غير قياسية أو غير مضبوطة Degenerated primers ولكن تقوم بخطوات لاحقة من التأكد من كفاءة البادئ بإجراء عملية الاصطفاف باستعمال برامج خاصة مثل BLAST .

• بعض البرامج تكون مرنة الاستخدام فهي تقدم تصميم للبادئ أصلي وفق المعايير التي حددها البرنامج (Default) ، ولكن تسمح للمستخدم التلاعب الى حد معين بالمؤشرات التي يرغب بها. ومثل معطيات هذه البرامج تكون المعايير الأصلية مصممة لإعطاء التصميم الأمثل الخالي من الاضطراب Noise وذلك بتقليل احتمالية البدء العشوائي Random priming .

• البرامج الجيدة هي التي تدمج أكثر من برنامج أو وسيلة لأداء مهامها فمثلا " على سبيل المثال لا الحصر البرنامج المرفق في NCBI لإيجاد تصميم البواى فهو يدمج بين برنامج Primers وهو أشهر برامج تصميم البواى مع عمليات الاصطفاف عن طريق برنامج أو خوارزمية BLAST .

• هناك برامج يمكن ان تتقبل أكثر من توالي استعمال لتحديد البواى لأغراض خاصة ، وتختلف البرامج في طول التوالي المدخل اليها فالبعض لا يقبل أكثر من 5000 قاعدة في حين أخرى تستوعب أكثر من 10 كيلوقاعدة .

الأسس العامة لبناء برامج تصميم البواى

البرامج الموجودة والتي ستظهر مستقبلا" يجب ان تأخذ بنظر الاعتبار المعلومات التجريبية . واغلب البرامج الموجودة في الوقت الحاضر تأخذ بنظر الاعتبار المؤشرات العامة المذكورة أعلاه . فضلا" عن ان بعض البرامج تضيف لها مؤشرات أخرى خاصة بأنواع PCR المرام عمله واغلب البرامج تقوم بإيجاد عدد من البواى ثم تجرى عليها عمليات تصفية وفق القيم الأصلية للبرنامج أو وفق طلب المستخدم . وبعد عمليات التصفية تظهر البواى مرتبة وفق درجات ترتيب خاصة تحسب من المؤشرات المستعملة وخاصة تلك المعتمدة على أساس الطاقة . ويمكن ذكر بعض هذه الأسس بشكل مختصر حيث مر ذكر بعضها فيما سبق من مناقشة ومنها :

• بعض البرامج تعطي توالي لبواى تلاؤم قطعة من البروتين بعد إجراء عملية الترجمة العكسية لتوالي الحوامض الامينية كما يحصل في تصميم البواى غير المنضبطة وهنا لابد ان تكون هناك مرونة في توالي البادئ ليواكب الاختلاف في الشفرات الوراثية فضلا" عن استيعابه للمفردات اللغوية لوحداث بناء الجزيئات وفق IUPCA.

• هناك برامج تقوم بمسح وتحليل التوالي المدخل وإعطاء التصميم الملائم والبعض منها يوفر معلومات حول طبيعة البادئ المصمم من حيث محتواه من GC % ودرجات حرارة الالتحام والانصهار والطاقة الحرة لكل من الطرفين 3' و 5' والوزن الجزيئي وحالات ظهور ماشيات الشعر وتكوين المزدوجات فضلا" عن التوالي نفسه وغيرها من المعلومات لتساعد المستخدم في اختيار الأفضل .

• يعتمد بعض مصممي البرامج الى إدماج أكثر من برنامج لتصميم بواى لعمليات PCR معتمدة مثل الكوثر المتعددة Multiplex PCR أو Nested PCR أو البواى الخاصة بالاكسونات . ثم يقوم البرنامج بفحص البواى الناتجة لتجنب المشاكل من تكوين مزدوجات البواى أو التراكيب الصليبية الشكل أو ماشيات الشعر وغيرها.

وكذلك الحال مع تصميم البواى الخاصة بعمليات مقارنة أنواع الجينومات وفيها يتم تجنب البواى التي فيها حالات تشابه عالية .

- بعض البرامج تصمم لاختيار أفضل المؤشرات عند التصميم مثل استعمال الطاقة الحرة ΔG لأنها أكثر مصداقية من درجات التطابق في حين ان البرامج العامة والمتوسطة الجودة تستعمل معادلة والاس لتحديد درجات الانصهار .
- العديد من البرامج تضع مؤشرات لقبول تصميم البوادي وقيم الفصل Cutoff لتحديد بعض المؤشرات مثل حالات ازدواج البوادي وغيرها.
- البرامج الخاصة بالحالات الدقيقة مثل حالات التشخيص والتعرف على SNPs وغيرها والتي يكون أساسها اختلاف في قاعدة واحدة من توالي DNA مثل هذه البرامج تقبل التواليات المضبوطة فقط وتقل كفاءتها اذا وجدت حالة عدم تطابق وان كانت واحدة .
- البرامج الخاصة بتصميم البوادي لحالات الوراثة اللاجينية تكون متخصصة جدا" وذلك لانه في هذه الحالة يكون تصميم البادي صعب جدا" نظرا" لوجود نواتج PCR غير مرغوب فيها وزيادة حالات عدم البدء الصحيح والتضخيم غير المتخصص نظرا" لكثرة T في التوالي المستهدف نتيجة للمعاملات التي تجري على مادة DNA خاصة عند استعمال مكنون النسخ للخلايا حقيقية النواة مثل الحيوانات الراقية والنباتات المهمة كما في حالة دراسة DNA الحاوي على المثل الذي يتم تحديد تواليه باستعمال Bisulphite ، ولذلك تقوم هذه البرامج بتحديد مواقع عدم البدء الصحيح .

تقييم البوادي الناتجة

تقيم البوادي المصممة بالحاسوب بالعمل التجريبي كأفضل طريق ولكن قبل القيام بالتجارب الفعلية يمكن مقارنة البادي المصمم مع مؤشرات مرجعية خاصة بالبوادي مثل حرارة الانصهار T_m ومحتواه من GC وغيرها من المؤشرات المذكورة سابقا" ، كما يؤخذ بنظر الاعتبار طول البادي في التقييم اذ ان جزءا" من التخصص يعتمد على الطول وبالنظر الى توالي البادي يمكن تجنب البوادي الحاوية على تكرارات كثيرة من القواعد لأنها ستظهر نواتج تضخيم DNA بشكل مسحات على هلام الفصل خاصة في حالة DNA الجينومي ولكن مثل هذه يمكن التغاضي عنها بعض الشيء عند تضخيم جين واحد لأنها ستعطي حزمة واحدة على هلام الفصل . فضلا" عن إجراء عمليات الاصطاف مع بوادي قياسية مثل تلك الخاصة بجينات الإدامة (على سبيل المثال) وتحديد الدرجات النهائية بعد طرح الجزاءات لتحديد صلاحية البادي .

ولتجنب التداخل وزيادة تخصص البادي تخضع البوادي لعمليات صف باستعمال BLAST وبعض البرامج تكون هذه العملية ضمن تصميمها أي Built-in والا وجب إخضاعها لعملية اصطاف منفصلة ومنها يتم اختيار التواليات الفريدة وليس لها شبه كبير مع القطع المراد تضخيمها . ويحدد ثبوت البادي بقياس عمليات البدء الكاذبة False priming. أي انه يحوي على طرف 5' ثابت ونهاية 3' ثابتة ، اذ ان الاختلال في الطرف 3' يمكن ان يؤدي الى ارتباط البادي الى مواقع مكملة له في التوالي غير مستهدفة وربما يؤدي الى تكوين حزم ثانوية . ولذلك كان من الكفاية وجود G او C في القواعد الثلاث الأخيرة من الطرف 5' لعمل ارتباط بين البادي والقطعة المستهدفة وهذه تمنع تكون الحزم الكاذبة .

ويمكن ان تقل كفاءة البادي بتأثير درجات الحرارة ، فعملية التضخيم تتم في العادة بدرجات حرارة أكثر من 50 °م ، ولكن تقل هذه التفاعلات وتحدث الاستطالة ويقوم البادي بعمل أخطاء غير متخصصة قبل بدء الدورات الحرارية فيما اذا تمت عملية خلط مكونات التفاعل بدرجة حرارة الغرفة ، لذلك يتم خلط مكونات التفاعل ماعدا إنزيم الكوثره مثل Taq ويتم مسخ الأشرطة المزدوجة ثم تحفظ في حمام ثلجي لبعض الوقت ويضاف إنزيم الكوثره بعدها لبدء التفاعل وكما

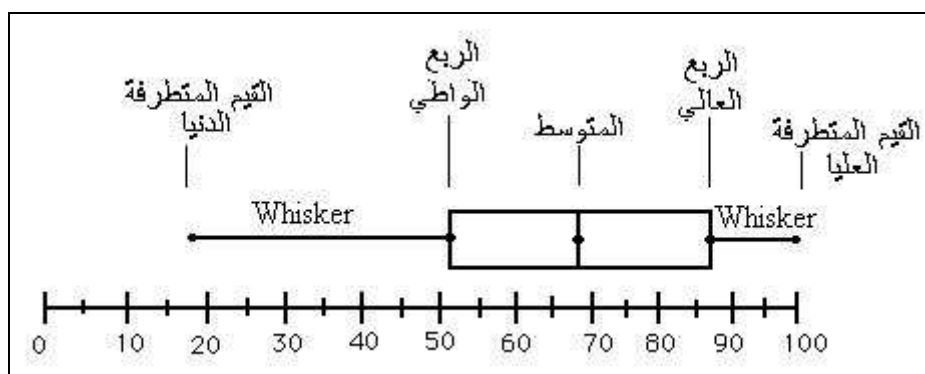
هو الحال مع تطبيق البرامج والوسائل في مجال المعلوماتية ينبغي على المستخدم ان يكون على دراية كاملة بالبرنامج او الوسيلة التي يستخدمها بالرجوع الى المراجع الأصلية التي تمخضت عنها البرنامج لمعرفة نقاط القوة والضعف وفيما اذا كانت النقاط التي بني عليها البرنامج تلاؤم توجهاته فضلا عن الأمثلة التي تضرب لاستعمال كل برنامج وإعادة تجربتها ليتم التأكد من صحة وملائمة البرنامج للغرض الذي يستعمل من اجله .

الجوانب التطبيقية الأخرى للمعلوماتية الحيوية في مجال الجينومات

بعد استعراض بعض الجوانب وخاصة المذكورة توا (بواى التضخيم) التي كانت مطية لبدء تجارب التضخيم ، تلعب المعلوماتية الحيوية دورا " مهما" في جوانب بحثية أخرى سيتم استعراض بعضها في هذا الجزء ، فهناك أكثر من 20 طريق لمقارنة الجينومات من كافة النواحي . البعض منها يعتمد على النواحي الإحصائية وأهمها استعمال

[رسم Box and whisker plot

وتسمى اختصارا " رسم الصندوق ، وهي نوع من رسوم Histograms لتوضيح البيانات وعلاقتها مع بعضها . والشكل التالي (87.5) يوضح طريقة رسم الصندوق



شكل 87.5 : طريقة رسم الصندوق Box and whisker plot

حيث يرسم مستطيل او صندوق عند المتوسط (M) Median 50% من البيانات للقيم المستعملة ويوصل بالقيم القيم العليا (Upper quartile 25%) والقيم الواطئة Lower quartile 25%) ومد الخطوط (Whisker) بينهما الى ابعد نقطة في القيم على ان لا تكون خارجية وتكون ضمن مدى معين من البعد عن المتوسط وبذلك يكون الرسم موضحا لتوزيع البيانات اما النقاط البعيدة فترسم بشكل نقاط منفصلة (Outliers) وتعرف على انها بيانات بعيدة وقيمها أكبر من 1.5 مرة من مدى المستطيل (الذي يمثل المتوسط).

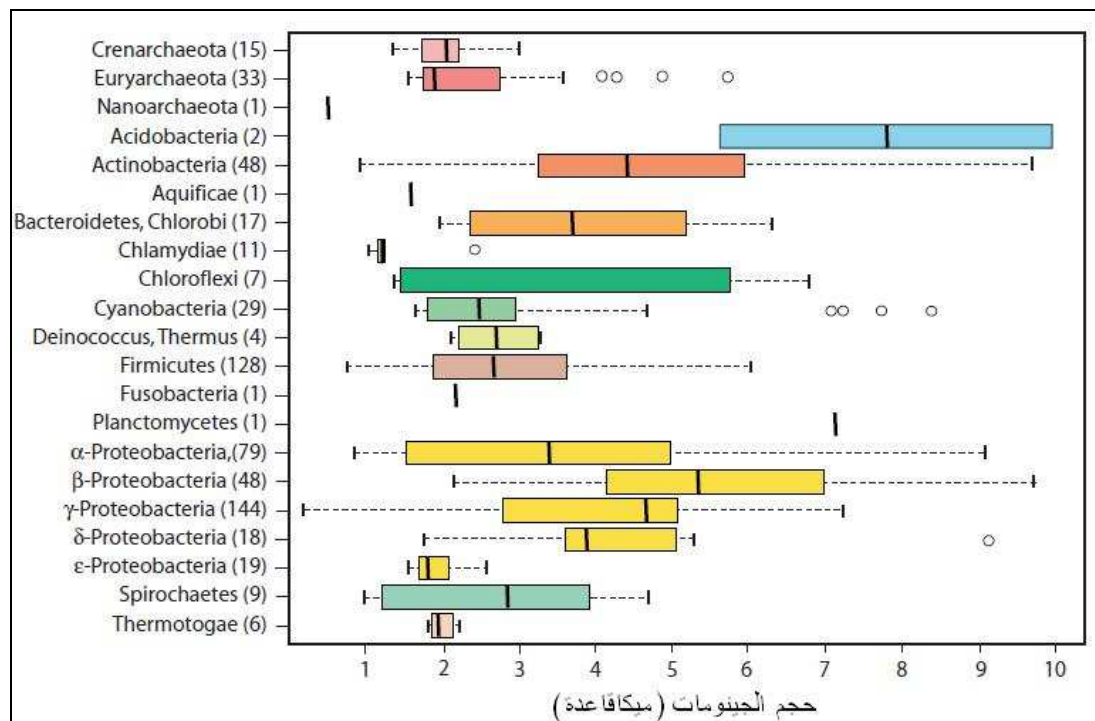
مقارنة الجينومات

المعلومات المخزونة في تواليات الجينوم يمكن ان تحلل على مستويات مختلفة ، وتساعد التحليلات في الحصول على قيم رقمية مثل الحجم او مكونات الجينوم من القواعد النتروجينية التي يمكن ان تقارن بين عدد من الجينومات باستعمال الطرق الإحصائية ، وفي هذا المجال

يصبح التمثيل الرسمي من الضرورات الأساسية وبواسطة هذه المقارنات يمكن معرفة مناطق انتقال قطع الكروموسوم Translocation او الانقلاب Inversion او حدوث عمليات الإقحام والحذف ، وهناك عدة طرق يمكن بواسطتها مقارنة الجينومات وفق الخاصية المطلوبة من ذلك.

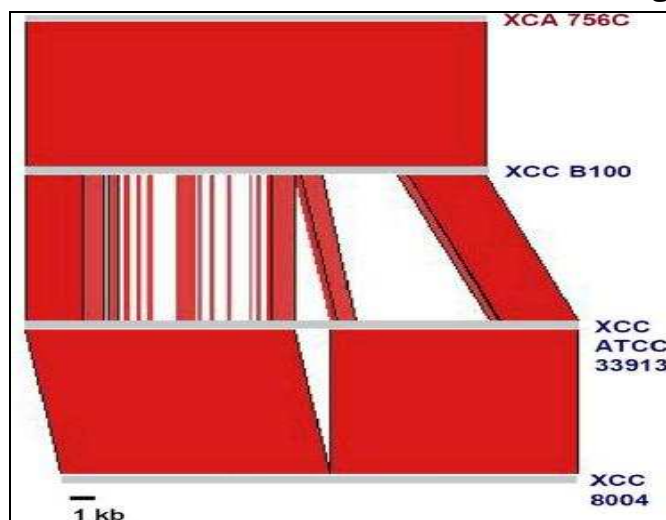
حجوم الجينومات وتشابهها

تختلف الجينومات في حجومها وقد مرت مناقشة بعض جوانب الموضوع سابقاً ولكن هناك بعض الجوانب الأخرى التي يمكن طرحها ، فمثلاً الجينومات الصغيرة التي تتراوح بين 490885 قاعدة في *Nanoarchaeum equitans* وتصل الى 9,105,828 قاعدة في *Bradyrhizobium japonicum* . وتختلف الحجوم بين السلالات المختلفة مثلاً في *E. coli* يمكن ان تختلف السلالات بمليون قاعدة ، وكذلك الحال مع *B. subtilis* التي يمثل 2.4 ميكاقاعدة الجينوم المركزي الثابت والباقي المتمثل بـ3 ميكاقاعدة من الكروموسوم يكون متغايراً . وفضلاً عن ذلك فإن السلالة الواحدة يمكن ان يحصل فيها التغيرات عند اختلاف ظروف التسمية . ويلاحظ ان الاختلاف في جينومات الخلايا بدائية النواة ذات الجينومات الصغيرة يمكن ان يصل الى 20 مرة ، في حين يصل الى 1,000,000 في جينومات الأحياء حقيقية النواة ، فمثلاً في الكائن المجهرى حقيقي النواة (طحلب) *Guillardia theta* يكون حجم جينومها 510,000 قاعدة ولكن بعض أفراد هذه المجموعة من الأحياء يصل حجم جينومها الى أكثر من جينوم الإنسان كما في *Amoeba dubia* الذي يصل الى 670,000,000,000 قاعدة (أي بمستوى التريليون) . وفضلاً عن ذلك فقد أوضحت دراسات مقارنة الجينومات ان المختزل منها يوجد في *Rickettsia* التي يعتقد انها الجد الأول للمايكوبلازما . وفي جميع الأحوال يكون المجموع الأدنى من عدد الجينات هو الضروري لحياة الكائن ، وحجم الجينوم يكون دليلاً غير دقيق على التطبع الحيوي ، وهناك عدة برامج يمكن ان تساعد في مقارنة الجينومات من هذه الناحية وقد تم حساب حجوم بعض الجينات وجدولتها والشكل التالي (88.5) يوضح حجوم بعض جينومات الأحياء .



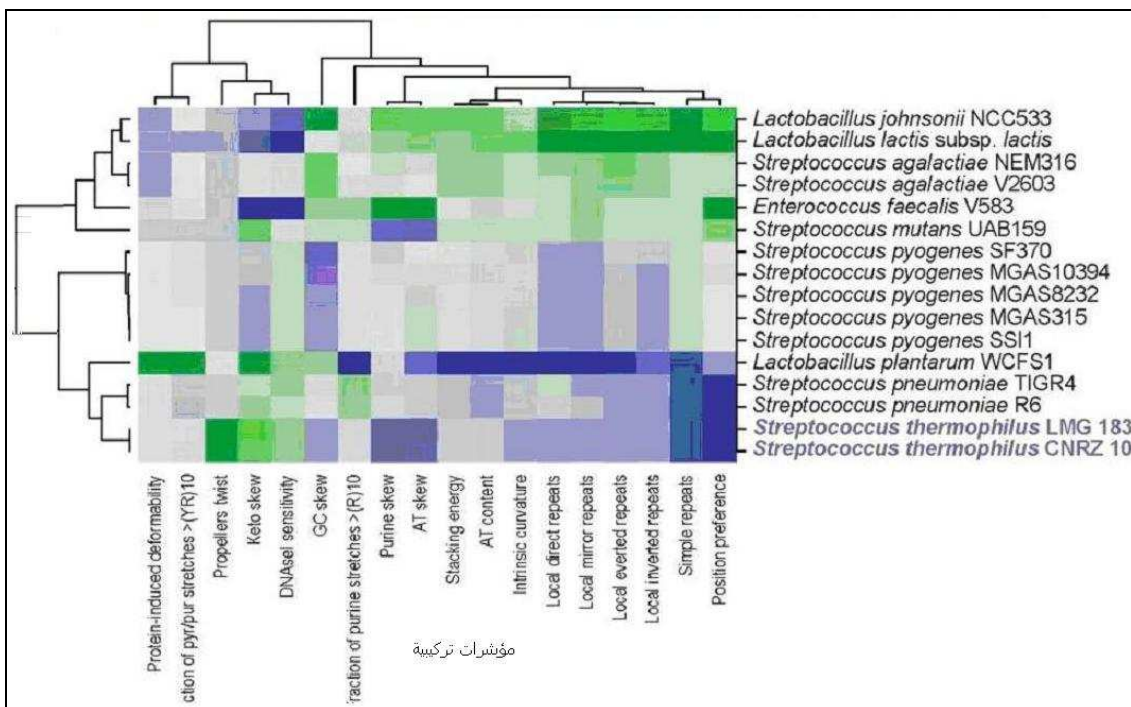
شكل 88.5 : حجم الجينومات لإحياء مختلفة

اما مقارنة التشابه على مستوى النيوكليوتيدات فيتم بشكل غالب باستعمال برامج BLAST . وقد يدمج الأخير مع طريقة Artemis comparison tool (ACT) لمقارنة عدة جينومات كما موضح في الشكل 89.5 :



شكل 89.5 : مخرجات برامج Artemis comparison tool لمقارنة الجينومات

2 D clustering of genomic properties العنقدة الثنائية لصفات الجينومات



شكل 90.5 العنقدة الثنائية لصفات الجينومات

والبرامج الخاصة لتحديد مثل هذه الصفات تكون مزودة ضمن أطالس الجينومات (التي سيأتي ذكرها لاحقاً) وتستعمل قواعد بياناتها .

ومن النواحي المهمة في مقارنة الجينومات العامة تحديد مواقع الجينات التي تستعمل أصلاً لتحديد مواقع الجينات على الكروموسوم الواحد في الأحياء حقيقية النواة ولكن يمكن ان تستعمل لمقارنة الجينومات في الأحياء الأخرى فموقع الجين Gene synteny يكون في الأحياء القريبة من بعضها متشابهة ولكنه يضطرب او يتغير عند حصول انقلاب أي تغيير اتجاه واحد او أكثر من الجينات او بعمليات الإقحام او الحذف ، وكل هذه يمكن ان تنتج من أخطاء أثناء التضاعف او تضاعف العناصر المتحركة مثل العاثيات البكتيرية او القافزات وغيرها . فالطفرات النقطية مع تغيير موقع الجينات تشكل القوى الغالبة لزيادة التنوع الوراثي الذي يقلله ويحد منه العيش في ظروف محددة .

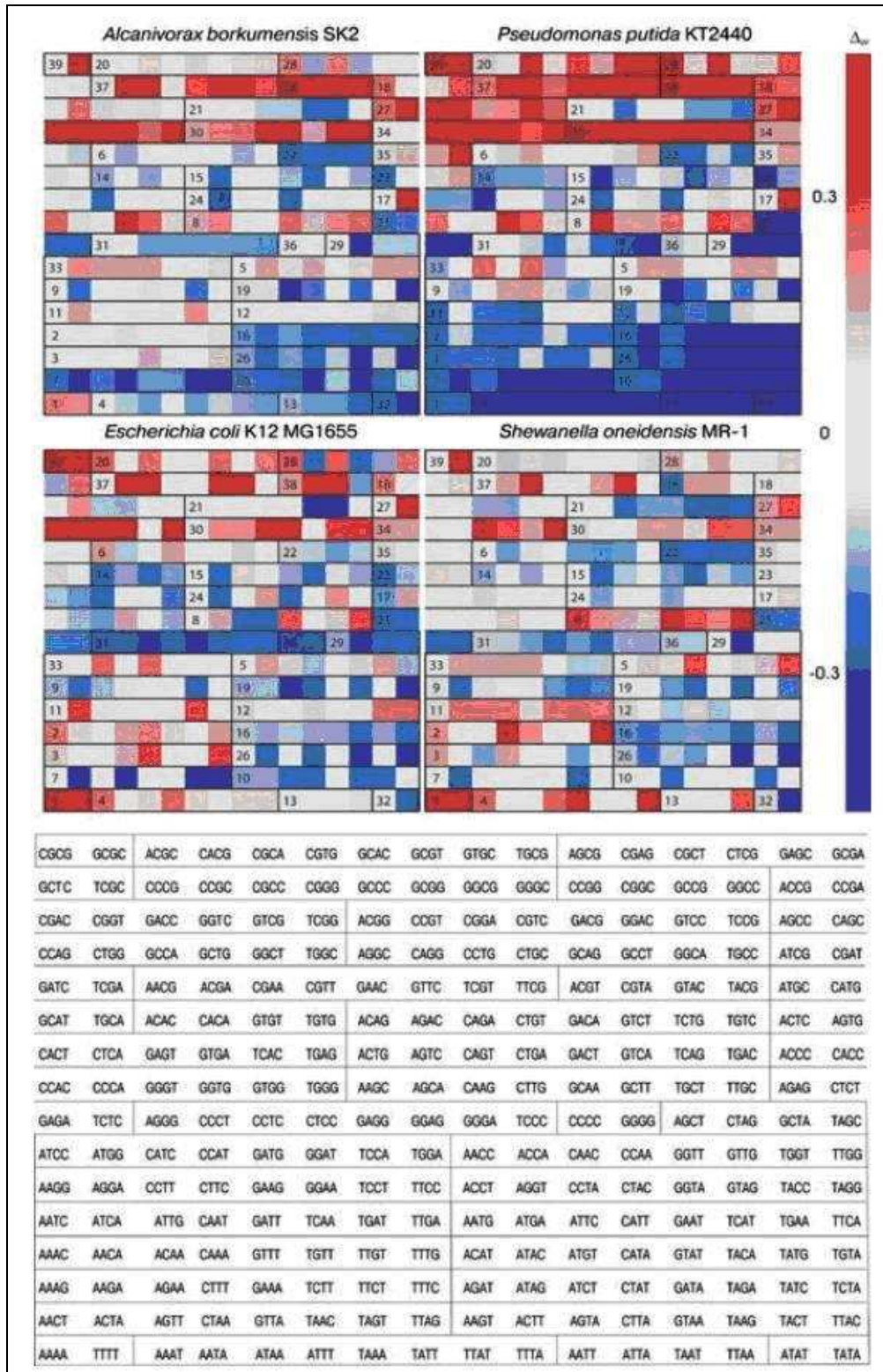
وفي هذا المجال تبرز مسألة الجزر الجينومية Genomic islands وهذه تحدث نتيجة للتطور وإعادة إدلاف الجينات او نتيجة الانتقال الافقي لتكون مواقع ثابتة نوعاً ما خاصة في الجينومات التي تكون المسافة الوراثية بينها معقولة . وهذه الجزر تحوي على جينات ذات علاقة ببعض الفعاليات مثل جزر الامراضية Pathogenicity islands او جزر التعايش Symbiosis islands وجزر الايض Metabolic islands مثل جزر ابيض الكبريت ، او جزر الجينات المسؤولة عن إنتاج عضيات داخل الخلايا التي تساعد البكتيريا في تنظيم خلاياها مع المجال المغناطيسي Magnetic field كما في Magnetotactic bacteria . ومثل هذه

الجزر يمكن ان تنتقل كاملة من كائن (بكتريا) الى أخرى لإضفاء صفة تساهم عند الدخول الى بيئات جديدة .

ومما يعيق مقارنة الجينومات بشكل مضبوط جدا" هو ان بعض الجينومات المودعة في قاعدة بنك الجينات والتي تشكل المرجع لمعظم المقارنات هو احتوائها على بعض الجينومات التي ليس لها موقع تصنيفي واضح وهذه كلها وضعت تحت مسمى Candidatus وخاصة للأحياء غير القابلة للزرع وهذه تأخذ أرقام تسجيل معينة ، ومثل هذه ينقصها بعض المعلومات المهمة حول النواحي البايولوجية والبيئية . ومستقبلا" يؤمل ان تتم معرفة التواليات وتحديد الجينوم المركزي Core genome الذي يمثل كل الجينات الموجودة في كل جينومات النوع الواحد وكذلك تحديد الجينوم العام Pan genome الذي يمثل كل الجينات الموجودة في كل الأنواع التابعة للجنس الواحد وتحديد الجينات غير الأساسية المسؤولة عن التغيرات بين العزلات .

مقارنة النيوكلووتيدات الرباعية Tetranucleotide usage patterns

يستعمل تكرار كل 4 نيوكلووتيدات في مقارنة الجينومات وهذه المجاميع يكون عددها $4^4 = 256$ وهذه النيوكلووتيدات الرباعية مرتبة في 39 صنف التي يكون لها تركيب متكافئ وتصنيفها هنا يعتمد على طاقة الرص . والبعض منها لبكتريا مختلفة موضحة في الشكل 91.5 :

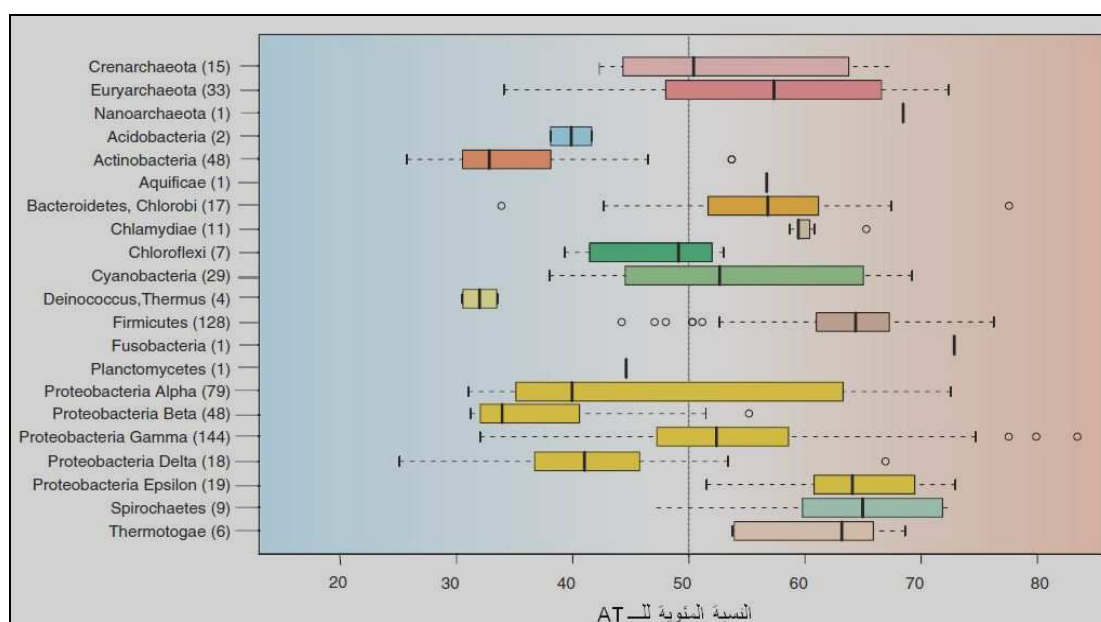


شكل 91.5 : ترتيب النيوكلووتيدات الرباعية المستعملة في مقارنة جينومات الأحياء المختلفة

الشكل يوضح نمط توزيع النيوكلوتيدات الرباعية لبعض البكتريات ، التي تبدأ بالصنف 39 من الجهة اليسرى العليا . ومثل هذا التوزيع يكون خاص بكل جينوم لذلك يستعمل في المقارنة.

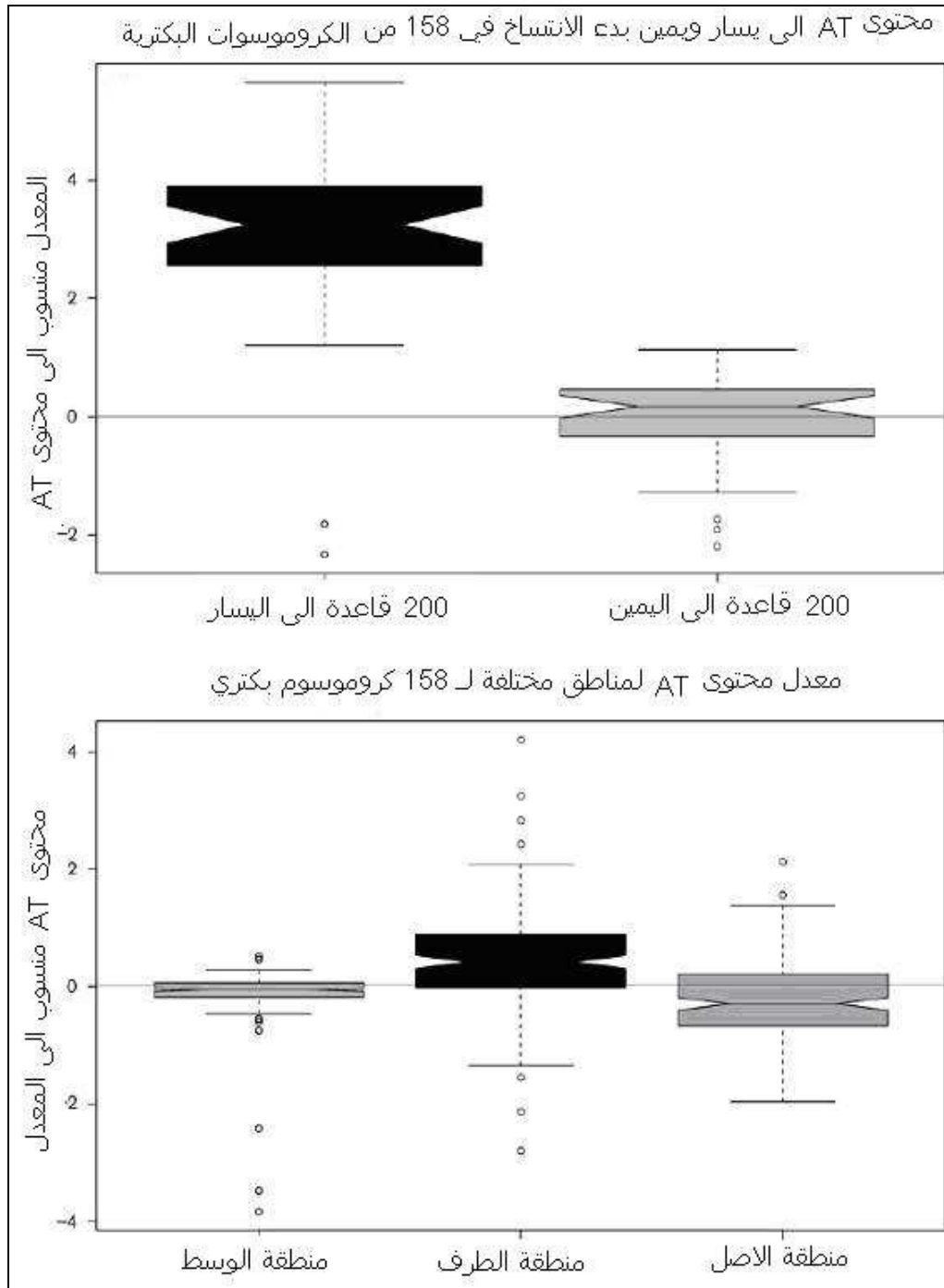
مقارنة محتوى الجينومات من AT

تختلف الأحياء في نسبة محتواها من AT فهي قد تكون منخفضة وتصل الى 27.9 % في *Streptomyces coelicolor* او تكون ذات مستويات أعلى كما في البكتريا المتعايشة داخل الخلايا مثل *Wigglesworthia glossinidia* التي تصل النسبة فيها الى 77.5 % . ويوضح الشكل التالي (92.5) محتوى بعض المجاميع من AT .



شكل 92.5 : الاختلافات في نسب AT بين الجينومات البكتيرية المختلفة

فالملاحظ ان الأجناس التابع لمجموعة Firmicutes تكون غنية بـ AT في حين تكون الأجناس التابعة لمجموعة Actinobacteria غنية بـ GC . ولكن الملاحظ ان توزيع AT يختلف باختلاف منطقة الجينوم كما موضح الشكل 93.5 :



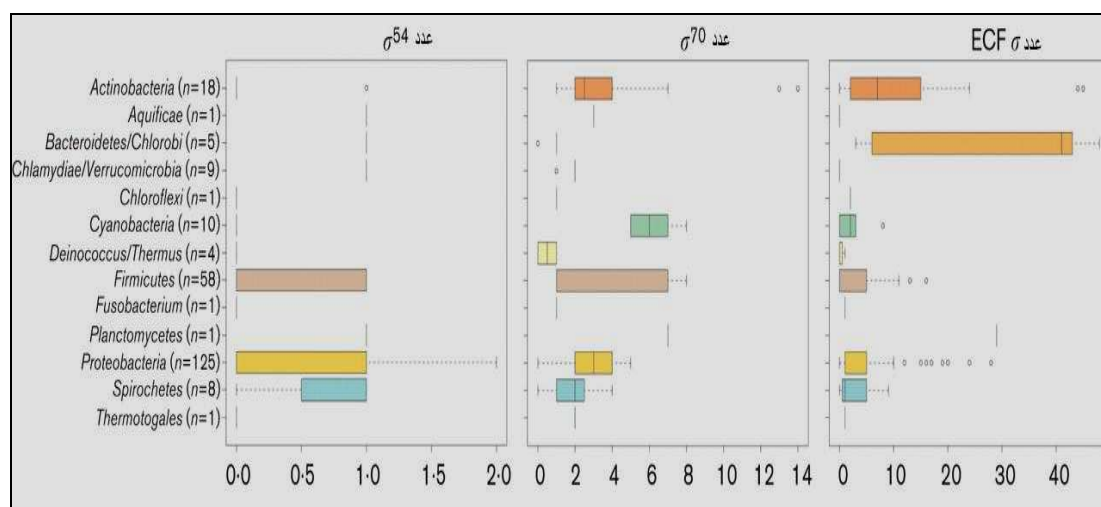
شكل 93.5 : توزيع AT على مناطق مختلفة من الجينوم

ان اختلاف توزيع AT على الجينومات البكتيرية لا يشمل المناطق المشفرة فقط وانما يشمل مناطق أخرى ، منها بشكل رئيس نقطة بدء التضاعف الذي يتوقع ان تكون غنية بـ AT ولكن وجد انها تكون غنية بـ GC مقارنة بباقي مناطق الجينوم واحد التوضيحات البايولوجية لهذه

الظاهرة هو لمنع عمليات بدء تضاعف كاذبة ، اما منطقة انتهاء التضاعف فتكون غنية بـ AT وبالتالي تحوي على DNA مقوس أكثر . وإضافة الى ذلك فإن الاختلاف في محتوى AT يمكن ان يطول المناطق البيئية التي وجد ان هناك 5% زيادة في محتوى AT مقارنة بمناطق الجينات ، وهذه الحقيقة بالنسبة لجينومات البكتيرية التي درست سواء كانت غنية AT او فقيرة . اما مناطق جينات rRNA فتكون فقيرة بـ AT ، ومن المظاهر الأخرى هو غنى الجينات بـ AT المشفرة للبروتينات التي تصدر الى خارج الخلايا مقارنة بعوائل الجينات الأخرى . وقد وجد ان الجينات الموجودة على البلازميدات تكون غنية بـ AT . ويظن البعض ان الجينات التي لها تركيب مختلف عن التركيب العام لجينوم بكتيرية ما يمكن ان تكون اكتسب أفقياً" .

تحديد عواما سيكما

وهذه تكون خاصة بالجينومات البكتيرية ، وتمثل عوامل تساعد ارتباط RNAP الى توالي محدد في الممهدات البكتيرية . وعدد الجينات المسؤولة عن عوامل سيكما (σ factors) يختلف وفق الظروف البيئية وتغايرات النوع الواحد . ويوضح الشكل التالي (94.5) عدد أنواع عوامل سيكما المختلفة σ^{54} ، σ^{70} ، ECF لكل جينوم من الأحياء



شكل 94.5 : عدد عوامل سيكما في البكتريات المختلفة

والملاحظ ان العامل سيكما 54 قد يكون غير موجود او يكون بقيمة 1 لكل جينوم وان كانت هناك بعض الشواذ التي يصل فيها وجود 2 من العامل 1 جينوم . اما بالنسبة للعامل 70 والذي يعد من الأساسيات فيكون هناك وفرة من الجينات المسؤولة عنه ، اذ يشترك في العديد من الفعاليات للبكتيريا ، اما عوامل سيكما Extracellular factor (ECF) فهي قد تكون غير ضرورية لعدد من الأحياء لذلك تفقد ، ولكنها وكما يلاحظ من الشكل تكثر في *Bacteroides* و *Streptomyces* اذ يمكن ان تصل الى 40 عامل اجينوم . اما في حال *Firmicutes* فأغلب الأجناس لا تحوي على هذه العوامل وهذا له علاقة مع وجود عدد من عوامل سيكما

70 ويوضح الشكل بعض الشواذ العامة في عدد العوامل البعيدة عن المتوسطات والتي تظهر بشكل نقاط خارجية (o) .

دورية القطيفات Motifs periodicity

ان وجود تردد نوع من القطيفات بشكل دوري يمكن ان يشير الى ارتباط البروتينات المنظم للـ DNA او إشارات لتقوس DNA . وهذه الدورات المكونة من 10 - 11 قاعدة لها القابلية على إضفاء صفة الانحناء Bendability والتقوس لجزيئات DNA ، فهناك Poly (T) و Poly (A) في تواليات DNA تعكس تركيب DNA في الكروماتين ، وقد أشير الى ان حلزونات ألفا من البروتينات تحت وجود دوريات في DNA وبدورات متشابهه ، والحث في هذه الحالة أي بواسطة البروتينات يشكل جزءا "بسيطا" من عمليات مؤثرة أخرى ، ومثل هذه الدورات توجد في الغالب في المناطق غير المشفرة من DNA ، ويشير وجودها الى العلاقة مع عملية الالتفاف الفائق للـ DNA في جينومات الخلايا حقيقية النواة . وتحديد عدد القواعد في هذه الدورات تعطى إشارات لعمليات الانحناء والالتفاف في الجسيمات النووية

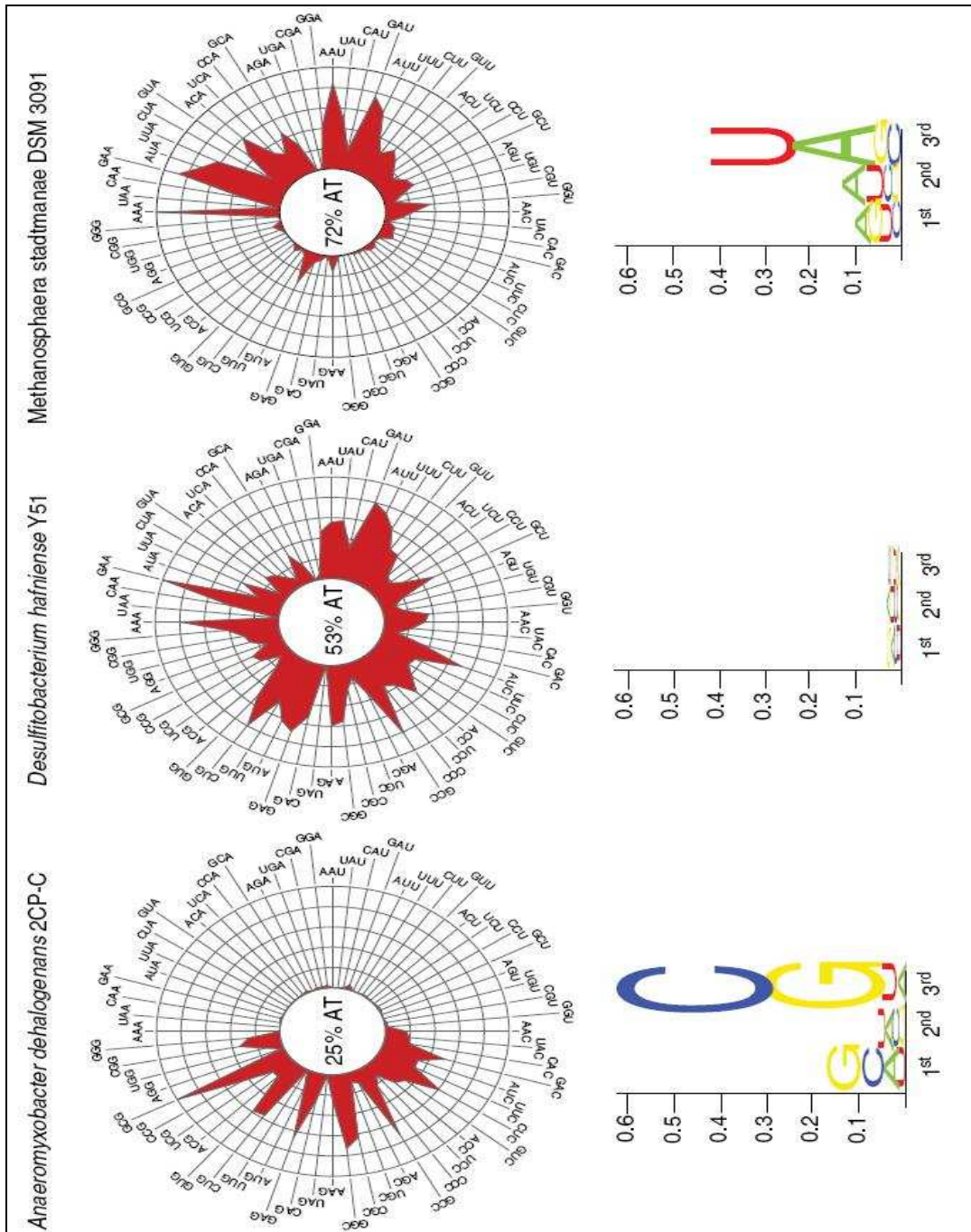
استعمال الشفرات Codon usage

يعد الاختلاف في استعمال الشفرات لإنتاج الحوامض الامينية من أوسع المواضيع التي حظيت بالدراسة لأنها تمثل نقطة إشكال في وضع البرامج وسلاستها ، فهناك 61 شفرة وحوالي 20 - 22 حامض أميني ، ووجود مبدأ المراوغة في عمليات الترجمة مما اثر في وضع البرامج والعلاقات الرياضية . ولكن الثابت ان استعمال الشفرات من قبل الأحياء ليس عشوائيا . وقد وضعت قاعدة بيانات خاصة بهذا الموضوع للتعريف بالانحياز في استعمال الشفرات لإنتاج الحوامض الامينية وتضم القاعدة كل الشفرات لكل الجينات الموجودة في قاعدة بيانات بنك الجينات الأكثر استعمالا" لكل كائن موجود في القاعدة . ويشير واقع الحال ان تواليات البروتينات المتناظرة أكثر تشابها" من الجينات المتناظرة المشفرة لها وهذا يعود بطبيعة الحال الى اختلاف الشفرات للحامض الاميني الواحد ، وهذا ينسحب عند إجراء الترجمة العكسية المستعملة لتصميم البودائ للبروتين نفسه ، فضلا" عن وجود تغاير في استعمال الشفرات بين الجينات المختلفة في الجينوم الواحد . وقد أشارت الدراسات وبعد تحديد التواليات للعديد من الجينومات ان التنوع الوراثي اكبر مما كان يظن وان الجينومات الكبيرة تكون غنية بـ GC وان الجينومات الصغيرة تكون غنية بـ AT فضلا" عن إيضاح ان ثملات الكوانين تكون هي الأكثر في الأشرطة القائدة وان توزيع ثملات الأدينين في الأشرطة القائدة تكون خاصة بالنوع وهذا كله يؤثر في تفضيل استعمال الشفرات في الأنواع المختلفة وفي حقيقة الأمر فإن الانحياز في استعمال الشفرات يرتبط بعوامل بايولوجية أخرى منها :

- مستوى التعبير الجيني
- طول الجين
- إشارات بدأ انتساخ الجين
- مكونات البروتينات من الحوامض الامينية وتركيبها
- وفرة جزيئات tRNA
- تردد حدوث الطفرات

- مكونات الجينوم من GC والانحياز في استعمال الشفرات يساعد في فهم تطور الكائنات الحية والتطبع البيئي والأساس في تحديد استعمال الشفرات يعتمد على تحليل الانحياز في استعمال الشفرات في الجينوم نفسه او بين الجينومات . ومن المعروف ان الحوامض الامينية يشفر لها بأكثر من شفرة ماعدا التربتوفان والميثايونين ، والباقي يشفر لها بشفرات مترادفة والتي لا تستخدم بتردد متساوي بين الأحياء ، وهذا يكون نتيجة الانتخاب الطبيعي أثناء عمليات التطور ويكون مرتبطا " بعدد من العوامل المذكورة أعلاه . ويتم تحديد استعمال الشفرات بأكثر من طريقة
- ❖ الطرق الإحصائية ويكون بقياس تفضيل استعمال الشفرات المعتمدة على حساب X^2 (Chi square) وتحليلها
- ❖ الطريق الثانية هو استعمال مجموعة من تواليات الجينات كمرجع .

وهناك عدد من البرامج الخاصة بتحديد استعمال الشفرات التي تكون مرتبطة بالانترنيت الى قواعد البيانات العامة وتكون نتائجها اما بشكل وردي وترفق معها نتائج برامج Logo فضلا عن إرفاق التحاليل الإحصائية ، ويمثل الشكل التالي (95.5) استعمال الشفرات في بعض الأحياء وتمثل الدوائر المتحدة المركز مدى تردد استعمال الشفرات .



شكل 95.5 : الاختلاف في استعمال الشفرات الوراثية في الأحياء المختلفة

وتواجه مسألة تحديد استعمال الشفرات وحسابها والرسوم الناتجة منها مشاكل وهو التباير في القاعدة الثالثة في الشفرة الوراثية ، والملاحظ من الشكل أعلاه ان مستوى GC له تأثير كبير

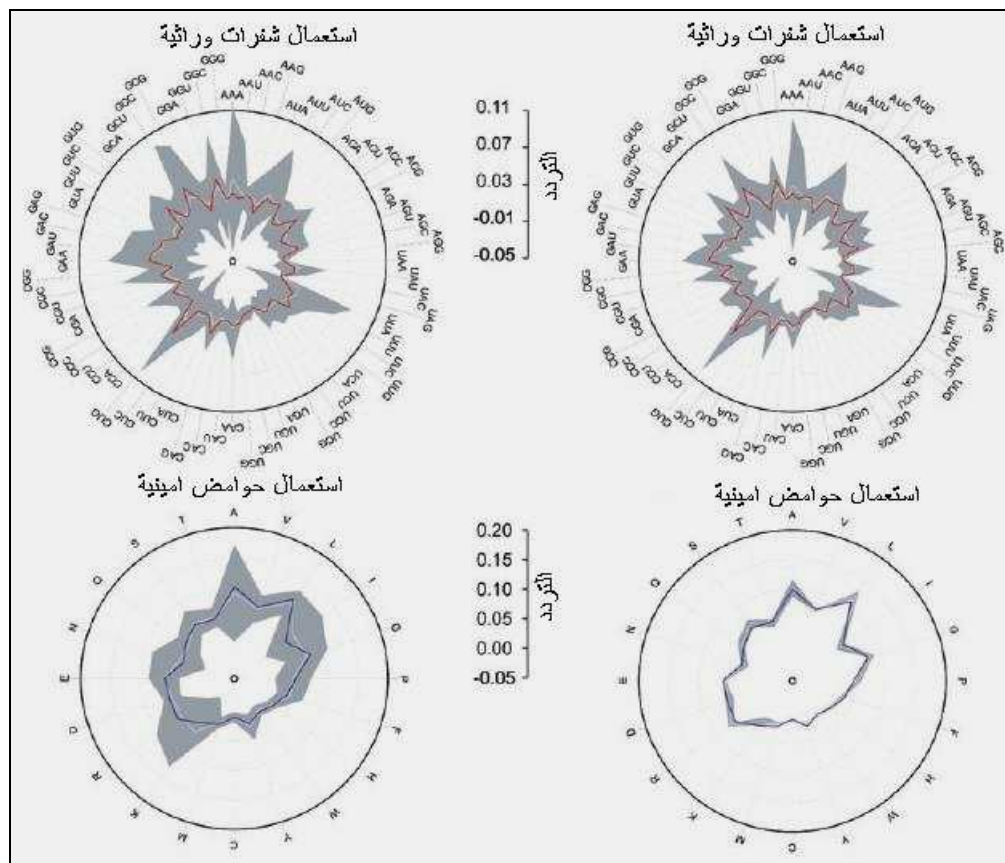
في استعمال الشفرات ومثل هذه الرسوم تساعد في استنتاج العديد من المعلومات مثل البيئة التي تعيش فيها الكائنات وغيرها من المعلومات .

ان استعمال الشفرات يكون مرتبطاً ارتباطاً وثيقاً بجزيئات tRNA وبالتالي الجينات المسؤولة عنها ، فمثلاً في البكتريا *Vibrio parahaemolyticus* السريعة النمو هناك 126 من الجينات الخاصة بـ tRNA والتي تقل لتصل الى 29 في *Mycoplasma pulmonis* ، وهذا يؤثر في استعمال الشفرات . وكما أشارت دراسة المثال الأخير (*M. pulmonis*) يكون استعمال الشفرات مختلفاً فمن الشفرات الست الخاصة بالحامض الاميني الليوسين يستعمل الكائن UUA بعدد 13272 مرة في حين يستعمل CUG 165 مرة أي اقل بحوالي 100 مرة وهذا له علاقة بمحتوى الجينوم من AT ، ففي هذا المثال فان الشفرة (CUN) تنحاز الى استعمال C او U او A بدلاً من (N) ، فالشفرة CUC تستعمل 399 مرة وهي منخفضة المستوى مقارنة باستعمال الشفرة CUU التي تصل الى 7523 والشفرة CUA التي تصل الى 3932 مرة . لذلك فالجينومات القريبة التشابه من حيث محتوى GC , AT ربما تشفر لحوامض امينية متشابهة وبالنتيجة تكون الحوامض الامينية الناتجة من الجينومات متشابهة ويكثر فيها الليوسين والالنين والكلايسين والسيرين ويقل فيها التربتوفان والميثايونين والهستدين .

ومن جهة ثانية فإن شفرات الوقف يمكن ان تشفر لبعض الحوامض الامينية في بعض الأحياء . ففي المثال أعلاه (*M. pulmonis*) فإن شفرة الوقف UGA تشفر للتربتوفان . وكذلك يشفر للحامض الاميني السستئين السيليني بشفرة الوقف هذه (UGA) في جينومات أخرى من البكتريا التي تخلق هذا الحامض . اما الحامض الثاني والعشرين المقترح Pyrrolysine فهو يستعمل شفرة الوقف الأخرى UAG .

لذلك يلاحظ الاختلاف الكبير في استعمال الشفرات للحامض الاميني الواحد بين الأحياء المختلفة (كما موضح في الشكل 54.3 السالف الذكر) يوضح هذه الحقيقة لثلاث من البكتريات التي تم تحديد توالياتها

وتلعب الألوان دوراً مهماً في إيضاح استعمال الشفرات اذ يمكن إدخال نظام التلوين للحصول على معلومات أكثر خاصة بالنسبة للقاعدة الثالثة في الشفرة الوراثية ويمكن استعمال مثل هذه الرسوم التي يطلق عليها او رسم الوردية او الرسم العجلي الشكل Rose plot او Wheel plot لتوضيح استعمال الحوامض الامينية فضلاً عن استعمالها للشفرات الوراثية كما موضح في الشكل 96.5 :



شكل 96.5 : استعمال Rose plot لتوضيح إنتاج الحوامض الامينية في الأحياء المختلفة

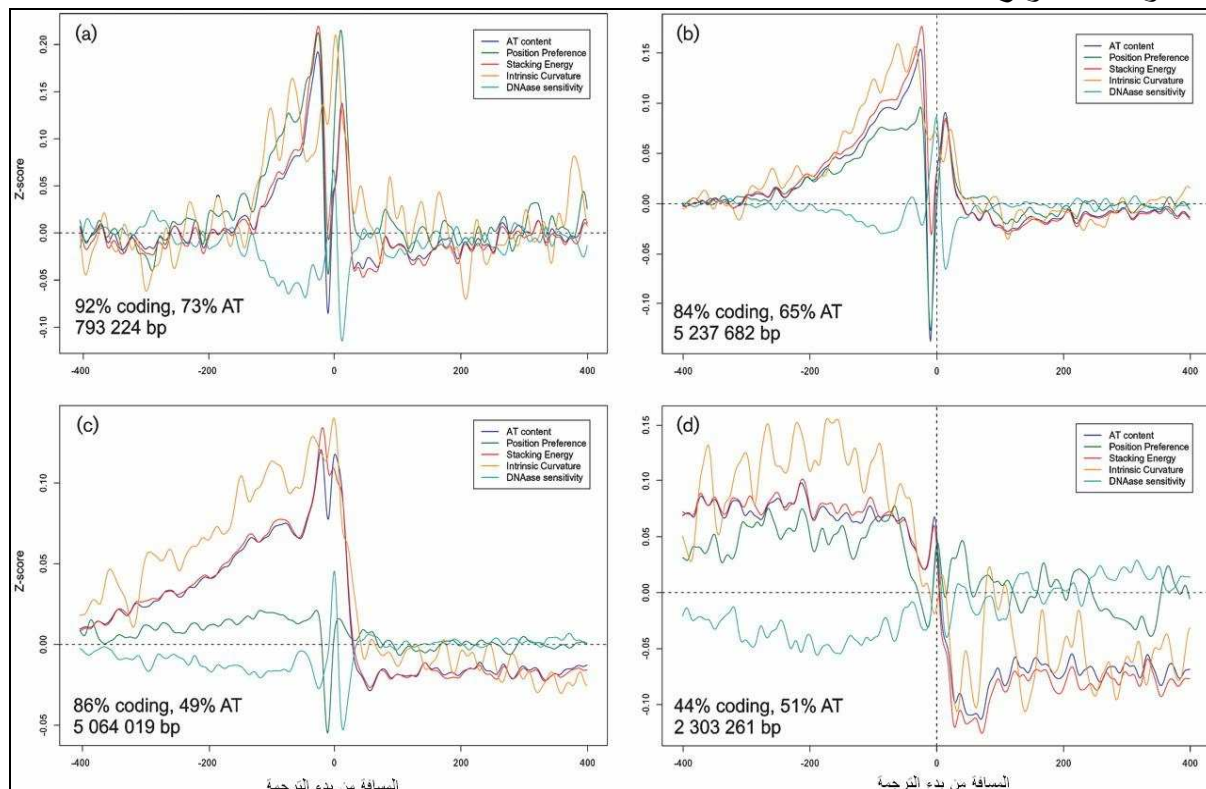
ان الرسوم أعلاه يتم الحصول عليها بمقارنة توزيع الشفرات في مجموعة من الجينات ذات التعبير العالي نسبة الى خلفية توزيع الشفرات في كل الجينات المشفرة في الجينوم وإجراء عمليات الاصطفاف بقيم واطئة جدا" من E. value تصل الى 10^{-6} .

تحديد نسق الممهدات

يكون من الضروري بعض الأحيان تحديد نسق منطقة الممهد وتوفير المعلومات حول المناطق التي تسبق والأخرى التي تلي منطقة بدء الانتساخ وفي هذه الحالة يتم إجراء عمليات الاصطفاف بدأ" من نقطة بدأ ويتم قياس نسق التركيب ثم جمع هذه النسق واستخراج المعدل باستعمال نافذة برنامج بطول معين ويقاس أيضا" الانحراف المعياري الذي يستعمل كأساس في المعادلات الحسابية المستعملة لإيجاد نسق الممهدات وتكون هذه معتمدة أيضا" على المراجع المتوفرة .

وفي العادة فإن منطقة بدأ الانتساخ تكون غنية بـ AT ولذا تكون أكثر قابلية للانصهار من باقي التواليات . والجينومات المختلفة تحوي نسق مختلفة ويعكس الظروف البيئية المختلفة وعوامل أخرى في ترتيب الجينومات . والشكل التالي(97.5) يوضح نسق الممهدات في أحياء مختلفة والمؤثرات المتعلقة بها بالاعتماد على حساب قيم Z . ومثل هذه الرسوم وتفسير نتائجها

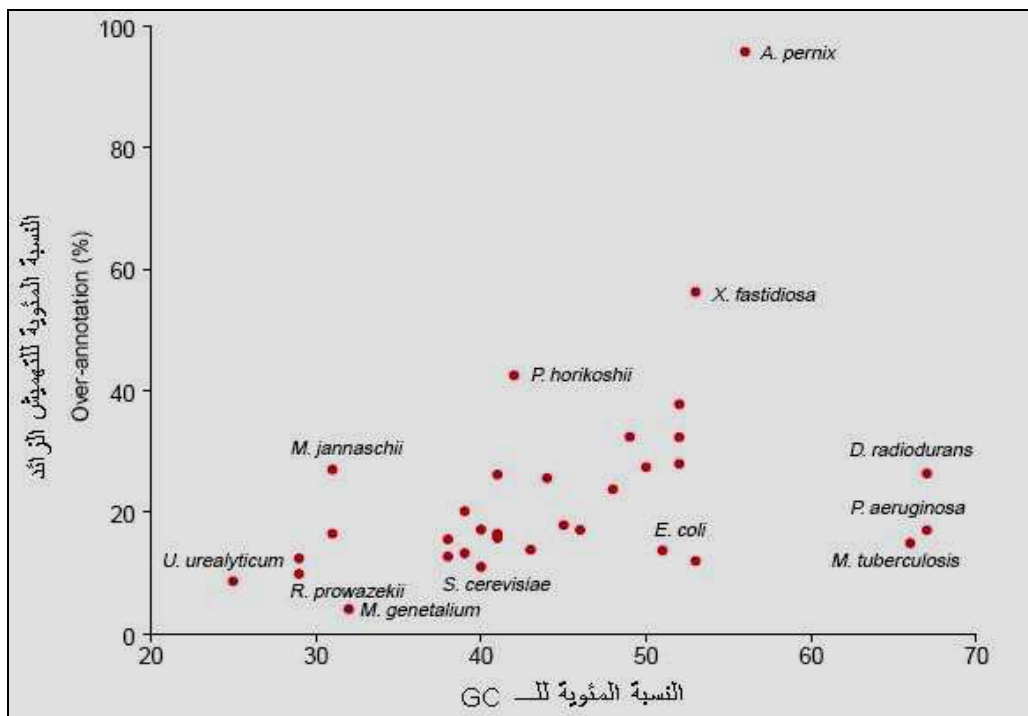
وإقرانها بحقائق أخرى حول الكائن يمكن ان تؤدي الى الحصول على معلومات مفيدة او زيادة المعلومات المتوفرة .



شكل 97.5 : تحديد نسق الممهدات في الجينات والمحتويات الأخرى

عدد الجينات

تحديد عدد الجينات يعد من المسائل التي كثر حولها الخلاف فمنذ إعلان اكتمال تحديد الجينوم البشري عام 2003 لا يوجد هناك رقم محدد من عدد الجينات في هذا الجينوم . ومن الواضح ان اختلاف عدد الجينات يختلف باختلاف حجم الجينوم فمثلاً " *Mycoplasma genitalium* ذات الجينوم الصغير يصل عدد جيناتها الى 480 جين وفي البكتريا حرة المعيشة *Bradyrhizobium japonicum* يصل الى 8317 وفي *Sorangium cellulosum* الذي يكون جينومها اكبر من 12 ميكاقاعدة يصل عدد جيناتها أكثر من 10,000 جين أي ان عدد الجينات يمكن ان يختلف بـ 20 مرة . والخطأ الشائع في تحديد عدد الجينات يعود الى إعلان اطر القراءة المفتوحة بشكل خاطئ على انها جينات أي حدوث حالة التهميش الفائق Over-annotation وهذا يعني ان العديد من الجينات المعلنة هي ليست حقيقية وانما هي اطر قراءة مفتوحة تحدث بالصدفة وهذا يعني ان عدد الجينات المعلنة هي أكثر من الجينات الحقيقية وخاصة الجينات القصيرة ، فمثلاً" المعلن ان جينوم *E. coli* يحوي على 4300 جين ولكن الحقيقية هي 3800 جين ومثل هذا التضارب موجود في جينومات الأحياء الأخرى المنشورة ، والشكل التالي (98.5) يوضح عدد الجينات المعلنة لبعض الأحياء وهي تمثل القيم الزائدة عن القيم الحقيقية .



شكل 98.5 : زيادة الإعلان وتهميش عن اطر القراءة المفتوحة بازدياد محتوى الجينومات من GC%

والسبب في كون الجينات المعلنة خطأ هي ان الجينات قصيرة وشفرات الوقف TAG و TAA و TGA تزداد في الجينومات الغنية بـ AT لذلك يكون ترددها عالي مقارنة بالأحياء ذات المحتوى العالي من GC ، ولذلك يتوقع ان تكون اطر القراءة المفتوحة أطول في الأحياء الغنية بـ GC وأكثر الطرق المقبولة لتحديد الجينات هو دراسة التشابه للبروتينات الناتجة منها في أحياء أخرى ، اما الجينات التي ليس لبروتيناتها تشابه مع بروتينات معروفة فيمكن ان تحدد باستعمال مقاييس إحصائية مع استعمال الشفرات المذكورة سابقا ، فالاختلاف في تردد الشفرات في الجين مقارنة مع يمكن توقعه بالصدفة من مكونات قواعد معينة يمكن ان يوضح بعض الحقائق ، لذلك فأن استعمال الشفرات يكون مؤشرا غير ملائم لأطر القراءة المفتوحة خاصة القصيرة ، ومثل هذه الحالة يمكن ان تحصل عند ظروف أخرى مثل زيادة قوة التشفير Coding potential وإحصائيات المكررات السداسية Hexamer statistics وغيرها .

ومن جهة ثانية فالقائمون على عمليات إعلان وتهميش الجينات يسمون الجينات غير الموثقة بالجينات الافتراضية ، والمستخدم لقواعد البيانات يتصور ان كل الجينات المذكورة في قواعد البيانات هي حقيقية وهذا يؤدي الى استنتاجات خاطئة . ولمعالجة مثل هذه الحالات أشارت الدراسات الى ان اطر القراءة المفتوحة التي تعطي 200 حامض أميني هي حقيقة وليست

صدفة ولكن في واقع الحال ان 15 - 20 % من البروتينات هي اقصر من 200 حامض أميني . ولذلك كان من الواجب تحسين إعلان الجينات وعددها وعندها تكون هناك :

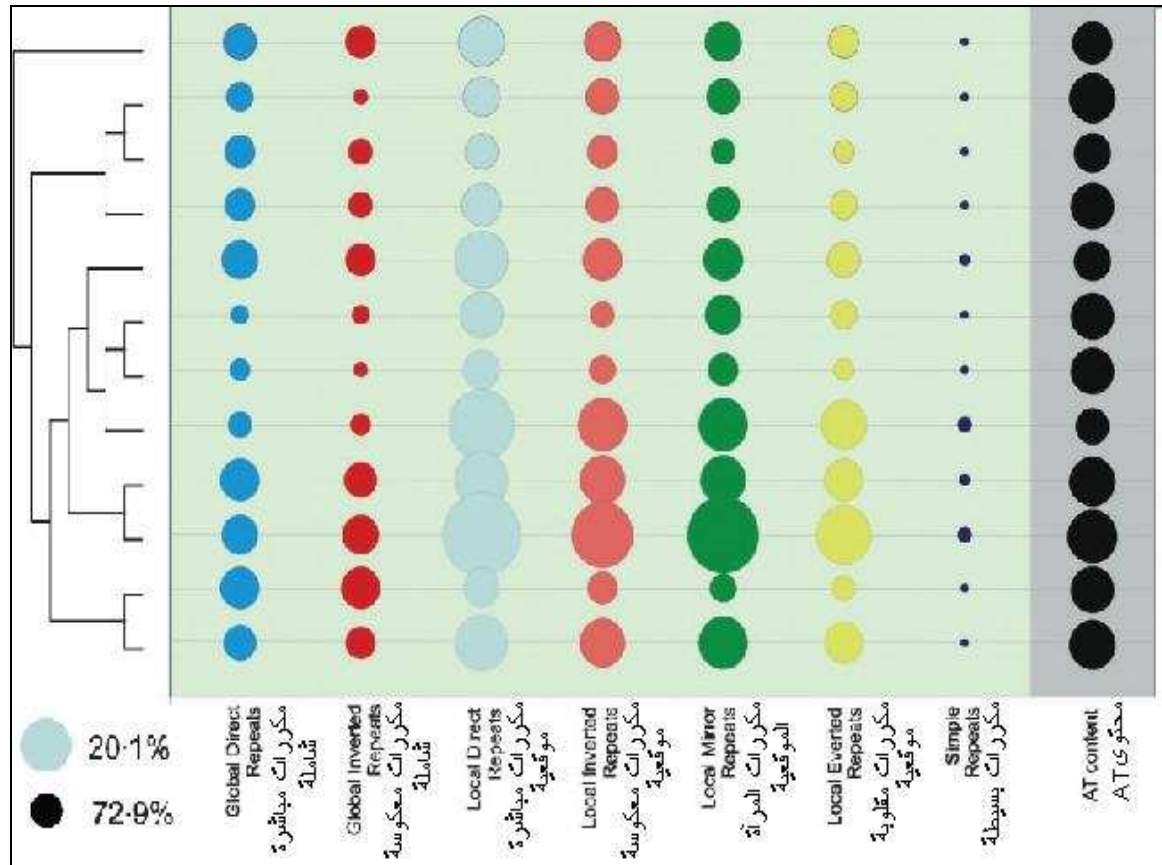
- حاجة الى أهمية إحصائية عالية للتنبؤ او حدس الجين
- اطر القراءة المفتوحة يجب ان تعلن وتهمش على انها مناطق تشفير CDS اذا كانت الأهمية الإحصائية كبيرة او انها تطابقت نواتجها مع بروتينات موثقة .
- فيما اذا أعلنت بعض الجينات على انها اطر قراءة مفتوحة فيجب ان توضح على انها افتراضية

والملاحظ ان مدى دقة التعرف على عدد الجينات يعتمد على قواعد البيانات .

العوامل المؤثرة في الجينومات البكتريا

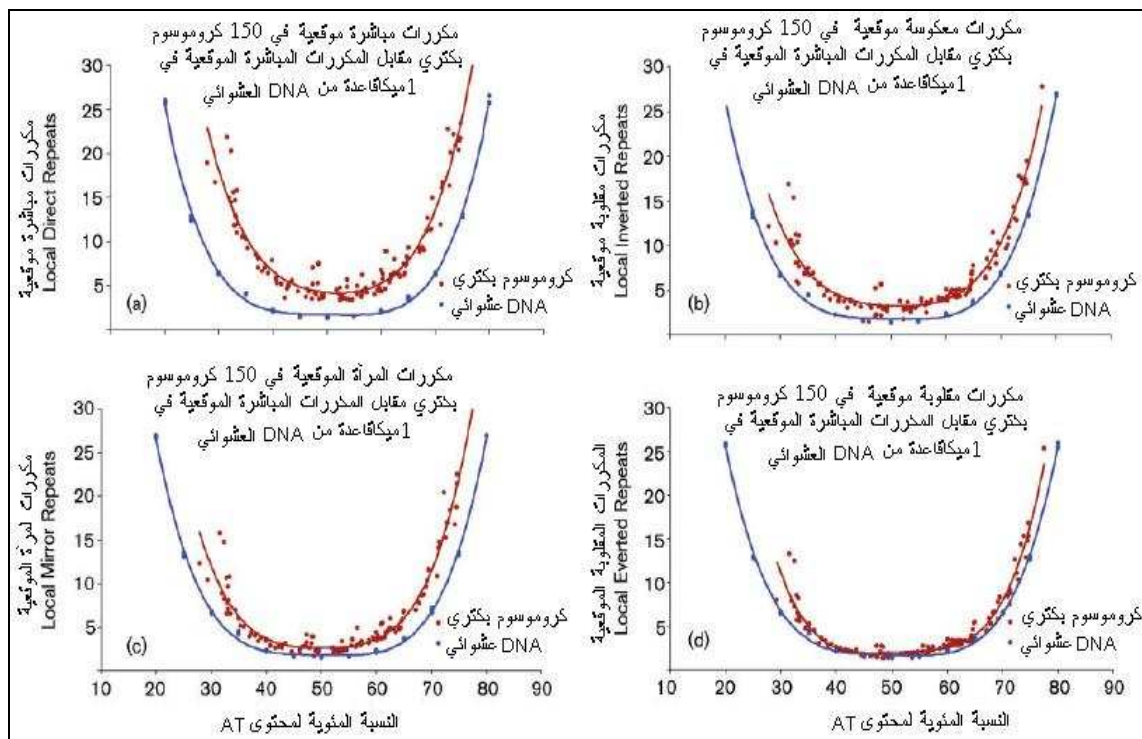
مما سبق مناقشته ان الجينومات البكتيرية تعد منطلقا لتحديد العديد من الصفات لجينومات الأحياء نظرا لبساطة تركيبها وصغرها (في العموم) مقارنة بالأحياء حقيقة النواة فضلا عن ارتفاع الكثافة الجينية لها . ولكن هناك بعض الجوانب التي لا بد من ذكرها في هذا المجال . فجينومات البكتريا تتأثر بشكل كبير بالبيئة التي تعيش فيها فمثلا " سلاسلات من *Prochlorococcus marinus* تعيش تحت شدة إضاءة عالية تكون جينوماتها اصغر من تلك المتطبعة على إضاءة قليلة ، كما ان الجينومات المختلفة تحوي على مناطق تقوس DNA curvature (الذي سيأتي ذكره لاحقا) في مناطق الممهدات مختلفة مما يشير الى تأثير الظروف البيئية في انتظام الجينوم وترتيبه .

وتتصف الجينومات البكتيرية بقلّة محتوياتها من المكررات وهي في العادة تعد مزايا انتخابية فهي يمكن ان تشجع التنوع وذلك بإيجاد مجاميع تختلف قليلا في مناطق محددة . وتوجد أربعة أنواع من المكررات منها المباشرة Local direct repeats التي تكون في مناطق صغيرة مثل 100 قاعدة وهذه يمكن ان تكون تراكيب لا تتطابق نتيجة الانزلاق مؤدية الى الحذف او التضاعف أي تؤدي الى حدوث طفرات إزاحة الأطر وتؤدي الى حالة الأشرطة المنزلة Slipped - strand DNA (S- DNA) ، اما مكررات المرآة Mirror repeats فهي تمثل تواليات DNA مقلوبة على الشريط نفسه ويمكن ان تؤدي الى تكوين شريط ثلاثي Triplex DNA او ما يسمى بـ H- DNA ، اما Everted repeats فهي التي تبدأ من 5' الى 3' من تواليات DNA مكررة على الشريط المقابل بالاتجاه من 3' الى 5' وتكون حلزونات متوازنة ، في حين ان المكررات المقلوبة Inverted repeats تستطيع تكوين تراكيب صليبية الشكل وقيم معدلات هذه المكررات حسب في 189 جينوم بكتيري تمثل 12 شعبة موضحة في الشكل 99.5 :



شكل 99.5 : أنواع المكررات في 12 شعبة بكتيرية

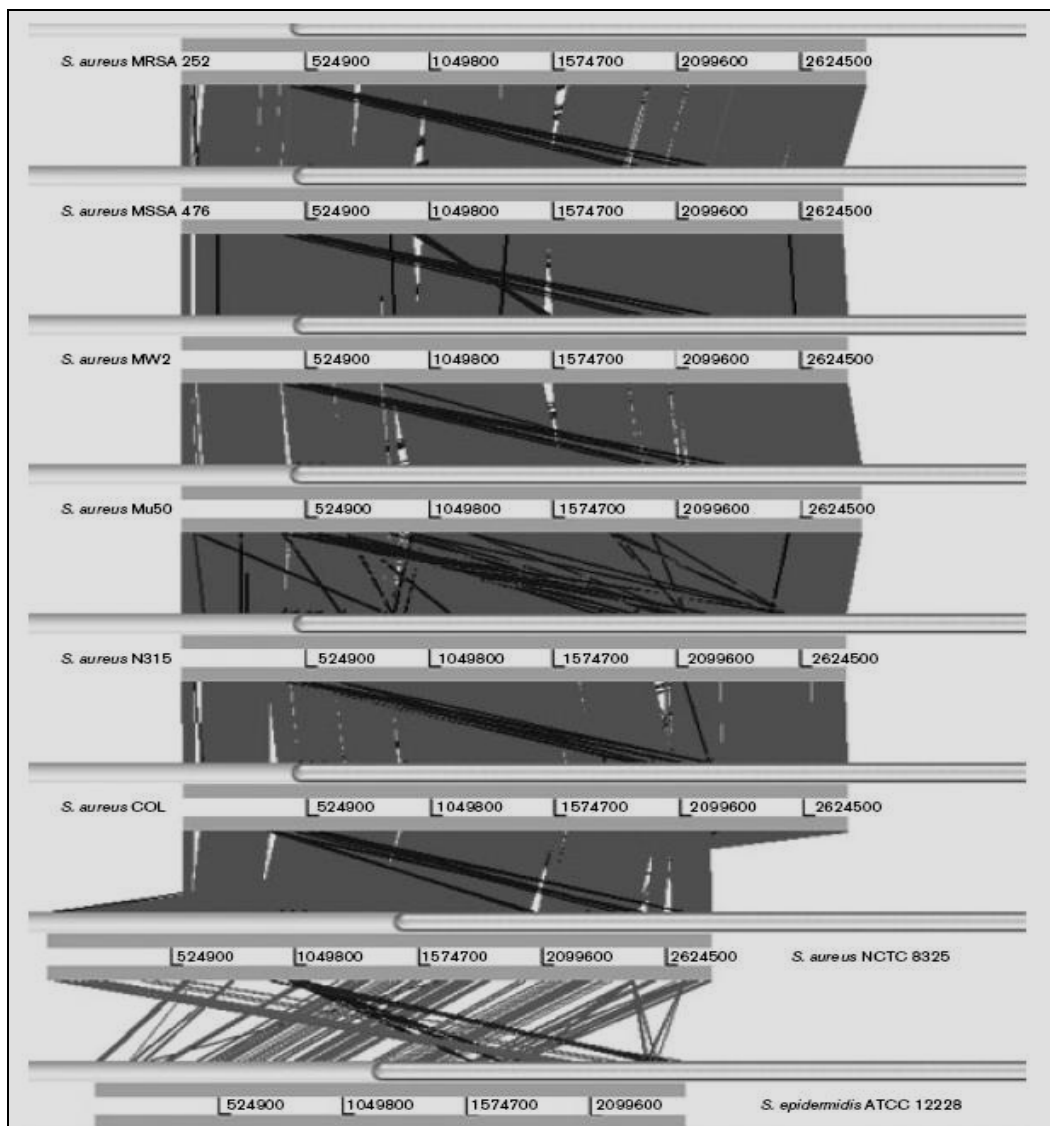
والملاحظ ان *Fusobacteria* لها أعلى ما يمكن من المكررات ، وتشير قيم المكررات مقارنة بمحتوى الجينومات من AT (الجهة اليمنى) الى ان التغيير في محتوى AT يؤثر في مستويات المكررات كما موضح في الشكل 100.5 :



شكل 100.5 : تأثير محتوى AT في أنواع المكررات الموجودة في الجينومات البكتيرية

والنقاط تمثل القيمة في كروموسوم واحد ، والملاحظ من الشكل ان اقل مستويات للمكررات تحصل عندما تقرب A + T من 50% ، في حين ان الجينومات ذات المستوى الاوطأ والأعلى من نقطة المنتصف تظهر فيها مكررات أكثر وهذا يشير الى العلاقة القوية بين محتوى الجينوم من AT والنسبة المئوية للمكررات المختلفة ، وتوجد قواعد بيانات خاصة بهذه الظاهرة تضم الأحياء التي تم تحديد تواليات جينوماتها .

وهناك عدة برامج لمقارنة جينومات البكتيريا منها السريعة وتوضح نتائجها بالصور كما موضح في الشكل 101.5 :



شكل 101.5 : مقارنة جينومات سلالات مختلفة من *Staphylococcus aureus* باستعمال برامج ACT

ومن الواضح ان الرسم باستعمال طريقة ACT (المارة الذكر) يظهر مقارنة الجينومات لسلالات مختلفة مثل *Staphylococcus aureus* ويظهر الشكل الاختلافات البسيطة بين جينومات السلالات والمناطق القليلة التي حصل فيها إقحام او حذف والمتمثلة بالمثلثات البيض . ويمكن ربط طريقة ATC المتوفرة بإصدارات مختلفة مع برامج أخرى للحصول على نتائج أفضل ، وبالعوم فأن مثل هذه البرامج المتوفرة على شبكة الانترنت مرتبطة بقواعد بيانات عامة مثل بنك الجينات وقواعد المركز الأوروبي وتقبل صيغ التواليات التابعة لهذه القواعد، وفضلاً عن إمكانية ربطها ببرامج المقارنة مثل BLAST و FASTA ، وبطبيعة الحال فأن مخرجات هذه البرامج تعد من عمليات التهميش للتواليات المستعملة .

خرائط وأطالس الجينومات

ان الزيادة في البيانات الجينومية بشكل هائل والخوف من ان تزيد عن سعة استيعابها دفع الكثير الى ايجاد وسائل لتلخيص وترتيب البيانات وإدارتها للحصول على المعلومات المفيدة منها بشكل ملائم للعمل مع أنظمة تشغيل الحاسبات المختلفة ومتصفحات الانترنت . ولهذا الغرض وجدت الخرائط والأطالس .

من أهم أهداف وضع خرائط او مشاريع تحديد التواليات هو التعرف على الجينات التي تمثل العناصر الفاعلة في الجينوم وبذا فهي تمثل الخطوة المهمة في فهم تركيب الجينوم والتي تعنى بتحديد المواقع النسبية للجينات والطفرات والصفات الأخرى على الكروموسوم او الجينوم . وهناك مشاكل تتعلق بإيجاد الجينات في الخلايا بدائية وحقيقية النواة التي تكون مختلفة في الحجم والكثافة الجينية . فالكثافة الجينية في الخلايا بدائية النواة تكون عالية وتصل الى 85 % وتحتوي القليل من المكررات كما ذكر في مواقع أخرى ، والقليل من جيناتها تحوي على الانترونات لذلك كانت الإستراتيجية في إيجاد جيناتها هو البحث عن اطر القراءة المفتوحة الطويلة ، كما انه يجب استعمال طرق خاصة للتعرف على الجينات غير المشفرة للبروتينات مثل جينات RNAs ، اما جينومات حقيقية النواة فهي تكون كبيرة وتتراوح في البسيط منها بحدود 13 ميكاقاعدة الموجودة في الفطريات الى 100,000 ميكاقاعدة الموجودة في النباتات المزهرة . والملاحظ ان تعقيد الجين والجينوم يزداد كلما زاد الكائن في التعقيد ، فمثلا في خميرة الخبز يكون 70% من الجينوم هو جينات وتنخفض النسبة الى 25 % في الدروسوفلا وتصل الى 3-1 % في جينومات الفطريات والنباتات الراقية ، اما باقي الجينوم فيكون من المكررات والذي يؤدي الى تعقيد تحديد التواليات ويحتاج الى وقت وجهد كبيرين ومعدات ووسائل عالية الإتقان .

والانترونات تعقد من تحديد التواليات وغيرها من الصفات ، ففي الإنسان هناك 95 % من الجينات تحوي على الانترونات ويصل عددها كمعدل بين 1 - 12 لكل جين ويشكل 10 % والأخرى تحوي على 20 انترون وأخرى يصل عدد الانترونات الى 60 انترون لكل جين ، واكبر الجينات في الإنسان والذي يصل حجمه الى 2 ميكاقاعدة هو جين Duchenne muscular dystrophy locus الذي يحوي على 79 انترون (المذكور في مواقع أخرى خاصة) وهذا يعني ان تحديد توالي الجين يقابل الجهد المصروف لتحديد توالي جينوم كامل لبعض البكتيريا .ومقابل هذا فأن حجم الاكسونات في الجينوم البشري يكون بحدود 150 قاعدة مع انترونات تختلف بشكل كبير وتتراوح من عدة مئات من القواعد الى عدة كيلوات . والانترونات يمكن ان تعترض اطر القراءة المفتوحة في أي موقع حتى داخل الشفرة ، ولذلك فأن البحث عن اطر القراءة المفتوحة في جينومات الخلايا حقيقية النواة غير ملائم وغير كافي وانما احتاج الى طرق متخصصة .

وفي مجال الخرائط الجينومية هناك بعض المؤشرات المستخدمة فيها مثل الواسمات الوراثية وهي تمثل جزءا "معينا" من الكروموسوم الذي يمكن تتبع نمط توارثه وللعديد من الخلايا حقيقية النواة فأن هذه الواسمات تمثل أنماطا" مظهرية معينة ، اما المصطلح الآخر وهو الوضوح الواطئ في عملية وضع الخرائط فيمثل وصف الترتيب والمسافات بين الواسمات الوراثية على الكروموسوم وضمن هذا المجال توضع أنواع من الخرائط الجينومية منها خرائط الارتباط

الوراثي وفيها تحدد مواقع الواسمات الوراثية على الكروموسوم والكيفية التي تتردد بها وتوارث الصفات سوية . فضلا" عن الخرائط الجينومية الأخرى مثل Physical maps و Cytological maps التي تصف الجينومات على درجات مختلفة والتي مر تفصيل بعض جوانبها .

ومن هنا يتضح ان البيانات الجزيئية الحيوية تكون متوفرة على مستويات مختلفة تمتد من مستوى الجزيئات الى مستوى الدراسات omics- ، والبيانات في المستوى الواحد يمكن ان تقسم الى عدة طبقات مثل تركيب الجينوم ، عمليات الانتساخ ، المكنون البروتيني ومكنون الايض والمسارات الكيموحيوية لذلك كان من الواجب إيجاد برامج للبحث في قواعد البيانات لإيضاح المعلومات وفي مثل هذه الحالة تفيد البرامج القابلة للتكبير لمناطق الجينوم وكذلك القدرة على جدولة البيانات خاصة عند البحث والتجوال في World Wide Web ، لذلك وجدت برامج ترتبط بـ Google maps للتجوال والبحث عن الجينومات ، في هذا المجال خصصت مواقع على شبكة الانترنت تخص جينومات البكتريا والتي تقدم خدمات كثيرة فضلا" على خريطة لجينومات البكتريا التي تم تحديد توالياتها ومنها المواقع المجانية للاستعمال مثل :

BacMap يقدم خرائط الكترونية لجينومات البكتريا والتفاصيل الكثيرة عنها فهو يوفر خدمة البحث عن الكروموسومات ، وتكون الخرائط قابلة للتكبير وباستعمال بطاقة الجين Gene card يوفر الموقع مجالات واسعة من البيانات عن الجينات والبروتينات التي تشفر لها . وضمن بطاقة الجين تعطى المعلومات المهمة بالاعتماد على ما موجود في قواعد البيانات العامة التي تحدث باستمرار وعلى مدى أسبوعي من سجلات بنك الجينات التابع للـ NCBI وكل بكتريا او اركيا مدرجة ضمن الموقع لها بطاقة النوع الذي يوفر المعلومات حول الكائن بصيغ جداول تشمل النواحي التصنيفية مثل صبغة كرام وعدد الكروموسومات ، فضلا" عن وصف بسيط لفلسجة الكائن والصفات العامة والبيئة التي تعيش فيها الكائن والعلاقة بأمراض الإنسان والحيوان وصورة للبكتريا في حالة توفرها ، وتحتاج الى تدخل عملي عند تحديثها . وتحوي واجه الموقع على إمكانيات متعددة مثل إظهار الخريطة الكروموسومية في نافذة جديدة وإمكانية إجراء المقارنات باستعمال BLAST بشكل مباشر ، وكذلك إمكانية الحصول على توالي الجينوم للبكتريا الكامل وبلازمياتها ، وإمكانية الحصول على أنواع البروتينات ونسبها فضلا" عن إعطاء أرقام التسجيل لها .

ويوفر الموقع إمكانية الحصول على نصوص البحوث المتعلقة بالجين او توالي محدد . والشكل 102.5 يوضح واجهة الموقع

BacMap
genome atlas

BacMap: An Interactive Atlas for Exploring Bacterial Genomes

[Home](#) | [About](#) | [Help](#)

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | All

[Acaryochloris marina MBIC11017](#)

Acaryochloris marina MBIC11017, complete genome.
NC_009925
6,503,724 bp

[Map](#) [Text Search](#) [BLAST](#) [Stats](#) [Download](#)

Acaryochloris marina MBIC11017 plasmid pREB1, complete sequence.
NC_009926
374,161 bp

[Map](#) [Text Search](#) [BLAST](#) [Stats](#) [Download](#)

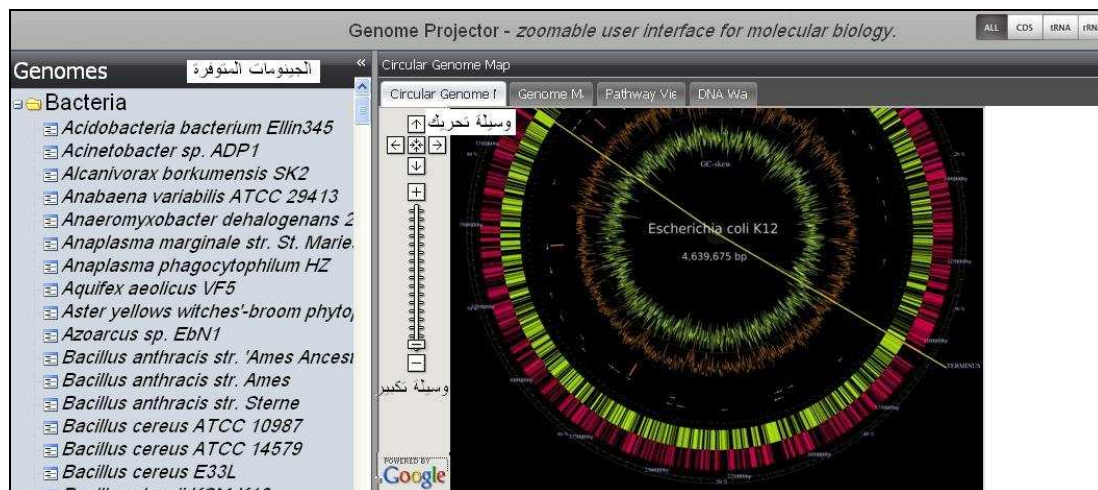
Acaryochloris marina MBIC11017 plasmid pREB2, complete sequence.
NC_009927
356,087 bp

[Map](#) [Text Search](#) [BLAST](#) [Stats](#) [Download](#)

شكل 102.5 : واجهة الموقع BacMap

Genome projector يمثل أحد المواقع المهمة التي يمكن الوصول إليها من محرك البحث Google . وواجهة هذا المتصفح تحوي العديد من الإمكانيات وروابطها والمحتويات فيه قابلة للتكبير لغرض الدراسة . ويضم الموقع العديد من البكتيريا التي تم تحديد تواليات جينوماتها . وتتوفر فيه خرائط لكل الجينومات الحاوية على المناطق المشفرة CDS , tRNA , rRNA .

والخرائط المتوفرة تكون للجينومات الحلقية الموضح فيها محتوى الجينوم من GC وكذلك انحيازها GC skew فضلا عن توفير خرائط للمسارات الايضية ومستويات التعبير ، إضافة الى ذلك يوفر الموقع وسيلة BLAST الملحقة به والتي يمكن ان تستعمل بطريقة بسيطة وإمكانيات أخرى . وهناك ملاحق لتوضيح الوسائل المستعملة وتخصص شفرات الألوان . والشكل التالي (103.5) يوضح واجهة الموقع .



شكل 103.5 : واجهة الموقع Genome projector

ومن الجدير بالذكر ان خرائط المسارات الايضية المستعملة هي خرائط شركة Roche الجدارية (Roche Biochemical Pathway wall chart) وهذه تكون أفضل من استعمال قواعد البيانات الخاصة بالمسارات الايضية لانها تظهر :

- التفاعلات
- الإنزيمات
- التراكييب الكيميائية للمركبات
- حالات التنشيط والتثبيط
- قابلية التفاعلات على الرجوع.

ولكن من الواجب ملاحظة ان ليس كل الإنزيمات والجزئيات الموجودة في الخريطة تكون موجودة في كل الأحياء ، ، ولكن باستعمال خاصية Semi-transparent overlay layer الموجودة في الجهة اليمنى العليا للواجهة يمكن ان تظهر مواقع الإنزيمات الموجودة في الجينوم المستعمل وحجب البقية . والإنزيمات الظاهرة تكون أرقامها EC المشتركة بين قواعد البيانات ذات العلاقة مثل قاعدة كيو تو ، ExPASy ، ENZYME . وهناك إمكانيات أخرى للدراسة في هذا الموقع الذي يعد بوابة مفيدة للدراسات (- omics) على مقاييس متعددة وطبقات متعددة .

أطالس الجينومات

أطالس الجينومات تضم أنواع من الأطالس (التي سيتم ذكرها لاحقاً) وهي رسوم دائرية او خطية للكروموسومات او البلازميدات تظهر فيها صفات جزيئية للـ DNA واغلبها معتمداً على الألوان وتدرجاتها.

والتطور السريع وغير المكلف لتحديد تواليات الجينومات فضلاً عن وجود مشاريع الجينومات البيئية Metagenomic projects تضيف المزيد التي يتوقع ان يتضاعف سنوياً أدت الى تجمع كميات كبيرة من البيانات الجينومية ، وأصبح التعامل مع البيانات وتحليلها هي التحديات

الرئيسية وصرف الأنظار عن عمليات الكشف الجديد وانحسرت الأخيرة ولو قليلا" لتحل محلها الاعتبارات التقنية للتعامل وخزن وتحليل التواليات ، لذلك كانت أول خطوات اكتشاف جينوم جديد هو المقارنة مع التواليات الموجودة لغرض تحديد المواصفات الجديدة الثابت منها والمتغير ، فضلا" عن تطوير العديد من البرامج للاكتشاف والتعرف عن ما يمكن ان يشفر له الجينوم الجديد من بروتينات . وتعد عملية إظهار الجينات من الأشياء المهمة لتزويد العامل في مجال علوم الحياة بنظرة شاملة حول الجينات وهل هي ضمن عنقود او اوبرون والكيفية التي يتم تنظيمه والحال التي توضح مكونات القواعد المحيطة به في باقي الجينوم . ولذلك تعد أطالس الجينومات وسيلة لجمع عدد من المواصفات والمعلومات في شكل واحد يلخص قصة طويلة ومفصلة عن الكروموسوم او البلازميد .

فوائد الأطالس

اغلب الأطالس خاصة بالأحياء المجهرية والبكتريا على وجه الخصوص وبلازميداتھا واغلبھا من بيئة اليابسة والبحار ، ومن المعروف ان هذه الأحياء تكون ذات كثافة تشفيرية عالية تصل الى أكثر من 80% وان كانت تقل عن هذه القيم في الأحياء التي تقطن بيئات معينة وتعتمد على نوعية الكائن . وباستعمال الأطالس يمكن ربط عدة صفات . ويمكن للأطالس إظهار الصفات الآتية

- طول الجينوم
- مكونات الجينوم من القواعد ومحتوياتھا من AT
- الكثافة الجينية وهل ان الجينات ضمن عناقيد او اوبرونات
- كيفية التعبير الجيني والتي تعتمد على المواصفات التركيبية للـ DNA .
- مقارنة مكونات الجينوم من القواعد النتروجينية .
- تحديد مناطق المكررات من النيوكلويتيدات لتحديد مناطق التنظيم والتي تستعمل في تفسير مكونات الجينوم .
- إظهار عدد جينات tRNA و rRNA .
- استخدام الشفرات
- انحراف محتوى القواعد Base skew فضلا" عن انحراف قطع صغيرة من النيوكلويتيدات Oligo - skew والتي ساعد في إيجاد نقطة أصل التضاعف في طول 1-8 من القواعد النتروجينية
- إظهار تقوسات جزيئة DNA (DNA curvature)
- إظهار طاقة الرص للقواعد النتروجينية Stacking energy

مقومات رسم الأطالس

تعد الأطالس وسائلًا توضيحية ذات فوائد جمة لتزويد المستخدم بنظرة شاملة حول محتوى الجينوم ، وأفضل الرسوم هي القابلة لتكبير مناطق ذات الأهمية . وتعتمد وسائل رسم الأطالس على الألوان وشفراتها لإظهار المعلومات الرقمية في أطالس الجينومات . وتشفير الألوان يكون:

- باستعمال مقياس خطي له قيمة دنيا وعليا ثابتة والقيم الباقية تتدرج بينها

• استعمال المقياس الديناميكي Dynamic scale المعتمدة على الانحراف المعياري ، وعند استعمال هذا النظام فإن شدة اللون تقل عند الوصول الى قيم المعدل ، لذا يتم التأكيد على المناطق التي فيها تباين مهم.

ويستعمل اللون الرمادي الباهت للمعدل وعندما تكون قيم المعدل تشذ عن الانحراف المعياري بواحد (1) يزداد اللون غمقا ، اما اذا كان بحدود 2-3 مرات فيكون غمق اللون أعلى بخمس مرات من الحالة (1) ، وعندما تكون القيم أعلى من 3 مقارنة بالمعدل يكون اللون اسود (او أغمق حالة من اللون المستعمل) ، أي ان القيم المحسوبة تمثل بشفرات اللون.

وفي العموم فإن المناطق الوسطية تأخذ اللون الرمادي الباهت لتتعلق منها تدريجات الألوان على الجانبين . وتدخل في هذه الوسائل قيم المتوسطات الموجودة في الكروموسوم ، فمثلا" عند عدم وجود فرق واضح في محتوى المادة الوراثية من AT فإنها تقارن مع متوسط ما موجود في الكروموسوم وعند هذه المنطقة فإنها تشير الى دخول DNA غريب مثل النقاط العاثيات او DNA غريب دخل بطرق أخرى .

وتوضح في الأطالس درجة الوضوح Resolution فمثلا" درجة 928 bp يعني ان أي خط واضح في الدائرة يمثل منطقة من DNA فيها 928 قاعدة وعلى ضوءها تعدل باقي القيم . ويوضح في الأطالس موقع بدء التضاعف Ori C وموقع انتهاء التضاعف وترفق مع الأطالس جداول للإيضاح وتبيان للعلاقات المستعملة .

المؤشرات المستعملة في رسم الأطالس

هناك العديد من المؤشرات التي تستعمل في رسم الأطالس اعتمادا" على نوع الأطالس بعد إخضاعها لعمليات إحصائية ، وقد وجد ان القيم الخاصة بتراكيب DNA تكون أكثر تطرفا" من تلك المتوقعة من مكونات النيوكليوتيدات

تقوس DNA (DNA curvature)

التراكيب الدقيقة لسلسلة DNA تشارك بقوة في محتوى ونوعية المعلومات من حيث ارتباطها بالبروتينات والتضاعف والتأشب والتحول في ظفر الأشرطة وبدء الانتساخ وإنهاءه وتحديد مواقع الجسيمات النووية وترتيب الكروماتين ، ومعظم هذه الصفات يسيطر عليها بالتقوسات التي تساهم بها التواليات DNA ، ولذلك فإن التعرف على الصفات الميكانيكية تكون أساسية لفهم بايولوجية DNA لان العديد من الصفات الفسلجية - البايولوجية للـ DNA تتأثر بالتقوسات الموضعية ومرونة الجزيئة . ولذلك فإن الجينومات المختلفة لها تقوسات في مهادتها مختلفة مما يشير الى تأثير الظروف البيئية في انتظام الجينوم وترتيبه والمناطق المقوسة من التواليات قد تكون أصلية ولا تتأثر بالاجهادات وهي التي يطلق عليها التقوس الذاتي Intrinsic curvature والتي تكون بعيدة عن التقوسات التي تحدثها حالات الإجهاد . وصفة تقوس DNA لها علاقة قوية بحركة قطع DNA في الهلام أثناء الترحيل الكهربائي ، فالقطع ذات التقوس العالي تتحرك ببطء على هلام Polyacrylamide gel مقارنة بقطع الواسمات المستعملة ذات الوزن الجزيئي نفسه ، ولذلك يعد الترحيل الكهربائي ابسط الطرق لدراسة تقوس جزيئات DNA ومرونتها ، والشذوذ في الحركة يجب ان يدعم ويحدد بالطرق التجريبية او استعمال Monte Carlo methods .

توزيع مناطق التقوس

تكثر مناطق التقوس في الممهدات التي تكون أكثر تقوساً و أقل ثبوتاً مقارنة بالمناطق المشفرة او المناطق البينية وقد تكون في الممهد أكثر من نقطة تقوس ، وهذه تشير الى الحاجة الى فتح التواءات DNA أثناء عملية الانتساخ ولذلك كان للتقوس دوراً أساسياً في تنظيم الجين ، وتوزيع مناطق التقوس في الممهدات يكون ثابتاً من الناحية التطورية وتحدد بشكل أساسي بحرارة البيئة . وتتوزع مناطق التقوس على جانبي المناطق المشفرة في الخلايا بدائية النواة ، والانحناءات الى اليمين في هذا النوع من الخلايا تكون في البكتريا المحبة للحرارة او المحبة لحرارة متوسطة متوفرة بشكل أكثر مقارنة بمناطق البدء (اليسار) مما يشير الى علاقة آليات التنظيم بين عمليات بدء وانتهاء الانتساخ وهذه تحدد بالظروف البيئية والعوامل الجينومية . والتقوس له علاقة بالمناطق الغنية بالنيوكليوتيدات الثنائية لذلك فإن التقوس في توالي معين يكون حاوياً على قطع من التواليات الأدينين (A) tracts ولكن التقوسات لا تحصل عند وجود الأدينين فقط ولذلك انصب الاهتمام على دراسة النيوكليوتيدات الثنائية والثلاثية باستعمال الأشعة السينية والرنين النووي المغناطيسي فضلاً عن استعمال الترحيل الكهربائي وحركات التحلق. Cyclization kinetics وغيرها ، وطرق استعمال النيوكليوتيدات الثنائية تكون أسهل من استعمال النيوكليوتيدات الثلاثية عند حساب قيم مختلفة تتعلق بالظاهرة .

وقد وجد ان التقوس يمكن ان يحصل نتيجة للانفتال الحاصل عند تلاقي هيئات DNA - B' وهو التركيب غير الطبيعي لقطع الأدينين مع B- DNA المعروف بأنه يملك تركيب متجانس و صلد للأشرطة الحلزونية المزدوجة ، ووجد ان له صفات تركيبية تلعب دوراً أساسياً في العديد من العمليات الحيوية . وان التواليات (A4 T4 CG)n تظهر ترحيل كهربائي غير طبيعي في حين ان (T4 A4 CG)n تكون طبيعية وهذه الصفات لا تتغير عند استبدال CG بـ GC . فضلاً عن ذلك وجد ان A tracts ليس لها هيئات موحدة وانما تعتمد على التواليات المجاورة . وقد قيس تقوس (A4 T4 CG) وكان 26 ° في حين (T4 A4 CG) تقوس بدرجة 8 ° . وهي متطابقة مع بيانات التحلق ، فالتوالي الأول وكما يلاحظ يتقوس بشدة في حين ان (T4 A4 CG) تبقى مستقيمة تقريباً وكل هذا يؤثر في تناظر الالتفاف الفائق .

العوامل المؤثرة في توزيع المناطق المقوسة

هناك بعض العوامل المؤثرة في توزيع المناطق المقوسة في المناطق بين الجينات ومنها

- درجة حرارة النمو
- محتوى الجينوم من AT
- حجم الجينوم
- وجود الماء الذي يؤثر بشكل كبير في تقوسات DNA

لذلك فالملاحظ ان تقوس DNA يختلف بارتفاع درجة الحرارة ، ويمكن ان تتغير فعاليات هذه المناطق بشكل يعتمد على انخفاض الحرارة .

وظائف مناطق التقوس

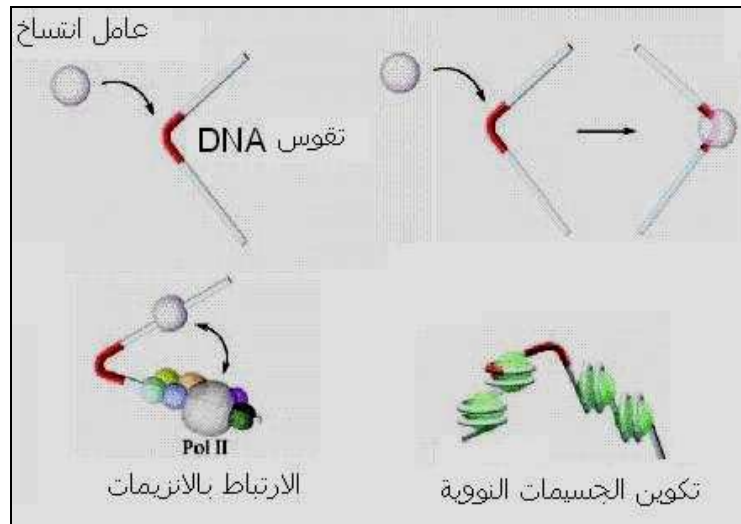
تشارك مناطق التقوس في عدد من الوظائف والتي تعتمد على مواقعها ، فضلا" عن ان وظائفها قد تكون مباشرة او غير مباشرة بواسطة التلامس مع البروتينات المنظمة او الإنزيمات ، ففي هذا المجال فأن التقوسات يمكن ان تعمل كمواقع لارتباط بعض البروتينات التي تساعد في إنهاء الانتساخ سواء تلك المناطق المعتمدة على ρ او المستقلة عنه كما في بكتريا *E. coli* و *B. subtilis*.

ومن الآليات الأخرى التي تساهم التقوسات فيها تشجيع الانتساخ بآليات مختلفة مثل تشجيع ارتباط إنزيم RNAP ، او زيادة ألفة معقدات الانتساخ بتكوين عروة كبيرة حول RNAP وبالتالي جلب مكونات الانتساخ الواقعة على مسافات بعيدة وغير متاحة للتجمع مع بعضها ، وهذا يشير الى ان تقوس DNA له ادوار مختلفة في تنظيم الجين ، فضلا" عن عملها كإشارات للتعرف على البروتينات المنظمة .

علاقة مناطق التقوس بالبروتينات

تعد المرونة والتقوس في DNA احد ضروريات تداخل DNA مع البروتينات ، واحد مظاهرها التأثير في الفعاليات الحيوية والتي تعود الى الصفات التركيبية والميكانيكية للـ DNA وهو تكوين العروات للـ DNA التي تتوسطها البروتينات والتقوس يقلل من كلفة الطاقة خاصة للعروات الثابتة والأخرى اللازمة لتغيير طوبوغرافية العروات . والعديد من البروتينات التي ترتبط الى DNA تؤدي الى ثنيه اذ ان البروتينات يمكن ان تميز أهدافا" مقوسة في التواليات وتؤدي الى تغيير تقوسها كما هو الحال في بعض الأدوية.

وتداخل البروتينات مع DNA له علاقة بتقوس جزيئة DNA سواء عند ارتباط عوامل الانتساخ كما موضح في الشكل 104.5 ، او تكوين الجسيمات النووية كما موضح في الشكل.



شكل 104.5 : تقوس DNA ومساعدته في الانتساخ وتكوين الجسيمات النووية

وقيم التقوس في الجسيمات النووية تصل من صفر في الحالة غير المنحنية الى (1) وهي وحدات افتراضية فعند هذه القيمة (1) فإنه يشير الى تقوس واقعا" ضمن جسيم نووي ، اذ ان هناك جداول خاصة بالقيم لكل نيوكليوتيد . ومن الجدير بالذكر أن صفة التقوس لها علاقة قوية مع طاقة الرص وقابلية التشوه المستحثة بالبروتينات (كما سيأتي ذكرها لاحقا") وكذلك الارتباط الى الأيونات الموجبة ، وتعتمد على الدورية Periodicity والتي تمثل 11 قاعدة في الخلايا حقيقية النواة و 10 قواعد في الخلايا بدائية النواة وتستعمل في تحديد الطمعة النيوكليوتيدية Nucleotide signature .

البرامج الخاصة بتحديد التقوس

هناك برامج خاصة لتحديد مناطق التقوس تعتمد على عدد من القيم مثل حالة الالتواء واتجاه زاويتها والتركيب ثلاثي الأبعاد لـ 21 قاعدة (مثلا" المستعملة في برنامج CURVATURE) الذي يستعمل نافذة بطول 21 قاعدة وهناك موديلات نظرية يمكن استعمالها والتي تعطي نتائج مشابه لما تعطيه البرامج المستعملة .

طاقة الرص Stacking energy

التضيد او الرص تعني ترتيب الجزيئات الكيميائية وخاصة الحلقية نظرا" للتداخلات بين ذراتها كما في قواعد DNA وكذلك تحدث في البروتين عندما تتداخل الحلقات اللاقطبية . ويثبت رص القواعد ببعض القوى مثل London dispersion forces والتداخلات بين الشحنات للحلقات المتجاورة وكذلك التداخلات الكارهة للماء وان كانت هذه المؤشرات لا يمكن ان تستعمل للكشف عن رص القواعد وانما تكون هناك حاجة لتوجهات او مؤشرات أخرى مثل التوجهات الهندسية . وعملية رص القواعد تؤثر في الفعاليات الحيوية ، والرص العمودي لقواعد النتروجين تعد من مصادر تثبيت ترتيب DNA وكذلك التركيب الثلاثي لجزيئات RNA اذ تلعب دورا" مهما" في عمليات الطوي وتكوين المعقدات ، ويحدث الرص بين القواعد المتجاورة في مناطق أشرطة الحلزونات المزدوجة . وطاقة الرص تعد من المؤشرات المؤثرة في صلابة تركيب DNA وهي تمثل مقياس لطاقة التداخل بين أزواج من القواعد المتجاورة في مزدوجات أشرطة DNA . وهي تقيس قابلية الانصهار ومدى سهولة انفصال أشرطة DNA في الحلزون أي انها تقيس القوة التي تتداخل بها مستويات القواعد الحلقية المتجاورة وتقاس بالكيلوسعرة ١ مول . ويمكن تحديدها من مجموعة من قيم النيوكليوتيدات الثنائية بأجراء حسابات الميكانيك الكمي Quantum mechanical calculations على التراكيب البلورية للنيوكليوتيدات الثنائية التي تستعمل في رسم الأطالس .

وتكون القيم سالبة لان رص القواعد يمثل التداخلات المفضلة من ناحية الطاقة التي تفيد في تثبيت الحلزونات المزدوجة ، فالقيم السالبة العالية تشير الى ثبوت عالي ولا تميل للانصهار او حصول حالة إزالة الرص مقارنة بالمناطق ذات القيم الواطئة . وتتراوح القيم من - 3.82 كيلوسعرة ١ مول التي تمثل حالة الانصهار السهل الى حوالي - 14.59 كيلوسعرة ١ مول اللازمة لفك تكسد او صهر الحلزون .

والمؤشرات المؤثرة في عملية الرص تشمل الحرارة وتركيز الأملاح مثل وجود 15-100 ملي مول من أيون الصوديوم ، فضلا عن وجود علاقة طردية بين هذه الطاقة ومحتوى DNA من GC ودرجة الانصهار .

والتذبذب او التمدج الحاصل في الهيئة الحلزونية الموضعية للـ DNA ضمن الظاهرة المعروفة بتنفس DNA (DNA breathing) تكون نتيجة للتذبذب في عمليات الرص تؤدي الى حصول أحداث لا تتكرر بكثرة حيث تنفتح أزواج القواعد مؤدية الى جعل المجاميع المدفونة بشكل طبيعي متاحة لعمليات التغيير والتداخل مع البروتينات ، ومثل هذه العمليات تتضمن تدمير الأواصر الهيدروجينية بين القواعد المتكاملة وتحرك القواعد بعيدا عن الرص الحلزوني . ومن جهة ثانية فإنه عند تلم DNA يتم تثبيت المنطقة بتداخلات الرص بين أزواج القواعد المحيطة بمنطقة التلم وهذه تكون معتمدة على تواليات المنطقة .

زعزعة مزدوجات DNA المستحثة بالإجهاد Stress – induced DNA duplex destabilization (SIDD) من المعروف ان DNA في جينومات الخلايا بدائية النواة يكون ذات الالتفاف سلبي داخل الأنظمة الحيوية ، وان هناك العديد من العمليات الحيوية ومنها بدء الانتساخ يعتمد على حالة الالتفاف الفائقة للـ DNA ، والاجهادات التي تستحث في DNA بواسطة الالتفاف السلبي الفائق ومنها SIDD يحث عدم ثبوت DNA الذي يعد من الآليات المسيطرة على عدد من العمليات الحيوية ، وتشير الإحصائيات التي جرت على أكثر من 42 جينوم بكتيري الى ان مواقع الزعزعة SIDD توجد بقيم أكثر من المتوقعة وتوجد بشكل محدد في المناطق المسؤولة عن التنظيم ونقل في المناطق المشفرة . وتستعمل مواقع الزعزعة للتنبؤ بالمناطق المهمة المسؤولة عن التنظيم وتوجد قواعد بيانات خاصة SIDD BASE تساعد في تحليل مواقع الزعزعة لجينومات البكتريا والاراكيا . وتقاس كمية وأعداد مواقع الزعزعة على DNA بمستويات تنافس إنزيمات Topoisomerases والأحداث التي تحدث في الموضع مثل ارتباط البروتينات الى DNA او انتساخه وكذلك بالتأثير في اختلاف أنماط ارتباطات الجسيمات النووية وأسئلة الهستونات وفعالية إنزيمات فتح الأشرطة وغيرها من الأحداث التي لها علاقة بارتباط البروتينات بـ DNA . وقد لوحظ ان هناك تغيرات آنية تحدث لحالة التحلزن وزعزعة DNA عند بعض الاجهادات البيئية مثل :

- الصدمة الحرارية
- صدمة البرودة
- تغيرات الرقم الهيدروجيني pH
- التغير في الضغط التناظفي
- التحول من حالة التهوية الى حالة غياب الهواء
- حدوث المجاعة

والمعروف ان مثل هذه الاجهادات تغير التعبير الجيني في البكتريا . فضلا عن ذلك فإن التفاعلات التي تحصل على DNA مثل التضاعف والانتساخ تؤثر أيضا في حالة التحلزن الفائق ، فعند اندفاع RNAP تندفع أمامه موجة من حالة الحلزونات الفائقة الموجبة وتترك خلفها حلزونات سالبة الالتفاف وهذا سيؤثر في التعبير عن الجينات القريبة . والطريق التي

تؤثر فيها حالة التحلزن الفائق السالب في التنظيم يكون بواسطة التأثير في عدم ثبوت الحلزونات المزدوجة في المناطق التي تكون عرضة لها في التوالي ، والحقيقة ان عدم الثبوت حتى بقليل من الطاقة (كيلوسعرة) والتي تكون اقل من تلك اللازمة لفتح المزدوجات يكون لها تأثيرا " واضحا " .

وبهذه الطريقة فإن تحويل التحلزن الفائق يمكن ان يؤثر في نمط التعبير الجيني الموضعي او الشامل لعدد من الجينات ، وقد وجدت طرق لتقييم نمط الإجهاد الذي يؤدي الى عدم ثبوت مزدوجات DNA وهي مطابقة للتجارب العملية المتوفرة.

وتوزيع هذه المناطق في البكتريا يكون بنمط محدد فهي تكثر في المناطق البيئية الحاوية على الممهدات وبشكل اقل في المناطق المشفرة وهي مشابهة لحالة الخمائر ولكن في الأخيرة تكثر في النهايات التي تحيط بالجينات وليس الممهدات ، وعلى العموم فإن معظم الجينات التي تستجيب لحالة الزعزعة لمزدوجات DNA تحوي على هذه المواقع في الجهة اليسرى ' 5 . وقد وجد انه في 80% من المناطق البيئية القابلة للانفتاح والتي تحصل فيها حالة الزعزعة تكون حاوية على الممهدات اما المناطق البيئية التي تكون فيها مواقع الزعزعة ضعيفة فيكون وجود الممهدات فيها منخفضا" ويصل الى 24% .

وتحدد مواقع الزعزعة بقياس الطاقة الحرة لهذه المناطق (Gx) وتمثل Gx الطاقة الحرة اللازمة لفتح زوج القواعد عند الموقع X وأكثر مواقع عدم الثبوت لها قيم Gx تصل الى 8 - 10 كيلوسعرة ١ مول وهي التي تمثل المواقع التي تبقى ثابتة عند ظروف التراخي وتمثل الغالبية العظمى في الجينوم ، في حين ان المواقع ذات الزعزعة القوية فتحتاج الى طاقة قليلة او حتى بدون طاقة إضافية للفتح والتي تقرب قيم Gx لها من الصفر ، اما المناطق غير الثابتة جزئيا فتكون قيمتها بين الاثنين .

وتصنف الجينات اعتمادا" على احتواءها على مواقع الزعزعة في الطرف ' 5 ضمن 50 قاعدة من موقع البدء ، كما ان مواقع الزعزعة التي تحدث بواسطة الطفرات يمكن ان تعمل كأصول تضاعف . وفي خلايا بدائيات النواة فإن مستوى الزعزعة يتأثر بأطوار النمو وكذلك الظروف البيئية مثل التغير في الضغط التناظفي ووجود بيروكسيد الهيدروجين والإجهاد الحراري وغيرها من الظروف التي تؤدي الى تغير التعبير الجيني لذلك عدت من الآليات الشاملة من عمليات التنظيم .

وتستعمل مواقع الزعزعة في أيجاد الممهدات وتكون أكثر دقة من الطرق الأخرى ، فمثلا" تصل دقة استعمال المكوثرات السداسية Hexamer sequences الى حوالي 53% في حين تصل دقة استعمال مواقع الزعزعة الى حوالي 80% . وباستعمال هذا التوجه وجد ان أقوى المواقع في *E. coli* سجلت في الجينات التي تشفر لمنظمات البروتينات الغشائية وبروتينات النقل المستعملة في أنظمة مختلفة مثل أنظمة النقل ثنائية المكونات ، كما أشارت الدراسات الى ان الجينات التي تستجيب للظروف الفسلجية والبيئية تحوي على مواقع زعزعة وعدم ثبوت في الجهة اليسرى أكثر من غيرها .

التشوهات المستحثة بالبروتينات Protein – induced deformability

تشوه يحصل في DNA نتيجة الارتباط للبروتينات وهي إحدى الصفات التي تدخل في رسم أطالس الجينومات . وهي نموذج للنموكولوجيات الثنائية التي تحدد قيمتها باستعمال التراكيب

البلورية لمعقدات DNA مع البروتينات ، وتمثل قياس لمدى حجم الفراغ المجسم الذي يغطي بـ DNA dimmers في المعقدات مع البروتين مقارنة بالـ DNA النقي.

زاوية الالتواء Propeller twist angles

إحدى الصفات المستعملة في تحديد تركيب DNA ورسم الأطالس وتمثل الزاوية بين المستويات لاثنتين من النيوكليوتيدات الحلقية والاختلافات تعتمد على الانحراف المعياري الذي يكون أساسا" لإزاحة زوج القواعد الى الاتجاه القائم لمحور الحلزون . وبصورة عامة فإن النيوكليوتيدات الثنائية تكون لها قيمة التواء صغيرة وتكون مرنة وزاوية الالتواء تعد مقياسا لصلادة DNA وهي ذات علاقة عكسية .

الحساسية لإنزيمات القطع والانحنائية DNase I sensitivity and bendability

يقوم إنزيم DNase 1 بالارتباط وقطع DNA المنحني او القابل للانحناء باتجاه الأخدود الأكبر Major groove في مزدوجات DNA . وتردد مواقع القطع بهذا الإنزيم للـ DNA العاري يمكن ان تفسر على انها مقياس كمي لمدى انضغاط الأخدود الأكبر او الانحنائية (القابلية على الانحناء) المتباينة Anisotropic bendability والتي تستعمل لحساب مؤشرات قابلية الانحناء .

والحساسية للإنزيم تقيس المرونة ، فالقيم العالية لوحداث افتراضية تشير الى مرونة كبيرة . وأشارت دراسات الأشعة السينية لمعقدات DNA مع البروتينات والتشوهات الحاصلة للـ DNA نتيجة الارتباط بالبروتينات الى ان القيم العالية تكون مرتبطة بمستويات عالية من المرونة . وتستعمل المواقع الحساسة للـ DNase 1 في رسم الأثر Footprint لعوامل الانتساخ او مواقع الجسيمات النووية ، وبما ان المناطق الحساسة للإنزيم تحتاج الى انحناء المنطقة لذلك فهي تعد مقياسا" لقابلية الانحناء او ما يسمى بالمرونة متباينة الأوضاع Anisotropic flexibility . وبتطبيق تجارب DNase 1 على نموذج النيوكليوتيدات الثلاثية يمكن الحصول على عدد من المؤشرات والقياسات التي تشكل مقاييس لهذه الصفة ، فالقيم العالية تشير الى تواليات مرنة أكثر .

ان الانحنائية او المناطق القابلة للانحناء تمثل مواقع ارتباط البروتينات الى DNA وتتميز وتعتمد على الانحناء الداخلي والذاتي للـ DNA ومرونته وهي أيضا" تحدد الجسيمات النووية وقد تم تأكيد ذلك بالعديد من التجارب . واستعملت طرق تعتمد على المسح المجهرى لتوضيحها وهي اكفا من استعمال الطرق التقليدية مثل استعمال الأشعة السينية ، لان مثل هذه المواصفات تكون مهمة في العديد من الفعاليات الحيوية ولها تطبيقات في التقنيات الحيوية .

الجسيمات النووية

يكون DNA عددا" من التراكيب بطاقات مختلفة البعض منها ينتج بمساعدة البروتينات وأخرى تكون خاصة وناجمة عن تواليات DNA نفسها ، وانتظام توالي من القواعد يحدد تركيب DNA فالتوالي TTT . AAA و ATT . AAT في الأخدود الداخلي الصغير تكون مواجهة للهستونات ، في حين ان التوالي المكون من GCC . GGC والتوالي AGCC.GCT في الأخدود الخارجي ويمثل الجزء البعيد عن المجمع الثماني من الهستونات ، والمرونة الناتجة من النيوكليوتيدات الثلاثية تؤثر في موقع وترتيب الجسيمات النووية في DNA أي تفضيل الموقع

Position preference ، فالقيم المحسوبة تحدد التفضيل اذ ان القيم الواطئة لها علاقة بالمرونة العالية ، وعليه فأن مواقع الجسيمات النووية يحدد مكانها بإشارات ثابتة وأخرى داينمكية تشفر بالتواليات في DNA . ومكان الجسيمات النووية يمكن ان يحدد بأطالس الجينومات وذلك بأجراء قياس للمرونة المعتمد على استعمال قيم لـ 32 من قيم النيوكلويدات الثلاثية الخاصة بالأخدود الصغير الخارجي ، وعند التقاف DNA على الجسيمات النووية فالقيم صفر تشير الى عدم وجود تفضيل بينما القيم المطلقة العالية تشير الى ان النيوكلويدات الثلاثية يكون لها تفضيل عالي .

مكونات القواعد وانحيازها

مر سابقا إمكانية تحديد محتوى الجينومات من القواعد النتروجينية وهذه المكونات تدخل كمؤشرات قوية في رسم الأطالس . وأبسط الطرق هو تحديد كمية كل من القواعد النتروجينية ولكن مجموعهم يكون واحد لذلك كان البديل استعمال A+T او A-T او G-C الذي يكون أسهل في التفسير في السياق البايولوجي وهذه تحدد في كل من الأشرطة القائدة والمتخلفة . فالمؤشر A+T يكون له علاقة مع المؤشرات التركيبية الأخرى خاصة طاقة الرص ، والمناطق الغنية بـ A+T يزال ثبوتها بسهولة وفيها مستويات عالية من النقوس الذاتي فضلا عن انها اقل مرونة . لذلك فأن محتوى A+T يحوي على كل المعلومات التركيبية التي تنشأ او ذات علاقة بالنيوكلويدات الأحادية .

اما مؤشرات G-C ، A-T فهؤلاء لهم علاقة بقياسات الانحرافات Skews التي تكون مهمة في تحديد مناطق بدء التضاعف وانتهاءه في البكتريا وتكون ضمن العلاقات الآتية

$$\begin{aligned} \text{GC skew} &= (G - C) / (G + C) \\ \text{AT skew} &= (A - T) / (A + T) \\ \text{Purine skew} &= (G + A - T - G) / N \\ \text{Keto skew} &= (G + T - A - C) / N \end{aligned}$$

ان الاختلافات في انحياز AT skew و GC skew يعتمد على المحتوى من A+T لذلك يحدث تذبذب كبير في GC skew في المناطق الغنية بـ AT والعكس صحيح بالنسبة للـ AT skew . والانحياز GC skew يكون أكثر فائدة واستعمالا في إيجاد أصل التضاعف ونهايته لذلك يستعمل في رسم الأطالس وباستعمال برامج خاصة

المكررات Repeats

من المعروف ان المكررات على اختلاف أنواعها يمكن ان تشارك في تحديد تراكيب DNA ولذلك فهي تساهم ويتم إظهارها في أطالس الجينومات وقد مر مناقشة أنواعها المختلفة سابقا .

العلاقات بين المؤشرات المستعملة في رسم الأطالس

هناك علاقات مختلفة سواء بسيطة او معقدة بين المؤشرات المارة الذكر . واغلب المؤشرات لها علاقة بمحتوى الجينوم من AT ويكون الارتباط بين المؤشرات في DNA البايولوجي هو أعلى من القيم الحسابية المستعملة في الحسابات فالقطع الطويلة تعطي نتائج مختلفة نوعا ما عن تلك المعتمدة على القطع القصيرة ، فضلا عن ان بعض المؤشرات تتأثر بالقواعد

المجاورة ولذلك اقترح ان أفضل الحسابات لإيجاد العلاقة بين المؤشرات ان تكون معتمدة على النيوكليوتيدات الثنائية والنيوكليوتيدات الثلاثية .

ومن العلاقات التي يمكن ذكرها على سبيل المثال لا الحصر والأمثلة التي سيتم ذكرها تتدرج من علاقة قوية الى ضعف هذه العلاقة ثم انعدامها كما في علاقة زوايا الالتواء مع AT ، وكذلك علاقة التشوهات التي تظهر في DNA نتيجة الارتباط بالبروتينات مع المحتوى من AT . ويلاحظ ان علاقة طاقة الرص تتناسب عكسياً مع زوايا الالتواء والتشوه المستحث بالبروتينات ، في حين يلاحظ ضعف العلاقة للمؤشرات المذكورة مع تقوس جزيئات DNA ، وقد تصل العلاقة الى الصفر بين الحساسية لإنزيم DNase 1 مع تفضيل الموقع

برامج رسم الأطالس

تختلف البرامج المستعملة في رسم الأطالس اعتماداً على الهدف المرجو منها ، فالبعض يهدف لإيضاح صفات DNA التركيبية او تحديد مواقع التواليات القصيرة المكررة او إيجاد مجالات Lanes خاصة مثل تلك المتعلقة بتغايرات النيوكليوتيد المفرد SNPs او بيانات المصفوفات الدقيقة او إجراء مقارنة للجينومات وغيرها من الأهداف .

وتتفق معظم البرامج بأن مخرجات الاستعلام لكل طريقة محددة بأرقام منسوبة الى تواليات مرجعية وتتفق أيضاً باحتوائها على وسائل للتخلص من المكررات والإبقاء على المقروءات الفريدة التي تطابق الجينوم المرجعي المستعمل أي انها تحوي على وسائل تصفية تختلف من برنامج الى آخر .

كما ان اغلب البرامج تعتمد القياسات الخاصة بالنيوكليوتيدات الثنائية والثلاثية في العمليات الحسابية . ومعظم برامج رسم الأطالس تستقي معلوماتها من قواعد المركز الأوروبي وبنك الجينات وكذلك المركز الياباني ، والبرامج الجيدة والحديثة تستقي معلوماتها أيضاً من قاعدة البيانات المباشرة GOLD وذلك لان هذه القاعدة تحوي الجينومات الجديدة التي لم تودع بعد في قواعد البيانات السابقة الذكر لتلافي أي نقص في المعلومات يحدث بين القواعد العامة وان كانت مرتبطة مع بعضها وفي هذه الحالة لا بد من ارتباطها بوسائل تصفية لإزالة المكرر والفائض من المعلومات عن التواليات .

والمستخدم لهذه البرامج والوسائط يفضل استعمال البرامج التي تكون مخرجاتها قابلة لتكبير لكي يمكن تفسير النتائج على مقاييس متغيرة . والبعض يفضل الأطالس المعتمدة على التواليات والتي تكون لها برامجها الخاصة والآخر يفضل الأطالس المعتمدة على الصفات التركيبية وغيرها من الصفات وفي رأي الكاتب يكون من الأفضل استعمال التوجيهين لأنها ستغطي الكثير من الجوانب التي ربما عجز احد التوجيهات عن تغطيتها وذلك لأن إظهار المعلومات متعددة الأبعاد التي تمثل بتوالي الجينوم فقط لازالت صعبة . ومن الجدير بالذكر ان البرامج المتوفرة حالياً البعض منها معتمد على برامج سابقة تم تطويرها لتواكب الحاجات المطلوبة .

وبرامج رسم الأطالس مكتوبة بأكثر من لغة ولها وسائل إظهار الجينومات وتحليلها على مستويات مختلفة وذلك لاستعمالها ملفات مختلفة لإنتاج أطالس لتحليل أفضل للجينومات وترتبط الى أكثر من موقع لاستقاء المعلومات وتضطلع القاعدة الدانمركية CBS أكثر من غيرها في معالجة ورسم الأطالس .

ولازالت بعض البرامج المستعملة مثل DNA plotter تقوم برسم الأطالس ولكن مخرجات هذه البرامج والمواقع توفر القليل من المعلومات مما يجعل الرسوم الناتجة صعبة التفسير فضلا عن انها توفر أطالس لجينومات مفردة ولا توفر إمكانيات المقارنة بين عدة جينومات . وقد تم تطوير برامج أفضل مثل GeneWiz الذي يضطلع برسم أطالس DNA المختلفة ويزود بمختلف القياسات فضلا عن احتوائه على وسائل وأنظمة تصفية جيدة ، ويعد أكثر البرامج استعمالا في الوقت الحاضر، وواجهة مزود الخدمة التي تضم هذا البرنامج توفر إمكانية اختيار البكتريا وسلالاتها فضلا عن إمكانية اختيار جزيئات DNA هل هي حلقية ام مفتوحة وإمكانية التحكم بدرجة الوضوح وكذلك إمكانية اختيار الصفات المطلوب إظهارها في الأطالس كما موضح في الشكل 105.5 :

GeneWiz browser 0.94 server

Custom atlas Supplemental data Pre-made atlases Linking

REFERENCE SEQUENCE AND ANNOTATION: Under test right now

Submit the reference genome in GenBank format directly from your local disk: [Browse...](#) [Example genbank file](#)

Search: (1160)

specify a genbank accession number from the scroll down list: [select organism] [select replicon]

'Nostoc azollae' 0708 (PID 30807)
Acaryochloris marina MBIC11017 (PID 12997)
Acetobacter pasteurianus IFO 3283-01 (PID 31129)

or - upload custom Fasta file [Browse...](#) [Example fasta file](#)
and annotation file [Browse...](#) [Example annotation file](#)

ATLAS OPTIONS

Main title and subtitle: Main title
Sub title

Modus: Circular atlas

Resolution: Medium

Atlas stamp: GENOME ATLAS

Reuse cached results: ☒

Lane order: BLAST LANES
MAPPING SHORT SEQ. READS
(feature)

شكل 105.5 : واجهة GeneWiz لرسم أطالس الجينومات

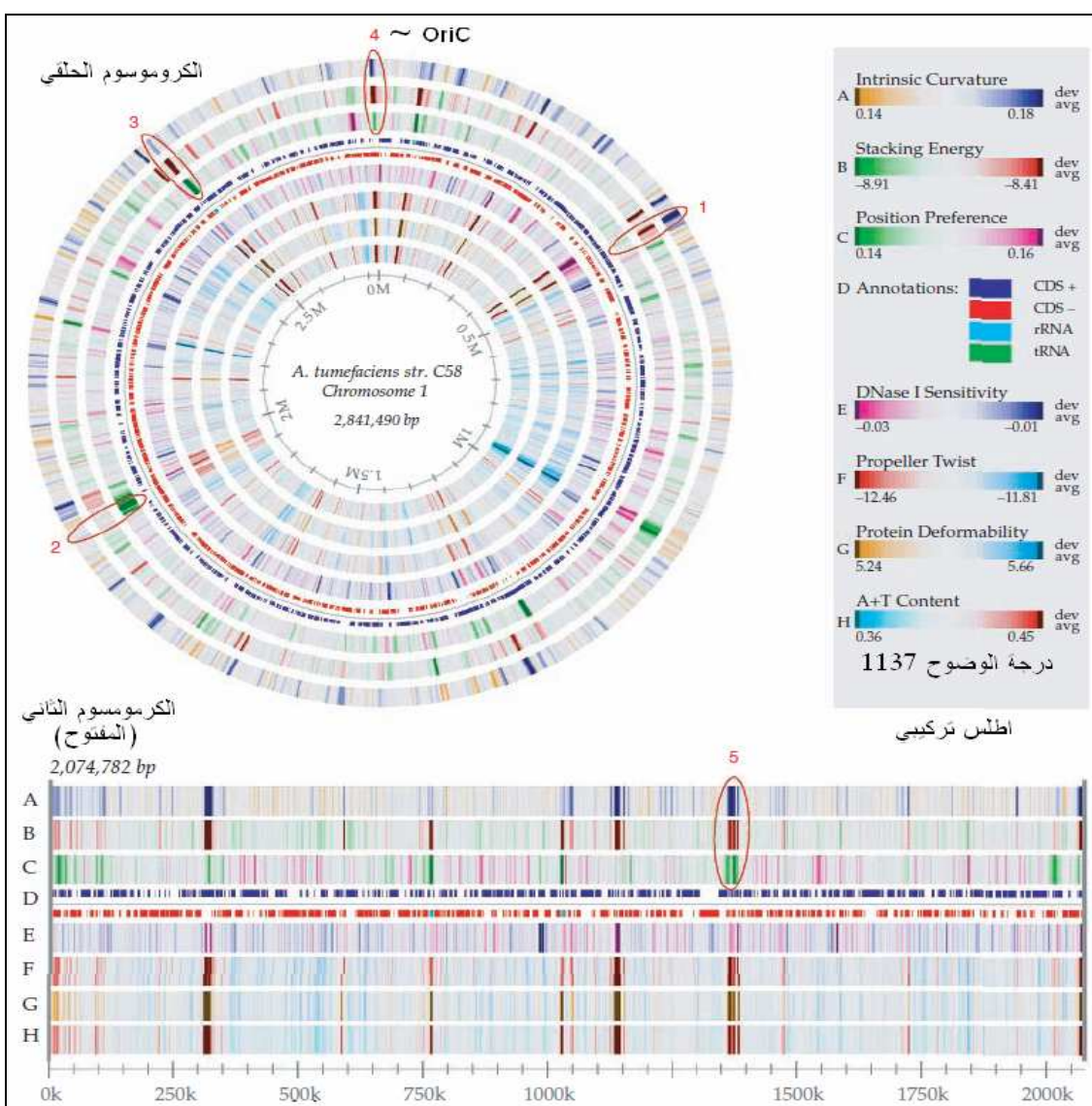
ومن الجدير بالذكر ان الأطالس الممكن رسمها باستعمال GeneWiz تكون للأحياء المدرجة ضمن قواعد بيانات القاعدة الدانمركية CBS ، اما في حالة رسم أطالس جينوم جديد فان ذلك يتطلب وجود ملف تهميش كامل Annotation file يوفر الكم الهائل من البيانات الإضافية عن الجينوم ، فضلا عن الحاجة الى المعلومات بصيغة بنك الجينات لتوفير المزيد من البيانات

أنواع الأطالس

هناك أنواع مختلفة من الأطالس البعض منها عام والآخر خاص ، فضلا عن إمكانية وجود أطالس تخص صفات معينة للجينوم الواحد والأخرى التي توفر إمكانية مقارنة الجينومات المتعددة لبعض الصفات ، وفيما يلي بعض أنواع الأطالس الممثلة

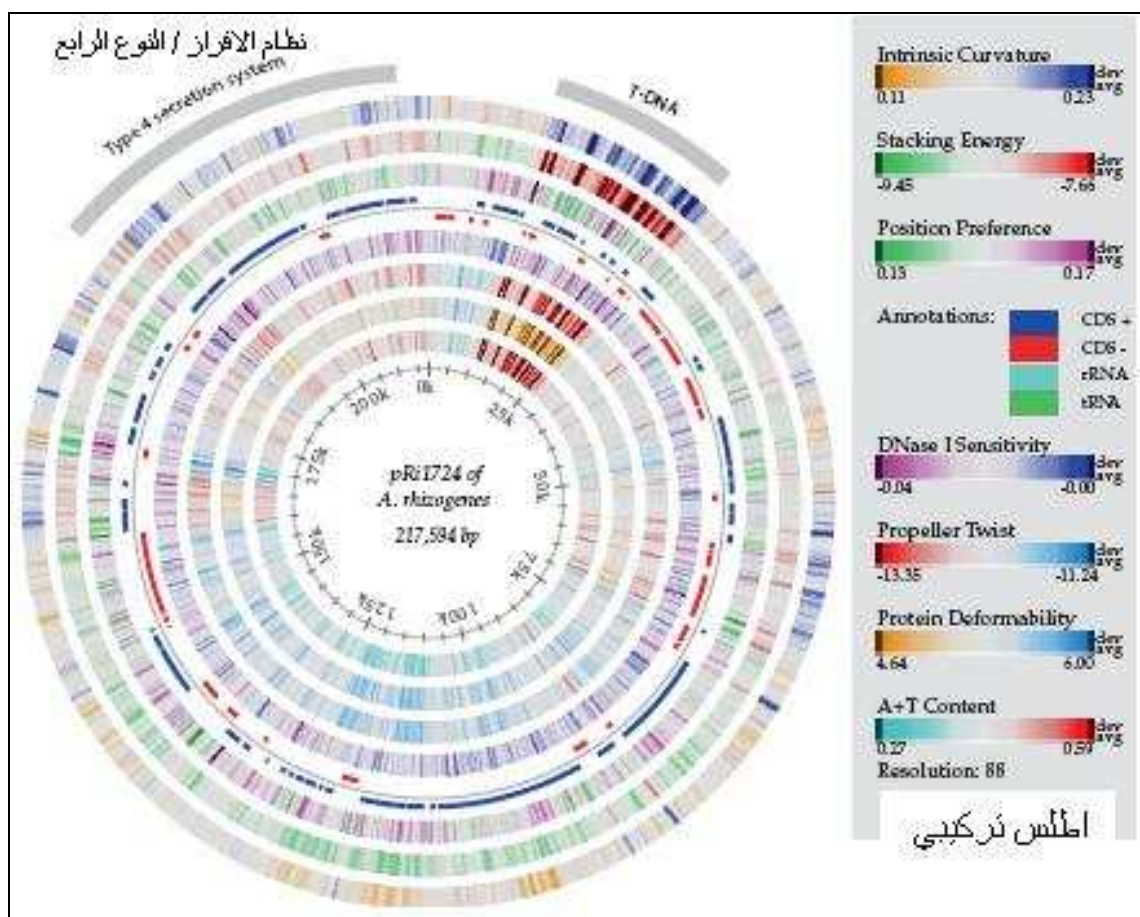
أطلس تركيب الجينوم Structural genomic atlas

الأطلس يكون بسيطا" يرسم بدوائر كما موضح في الشكل التالي (شكل 106.5) الذي يمثل كروموسومات البكتريا *Agrobacterium tumefaciens* الحلقي والمفتوح



والملاحظ من الشكل ظهور حجم جينوم السلالة مقدرا " بعدد القواعد فضلا" عن ظهور التقسيمات في الدائرة الداخلية التي تمثل الجينوم مقسما" على الدقائق الزمنية ،والصفات ظاهرة في المقطع الملحق بالأطلس الذي يوضح الصفات الظاهرة فيه ، ويظهر الأطلس موقع بدأ التضاعف Ori C . وتدرجات اللون موضحة في الشكل ويجب الانتباه الى ان الصفات المدرجة في التوضيح الجانبي ليس بالضرورة ان تكون موافقة لترتيب الدوائر في الأطلس لذلك لابد من العودة الى ملفات التوضيح ، ويظهر الأطلس بعض الجينات المهمة ودرجة الوضوح ويمكن تكبير اي منطقة لدراستها وتوضيحها .

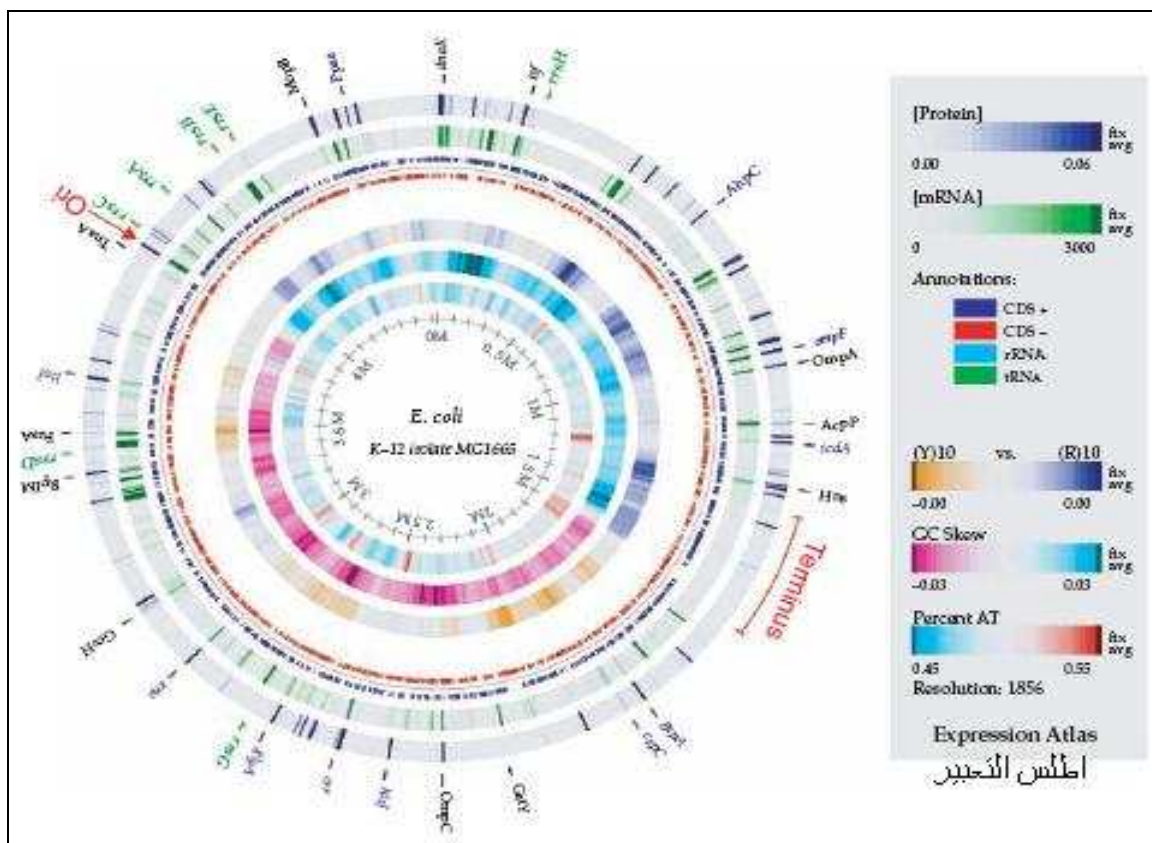
والأطلس لا تقتصر على الجينومات البكتيرية وانما يمكن ان تستعمل لدراسة البلازميدات ، والشكل التالي (107.5) يوضح الأطلس التركيبي للبلازميد الذي يسبب الشعيرات الجذرية للبكتريا *Ag. rhizogenes*



شكل 107.5 : الأطلس التركيبي للبلازميد pRi1724 لبكتريا *Ag. rhizogenes*

ويوضح الأطلس المؤشرات التركيبية الموضحة بالرسم الملحق (الأيمن) فضلا" عن إظهار المنطقة الخاصة ب T- DNA ونظام الإفراز الرابع .

أطلس التعبير الجيني Gene expression
يكون بمثابة وسيلة للحصول على المعلومات حول التعبير الجيني للأنواع والشكل التالي
(108.5) يوضح أطلس التعبير الجيني لإحدى سلالات *E. coli*



شكل 108.5 : أطلس التعبير الجيني لبكتريا *E. coli* K12

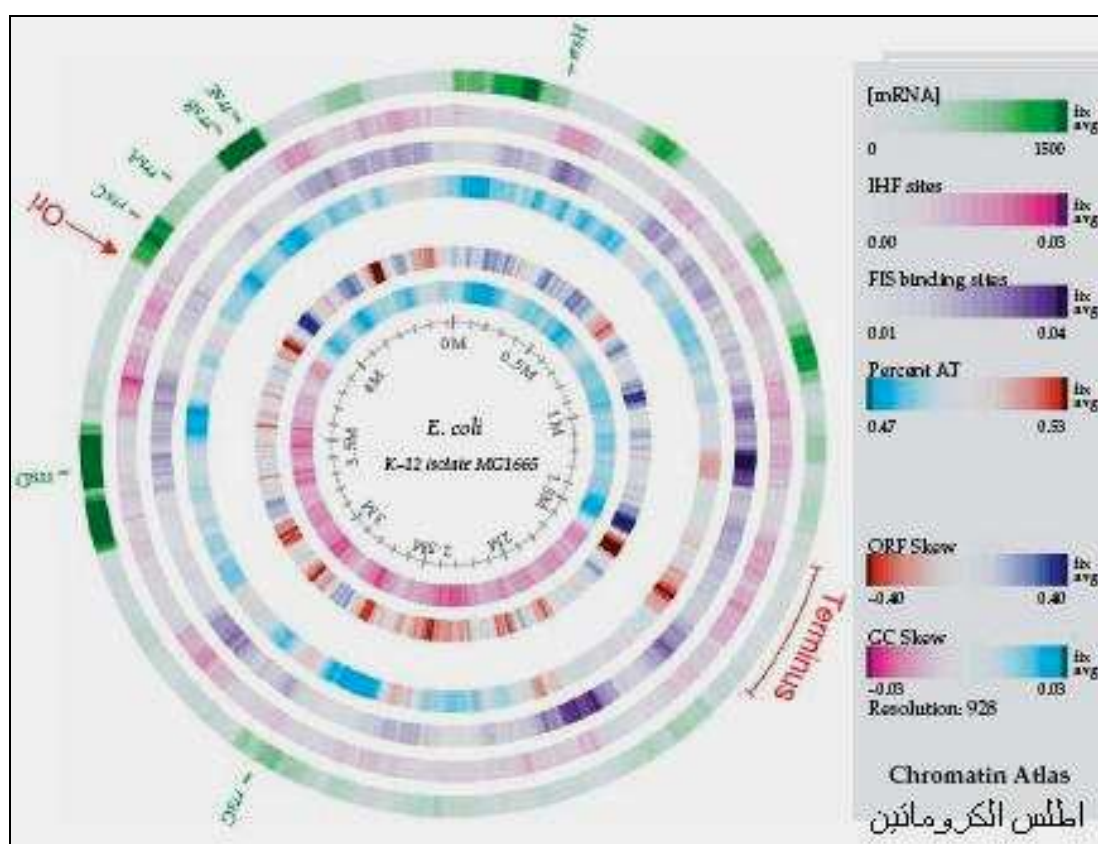
ويلاحظ الاختلاف في الصفات التي يغطيها الأطلس عن تلك المذكورة في الأطلس السابق ، فهو يوضح المناطق المشفرة وغير المشفرة من الجين ، وكذلك جينات tRNA و rRNA فضلا عن إظهار المواقع التقليدية مثل نقطة بدأ الانتساخ وانتهائه وبيانات أخرى مثل تراكيز البروتين المعتمدة على طرق خاصة (بالدراسة) وكذلك كميات mRNA المقاسة باعتباره الصفة المميزة للتعبير الجيني .

(و أطلس التعبير المذكور بياناته معتمدة على قواعد مركز الاتحاد الأوروبي EMBL) . ويمكن باستعمال طريقة رسم أطالس التعبير الجيني باستخدام كلمات استعمال خاصة مثل ذكر اسم الجين لتظهر النتيجة الخاصة بالتعبير عنه ، أو ذكر الظروف فتظهر الجينات المتأثرة بها ، أو يمكن ذكر اسم المرض لتظهر الجينات المتأثرة ومدى التغير في التعبير عنها مقارنة بالحالة الطبيعية ، ويمكن الربط بأكثر من استعمال لتظهر النتائج مرتبة اعتمادا على مؤشرات إحصائية مثل P- value .

والطرق المستعملة في تحديد مخرجات مثل هذه البرامج هي طرق المصفوفات الدقيقة والطرق السريعة الأخرى . وقد جمعت البيانات الخاصة بالتعبير الجيني في قواعد بيانات ثانوية يمكن ان تستغل في رسم الأطالس ، اذ يرتبط فيها كل جين مع الظروف التي يتعامل معها ، ويمكن للمستخدم إدخال أسئلة معقدة للوصول الى النتائج المطلوبة

أطلس الكروماتين Chromatin atlas

يوضح الشكل التالي (109.5) الأطلس الخاص بالكروماتين للبكتريا *E. coli* والسلالة المستعملة في الشكل السابق .

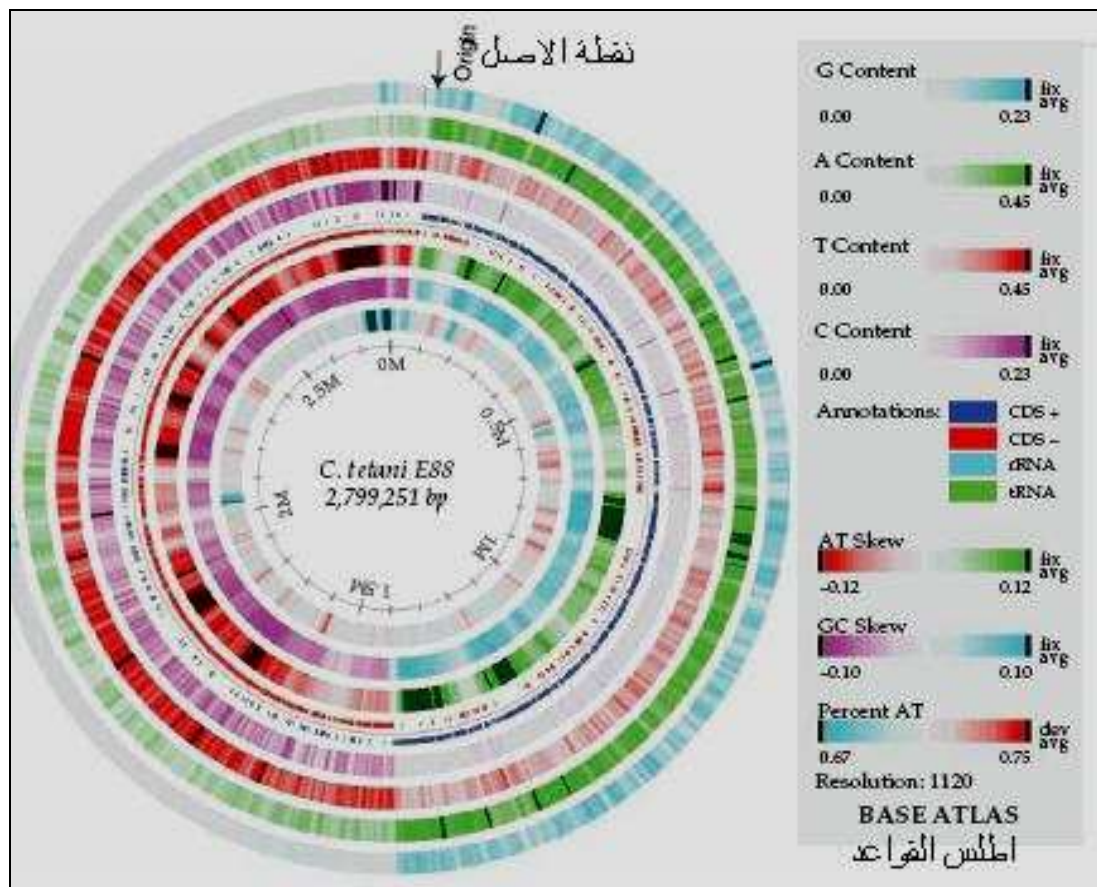


شكل 109.5 : أطلس الكروماتين للبكتريا *E. coli* K12

وفيه تظهر المظاهر التقليدية العامة فضلا عن إظهار مواقع وشدة وجود بعض مشابهاة الهستونات (الكروماتين) في البكتريا (*E. coli*)

أطلس القواعد Base atlas

نوع من الأطالس توضح فيه مكونات الجينوم من القواعد النتروجينية وإظهار الانحيازات سواء للـ GC أو AT والشكل التالي (110.5) يوضح احد هذه الأطالس للبكتريا *Clostridium tetani* السلالة E 88 .

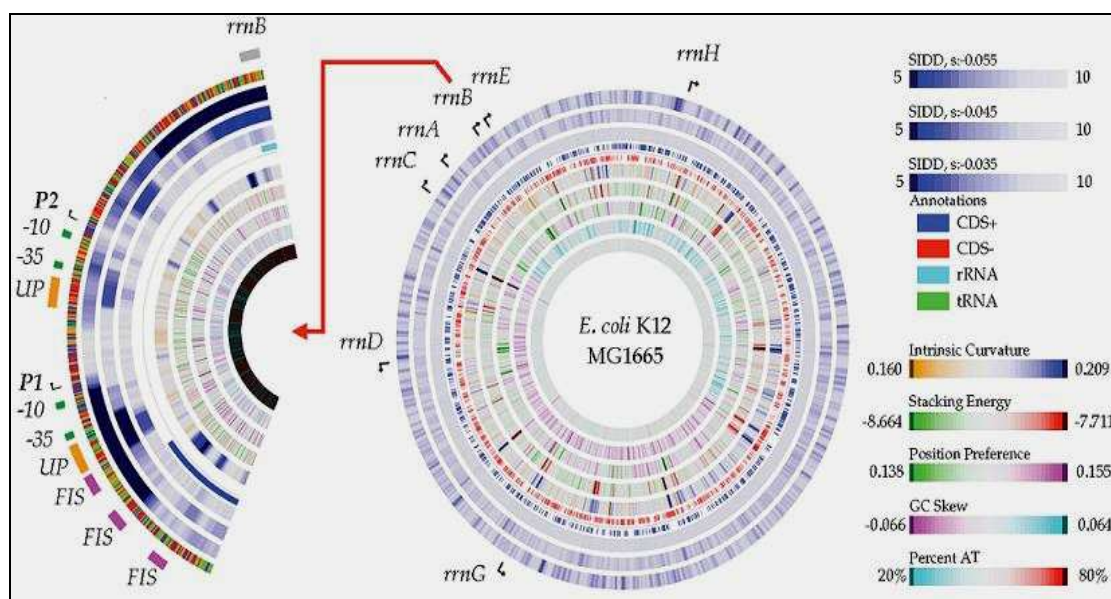


شكل 110.5 : أطلس القواعد النروجينية للبكتريا *Clostridium tetani* السلالة E 88 .

ويوضح الشكل المواصفات العامة فضلاً عن إيضاحه لميل الجينومات الى احتواء الكوانين في الأشرطة القائدة والأدينين في الأشرطة الأخرى .

أطلس الزعزعة SIDD atlases

أطالس تهتم بتوضيح الممهدات وعناصر التنظيم ، لذلك فهي تساعد في التنبؤ بالممهدات وعناصر التنظيم وتستعمل قاعدة بيانات خاصة لهذه الصفة كما ذكر سابقاً . والشكل التالي يوضح أطلس الزعزعة SIDD لإحدى سلالات *E. coli* ، وتكون مثل هذه الأطالس قابلة للتكبير كما هو الحال مع غيرها من الأطالس كما موضح في الشكل (111.5) وتحسب فيها قيم SIDD استناداً الى قواعد البيانات المستعملة ، ويظهر الشكل الصفات التركيبية ومواقع rrm

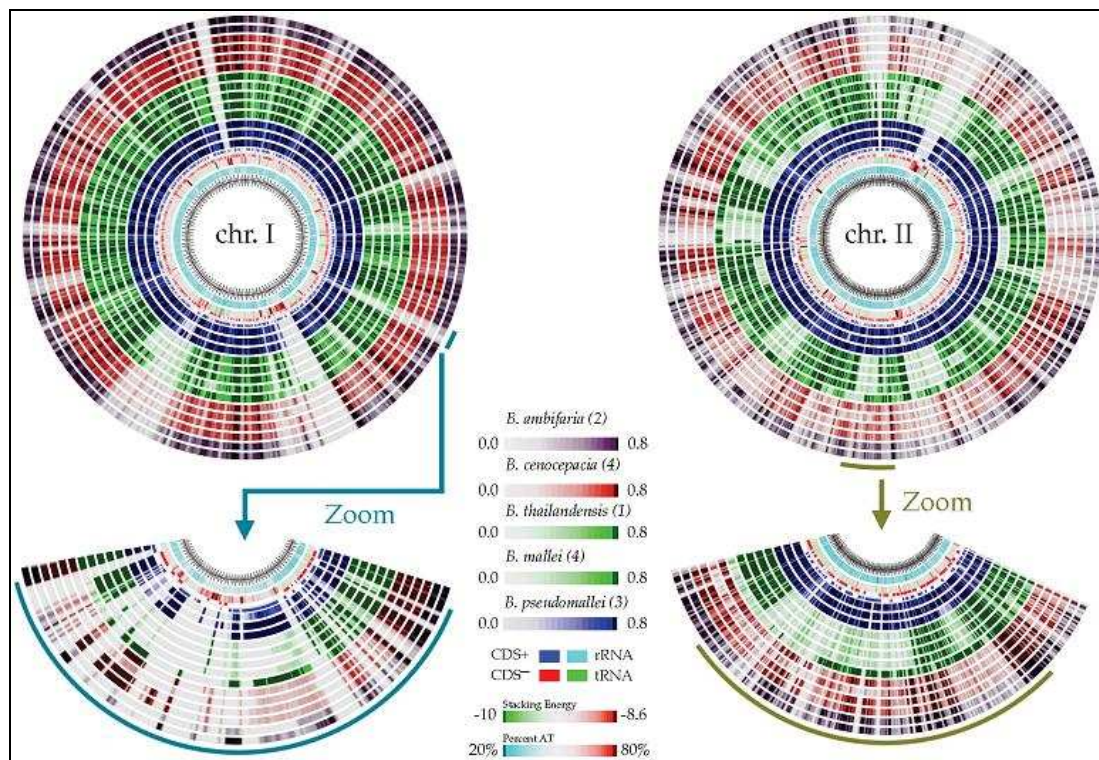


شكل 111.5 : أطلس الزعزعة SIDD للبكتريا *E. coli* K12

وتشير الأرقام الى قيم الطاقة اللازمة لفتح الأشرطة المحسوبة باستعمال خوارزمية SIDD

أطالس المقارنة Comparative atlases

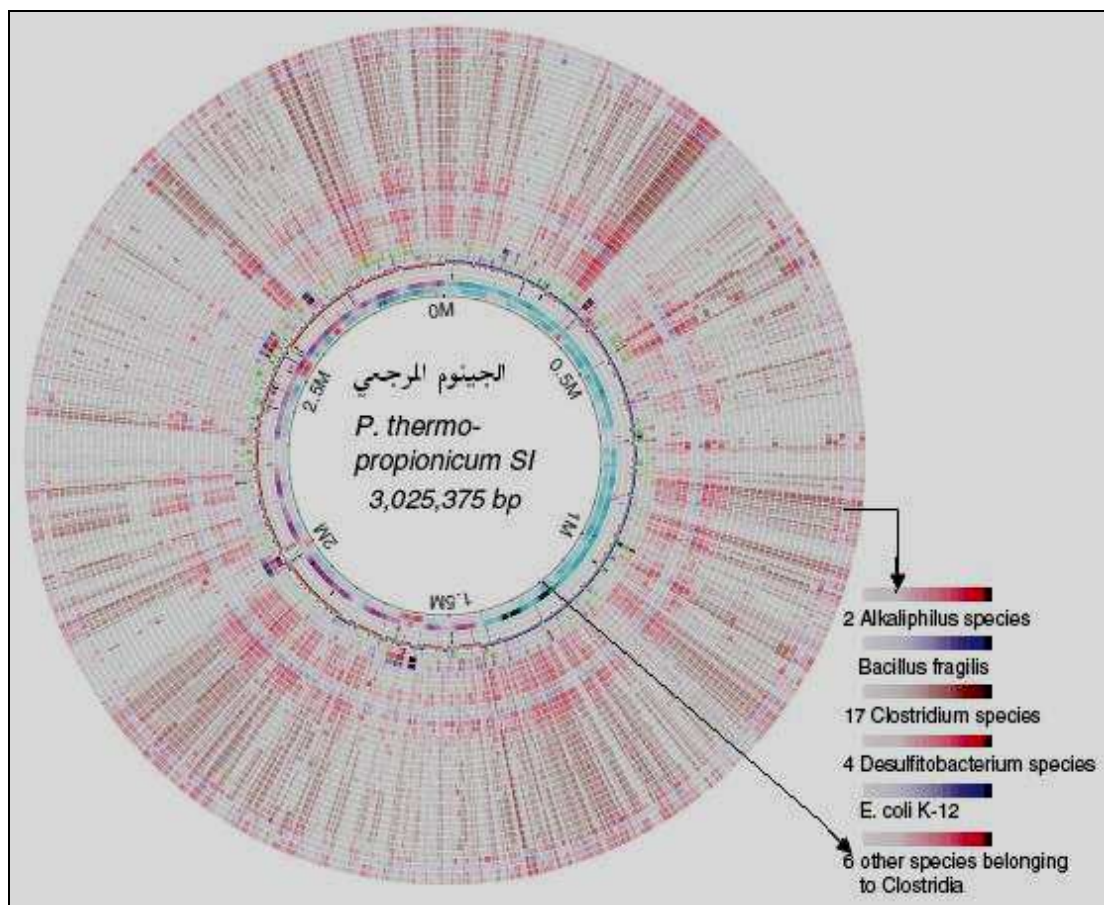
ذكر أعلاه بعض أنواع الأطالس وهناك أطالس أخرى يمكن ان تمثل عليها صفات أخرى ، وما ذكر كان خاصا" بكائن واحد لإبراز عدة صفات ، ولكن هناك أطالس أخرى تهدف الى إجراء المقارنات بين السلالات المختلفة ، او مقارنة الجينومات أحياء مختلفة . والغرض من وضع هذه الأطالس هو إجراء المقارنة لأغراض معينة مثل مقارنة جزر الامراضية بين السلالات البيئية والسلالات المرضية التي تعود الى النوع نفسه من البكتريا وكذلك التعرف على مناطق إقحام العاثيات او فقدان بعض المواد الوراثية المهمة ، ومواقع العنقدة وغيرها . ومثل هذه الأطالس تكون مهمة لمقارنة العديد من سلالات النوع الواحد التي تم تحديد توالياتها ومنها الموضح في الشكل 112.5.



شكل 112.5 : أطلس المقارنة لأنواع مختلفة من جنس البكتريا *Burkholderia*

ومن أهم الأطلال المقارنة هي BLAST atlas ، إذ أن من المعروف أن BLAST انشأ لإجراء المقارنات لذلك ربط مع برامج رسم الأطلال لتحقيق هذا الغرض . ويمكن بواسطة هذا البرنامج وضع خرائط للنوكليوتيدات أو للحوامض الامينية التي يشفر لها بعد الربط بالبرامج الخاصة بها مثل BLASTn للنوكليوتيدات أو BLASTp للبروتينات أو غيرها من إصدارات برنامج BLAST وفق الغرض المطلوب . والبرنامج يمكن أن يربط مع ACT لمقارنة 2 أو أكثر من الجينومات ، وتتفق وسائل رسم أطلال المقارنة بوجود جينوم مرجعي وبذلك يمكن تحديد الجينوم الشامل Pan genome والجينوم المركزي Core genome . وفي هذا النوع من الأطلال يتم تحديد التطابقات ودرجاتها بالاعتماد على الشفرات اللونية التي تستعمل نظام RGB على مقياس 0-10 وتكون القيم المتوسطة بلون رمادي فاتح . وفي مثل هذه البرامج يمكن التحكم بالصفات المراد دراستها بما توفره في خيارات مثل نوعية جزيئة DNA هل هي حلقية ام مفتوحة واطر القراءة المفتوحة ومتجاورات الكروموسوم وصفات أخرى على أن تكون متوفرة في الجينوم المرجعي لذلك يكون من الأفضل استعمال أكبر جينوم موجود كمرجع وكلها تجمع وتعالج بالبرنامج دون الحاجة الى تدخل المستخدم . بعض الأطلال تظهر بشكل دوائر متحدة المركز (في أغلب الأحيان ووفق الطلب) ومعها المؤشرات المستعملة . وفي الوقت الحاضر توفر هذه الوسائل مقارنة عشرات الجينومات ولكن الطموح الى المئات والآلاف من التواليات قائما .

ويمكن استعمال مثل هذه الأطالس لرسم أطالس الجينومات البيئية Metagenomes ورسم الجينات بغض النظر عن أصولها بالاستناد الى جينوم مرجعي والشكل التالي (113.5) يوضح احد هذه الأطالس باستعمال جينوم *Pelotomaculum thermopropionicum* كمرجع لانه من الجينومات الكبيرة .

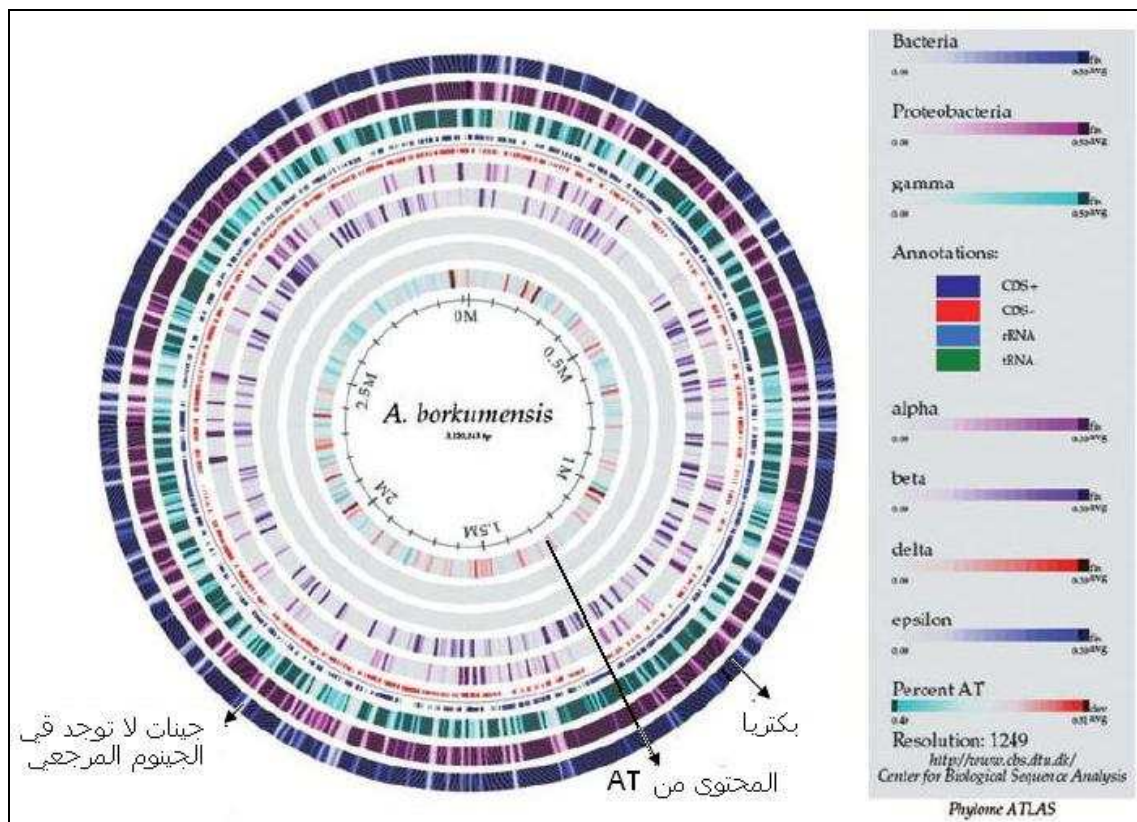


شكل 113.5 : أطلس لمقارنة 31 سلالة من بكتريا مختلفة

ونظرا " لكثرة الجينومات المستعملة فضلا" عن قلة التطابقات (BLAST hits) الموجودة في النماذج البيئية يلاحظ ان الألوان تكون باهتة . ويظهر الشكل مقارنة 31 نوع من أجناس بكتيرية مختلفة

أطالس العلاقات التطورية Phylome atlas

نوع آخر من أطالس المقارنة والذي يعتمد الى استعمال جينوم مرجعي لنسب التغيرات اليه . والشكل التالي (114.5) يوضح احد الأطالس المستعملة لإيجاد العلاقات بين الأحياء



شكل 114.5 : أطلس العلاقات التطورية بين مجاميع بكتيرية باستعمال *Alcanivorax borkumensis* كجينوم مرجعي

ويوضح الشكل وفق تدرج البيانات (الأيمن) البكتيريا عامة وأقسامها اما الدائرة الداخلية فتوضح المحتوى للـ AT ، اما الخطوط الباهتة فتشير الى وجود جينات ليس لها متشابهات في الجينوم المرجعي .

المعلومات الحيوية والبروتينات

تؤدي البروتينات أدواراً أساسية في الآلاف من العمليات الحيوية التي تحدث في الخلية ، وكذلك تشارك كمكونات أساسية في التراكيب الخلوية مثل الأغشية الخلوية والإسناد الميكانيكي المتمثل بالهيكل الخلوي Cytoskeletons .

وتوالي الحوامض الأمينية فيها يساعد في تكوين ملايين الأشكال المختلفة تساهم فيها مجموعة متخصصة من البروتينات وهي الوصيفات التي تساعد في طوي البروتينات بشكل صحيح لأداء وظائفها وتساعد المعلوماتية الحيوية في تحليل وإيجاد الجوانب المختلفة للبروتينات وإجراء الإحصاء الحيوي للبروتينات لتحديد أشجار العلاقات التطورية وتحديد نسق التعبير الجيني ، وتداخلات البروتينات مع بعضها ، وهذا ما سيتناوله هذا الجزء من الكتاب .

أهمية دراسة البروتينات

هناك عدة جوانب تحتم دراسة البروتينات وإدخالها في برامج مختلفة لاستقاء معلومات خاصة عنها . فالتعبير في توالي الحوامض الامينية يمكن ان يؤدي الى الأمراض مثل فقر الدم المنجلي Sickle – cell anemia والتحوصل الليفي وغيرها من الأمراض يمكن ان تنتج من تغيير حامض اميني واحد في البروتين . ومن جهة ثانية فأن تحديد توالي الحوامض الامينية يوضح العلاقة التطورية التاريخية بين الأحياء ، فعند تشابه توالي بروتين مع بروتين آخر فمن المحتمل أنهم يشتركون بأصل واحد ولذلك فأن التغييرات الجزيئية في حالة التطور يمكن تتبعها من تواليات الحوامض الامينية كما هو أساس البحث في علم أشكال الحياة القديمة Molecular paleontology .

معلومات دراسة البروتين

في الأنظمة الحيوية هناك أربع قواعد نتروجينية تشكل كل ثلاث منها شفرة وراثية لترجم الى حامض اميني ، ولكن بعض الحوامض الامينية يشفر لها بأكثر من شفرة ، فضلا" عن ان فيها مجاميع مختلفة من حيث الطبيعة الكيماوية وتفضيل بعض الحوامض ان تكون ضمن تراكيب معينة والتي سيأتي ذكرها لاحقا" (فضلا عن ان بعض الجوانب قد مر ذكرها) .

وبالرغم من وجود عدة طرق لتحديد تركيب البروتين الا انه من الصعوبة تحديد تركيب البروتينات المكتشفة حديثا" ، لذلك يتم اللجوء الى طرق المعلوماتية الحيوية والطريقة التقليدية في هذه الحالة هي مقارنة توالي البروتين مع قواعد البيانات كخطوة اولى وفي حالة الحصول على تشابه قوي لبروتين الاستعلام مع بروتين له وظيفة معروفة فهذا يشير الى تشابه بروتين الاستعلام ومن ورائه توالي الجين فيه مع الكائن الحاوي عليه اي كائن المقارنة .

ولكن هذه ليست الحقيقة اذ ان نتائج الحاسوب يجب ان تدعم بالتجارب المختبرية الخاصة بتحديد الوظيفة ولكن هذه العملية لا تبدو سهلة في ضوء عدد البروتينات الهائل التي تشفر بالجينومات سواء البكتيرية او خلايا حقيقية النواة . فنشير الإحصائيات ان تماثل البروتينات بمستوى 50% يعد مهما" بشكل كبير ولكن هذه واجهة اما التفاصيل فيمكن ان تشير الى غير ذلك لان التماثل بنسبة 30% او اقل يمكن ان يؤدي الى أداء الوظيفة نفسها وعليه فلا يمكن إعطاء نسبة فاصلة في هذا المجال . وهذا نابعا" من حقيقة ان الجينات ومن ورائها البروتينات يمكن ان تشترك بالوظيفة نفسها مع ان أصولها الوراثية مختلفة وهذه تنتج من عمليات التطور المتوازية وليس التماثل او التماثل الجذري .

ومن جهة ثانية فأن برامج BLAST الذي يستعمل بكثرة في إجراء المقارنات يقوم ايضا" بعمل مقارنات موضوعية للبحث عن التشابه في مواضع معينة وبالتالي يمكن ان يكشف عن الدومينات الفعالة وفي هذا المجال ولتحديد وظيفة البروتين كان من الواجب الانتباه الى وجود الدومينات الثابتة الفعالة التي توضح وظيفة البروتين الجديد . كما ان البحث عن وظيفة بروتين معين يكون ضمن افتراض ان البروتين يؤدي وظيفة واحدة محددة في الخلية وهذا غير صحيح نظرا" لان البروتين الواحد يمكن ان يؤدي عدة وظائف مختلفة تحت الظروف المختلفة لذا كان الحذر واجبا" من تحديد وظيفة البروتين بالاعتماد على تحديد التوالي لوحده .

توجهات التنبؤ بالبروتينات

عند التعامل مع البروتينات يؤخذ بنظر الاعتبار ان وظائف البروتينات تتحدد بتراكيبها على مختلف المستويات لذلك كانت المعلوماتية الحيوية ضرورية للتعامل مع البروتينات وهناك اكثر من طريقة يمكن إتباعها . ولكن عندما يكون العمل مع تواليات نيوكلوتيديّة والمراد التعامل معها في مجالات البروتين ، يجب في البداية اكتشاف هل ان التواليات مشفرة للبروتينات ام لا ، ويكون التحري في هذه الحالة عن :

- مكونات التوالي من النيوكلوتيديات وخاصة G+C حيث ان الانترونات تكون عادة غنية بـ A+T مقارنة بالاكسونات خاصة في النباتات .

- مكونات الشفرات الوراثية

- تكرار المكوثر السداسي من استعمال أعداد كبيرة ومختلفة من مقاييس التشفير التي اختبرت عن استعمال المكوثر السداسي (اي استعمال توالي بطول 6 نيوكلوتيدي) وهذه تكون ذات قابلية تمييزية عالية بين المناطق المشفرة وغير المشفرة.

والطرق هي :

الاصطفاف ومعرفة التشابه : وهي الطريق الشائعة بعد تحديد توالي البروتينات او الببتيدات والأساس في هذه الطريقة هو ان تشابه البروتينات يعني في الغالب أدائها لوظائف متشابهة بعد تحديد نسب التشابه ومقارنتها بالقيم الإحصائية . ولكن هذه الطريقة غير مأمونة الجانب عند مقارنتها باستعمال تواليات DNA ، ففي الحالة الأخيرة توجد 4 قواعد نتروجينية وان التشابه العشوائي يمكن ان يكون 25% ، في حين ان التوزيع العشوائي للحوامض الامينية فأن فرصة الحامض الاميني تكون بحدود اقل من 5% نظرا لوجود 20-22 حامض اميني . وان احتمالية استنتاج توالي البروتينات من توالي DNA بالاعتماد على الشفرات الوراثية لا يكون صائبا نظرا لوفرة وزيادة الشفرات الوراثية المترادفة للحامض الاميني الواحد وبذا فأن عدة تواليات من DNA يمكن ان تعطي توالي بروتين واحد ، وان DNA المتشابه بنسبة 50% يمكن ان يعطي بروتين بنسبة تشابه أعلى . أي ان التشابه على مستوى البروتينات يكون أعلى من تشابه النيوكلوتيديات .

وان حصول إزاحة في موقع واحد من اطر القراءة في DNA يؤدي الى اختلاف كبير في الحوامض الامينية المترجمة منه وبذا فان التشابه على مستوى البروتين يمكن ان يتغير بشكل كبير ، لذا فأن عمليات اصطفاف تواليات مشفرة للبروتين والتي تتم باستعمال تواليات DNA واصطفاف الحوامض الامينية لا تعطي نتائج متطابقة . وبالتالي فأن بروتين واحد يمكن ان ينتج من عدة تواليات للـ DNA وهذا يعني انه عند العمل مع تواليات البروتين فهناك نسبة من المعلومات الموجودة في تواليات DNA او RNA تضيع .

وهذا يظهر جليا في حقيقة ان النيوكلوتيدي في DNA تكون بوضعيتين اما متشابهة او مختلفة ، مقارنة بثلاث حالات تحصل في الحوامض الامينية فهي اما ان تكون مختلفة او متشابهة او متشابهة ولكن غير متماثلة . اي ان بعض الحوامض الامينية تكون متشابهة بالتركيب الكيماوي مثل السيرين والثريونين كل منهم يحوي على مجاميع من الهيدروكسيل (OH-) وكذلك الليوسين والايزوليوسين لهم صفات كيماوية متشابهة والكلوتاميك والاسبارتيك كلاهما حامضي

التفاعل . فعند استبدال الاسبارتات والكلوتامات في إنزيم ما يحتاج الى حامض اميني في الموقع الفعال سوف لا يغير من فعالية الإنزيم بشكل كبير وكامل ، ولكن استبدال الاسبارتات بحامض اميني حلقى مثل التايروسين يمكن ان يؤدي الى تدمير فعالية الإنزيم . وعليه فأن الدرجات عند الاصطفاف للاسبارتات والكلوتامات تكون متشابهة ولكنها مختلفة مع التايروسين .

وعند الأخذ بنظر الاعتبار ان الحوامض الامينية مصنفة الى مجاميع وان أفراد المجموعة او الزمرة تكون متشابهة (الى حد ما) فإنه يجب الأخذ بنظر الاعتبار وجود نوعين من الدرجات المسجلة في عملية الاصطفاف الأولى هي درجات التماثل Identity scores والثانية هي درجات التشابه Similarity scores . وللتشابه درجات تؤخذ بعين الاعتبار مثلاً الحوامض الامينية الانين والايزوليوسين والليوسين والفالين كلهم من الحوامض الامينية الاليفاتية ويمكن ان يحل الواحد محل الآخر في البروتينات العامة مثل البروتينات الكروية بدون اي اختلاف في الشكل النهائي للبروتين ولكنها يمكن ان تختلف في الحجم بحيث يمكن ان تؤثر حالة الاستبدال هذه في المواقع الفعالة للبروتين اذا كان إنزيمياً. والاختلاف يكون كبيراً في حالة الاستبدال بين الحوامض الامينية المشحونة اذ انها تؤثر في القوى العاملة في تثبيت تركيب البروتين . وتكون هذه الحالات موجودة في الحوامض الامينية لذلك يفضل ان تعطى أهمية كبيرة لنسبة التماثل أكثر من نسبة التشابه .

ان تحديد توالي البروتينات وإيجاد بصمة البروتين او الببتيد تتم في بعض الأحيان باستعمال الإنزيمات الهاضمة للبروتينات المتخصصة في مواقع فلقها للبروتين ، او بواسطة الهضم الكيماوي مثل استعمال بروميد السيانيد (CNBr) ، وبعد إعداد التوالي يتم إخضاعه لبرامج المعلوماتية الحيوية لمقارنته مع بروتينات موجودة في قواعد البيانات (المسجلة على انها مهضومة نظرياً) لإيجاد التوائمات او التطابقات المضبوطة او القريبة منها . ومرة أخرى تواجه عملية البحث عن التطابق المنشود ربما تكون بعيدة المنال وذلك لان الهضم بالبروتينات نادراً ما يكون مضبوط بشكل جيد ويؤدي الى إنتاج منتجات مهضومة جزئياً نتيجة لعمليات القطع الخاطئة في مواقع مختلفة والتي ربما تؤدي الى إنتاج قطع مشحونة وتكون هناك زيادة كتلتها او حجمها قليلاً ، ولذلك ولزيادة القابلية التمييزية للبحث في قواعد البيانات فأن الوسيلة المستعملة للبحث يجب ان تسمح بشئ من التفاوت في عملية التطابق للكتل الجزئية للببتيدات في حالة القطع الخاطئة او التحويلات الناتجة عن الشحنات ، ولذا فأن المستخدم يحتاج الى إدخال اكبر قدر ممكن من المعلومات للتواليات البروتينية او الببتيدات مثل الوزن الجزيئي ونقطة التعادل الكهربائي (pI) للبروتين الكامل وربما احتاج الى إدخال اسم الكائن المشتق منه البروتين لغرض التعرف عليه بشكل صحيح ، وبذا كانت بعض عمليات البحث الخاصة عن بروتينات لكائن معين في قاعدة البيانات تتطلب توفر تواليات كل البروتينات الخاصة بالكائن الحي ، وعليه فأن هذه الطريقة تعمل بشكل جيد مع الكائنات التي تمثل نماذجاً دراسية تم تحديد توالياتها وتهتمش جيناتها وبروتيناتها بشكل جيد وتكون قليلة الفائدة مع الأحياء الأخرى التي لا تمثل نماذج دراسية . وفي هذا المجال يمكن التذكير فيما ذكر سابقاً انه من الأفضل التعامل مع قواعد البيانات الخاصة بالكائن في حالة توفرها.

الطرق المعتمدة على التركيب الداخلي : وهذه الطريقة يطلق عليها طريقة البداية Ab initio المذكورة سابقاً وفيها يتم اعتماد المعلومات الخاصة بالبروتين مثل ثبوت البروتين

والهيئات التي يمكن ان تنتج من البروتين ، وتحولات الطاقة الدقيقة للتداخلات الضعيفة التي تحصل في المحاليل المائية . لذلك فهي خليط من الحقائق العلمية والهندسية ، الأولى هو فهم كيفية الوصول الى التركيب الثلاثي للبروتين والجانب الهندسي هو استنتاج التركيب الثلاثي الاتجاه الذي يمكن ان ينتج عن التواليات . وفي هذا التوجه يتم استنباط دالة لتسجيل الدرجات التي تميز بين التركيب الصحيح (الطبيعي او ما يشابهه) وبين التركيب غير الصحيح (غير الطبيعي) . وفي العديد من توجهات طريقة البداية فإن الحقيقتان تجمعان سوية بحيث تؤدي الى استنتاج التراكيب الطبيعية ، وهناك طرق مختلفة من نوع طريقة البداية التي تعتمد على مؤشرات أخرى .

قواعد البيانات والصيغ المستعملة

عند التعامل مع البروتينات والجديدة منها على وجه الخصوص تتم العودة الى قواعد البيانات المتوفرة واختيار قواعد البيانات للبحث يعتمد على المعلومات المخزونة ، لذلك ينبغي على المستخدم ان يكون على إلمام بسيط بها والتي يمكن ان تظهر من الواجهة الخاصة بالقاعدة ، البعض منها يهتم بالتراكيب والأخرى بالوظائف ، والبعض الآخر يجمع الغرضين سوية . ومن أهم المواقع التي تتعامل مع البروتينات او المكونون البروتيني هي ExPASy التابعة للاتحاد الأوروبي ويوفر مزود الخدمة هذا إمكانيات شاملة لدراسة المكونون البروتيني مع برامج ملائمة للبحث عن الببتيدات والبروتينات ويعتمد المزود على قواعد بيانات مهمة مثل قاعدة البيانات السويسرية و TrEMBL . ويضم برامج مختلفة للتعريف بالبروتينات وتشخيصها اعتمادا" على نقاط التعادل الكهربائي (pI) او الوزن الجزيئي ، والمكونات من الحوامض الامينية وغيرها من المعلومات ويقارنها مع القيم النظرية لهذه المكونات مع كل البروتينات الموجودة في قواعد البيانات التي يستند اليها ، ويمكن ان يقلص البحث في هذا المزود باختيار الكلمات المفتاحية الملائمة .

أسس بناء قواعد بيانات البروتينات والصيغ المعتمدة

عند إنشاء قواعد البيانات هذه يؤخذ بنظر الاعتبار العديد من المؤشرات التي مر ذكر معظمها في فصل سابق . وعند إنشاء قواعد خاصة بتراكيب البروتينات يعتمد الى إنشاء جدول فيه إحداثيات X , Y , Z التي تمثل موقع كل ذرة في التركيب وكذلك قياسات كل من الزوايا ϕ و ψ الخاصة بموديل راماجاندران وهذه مودعة في بنك معلومات البروتينات PDB والتي يعد استعمالها شرطاً "مهما" في نشر البحوث في المجالات العلمية ، وهناك قواعد بيانات أخرى خاصة بالبروتينات لكل منها مزاياها التي تفضل من قبل المستخدم وفقاً للغرض المطلوب ، ففي قواعد البيانات هذه روابط ومداخل كثيرة .تساعد في التعرف على الأهمية البايولوجية للبروتينات وأنماطها والنسق الذي توجد فيه والعوائل التي تنتمي اليها بروتينات الاستعلام . وقواعد البيانات هذه يتم تقييمها ومراجعتها بشكل منظم من قبل المؤسسات المعنية . ومن قواعد البيانات هذه على سبيل المثال لا الحصر :

- Protein - information resource - protein sequence database (PIR- PSD) وتكون بمثابة قاعدة بيانات للعوائل البروتينية التي تم تدقيق وتشذيب المعلومات فيها ، لذا تسهل البحث عن العوائل البروتينية .

• UniProt

قاعدة بيانات أنشأت من دمج قواعد أخرى (كما مر في فصل سابق) وتعد الأساس في التعامل مع البروتينات نظرا " لسهولة التعامل معها ولها تطبيقات واسعة . والمعلومات فيها تكون على ثلاث مستويات

*المستوى الأول UniPort Knowledge database (UniProt KB)

ويعد هذا الأساس أو المنطلق للوصول الى البيانات المدققة حول البروتين من حيث الوظيفة والتصنيف وغيرها من النواحي ، وهذه تكون مربوطة بقاعدة البيانات السويسرية و TrEMBL وتختلف كمية المعلومات والتطابقات وفق القاعدة المستعملة وبشكل خاص فأن قاعدة البيانات السويسرية تكون الأفضل لأنها الأوسع .

*المستوى الثاني (UniRef) Reference cluster database

وهذا المستوى يزود بمعلومات حول المجاميع العنقودية للبروتينات الموجودة في المستوى الأول التي لها تماثل في التواليات يتراوح بين 100% ثم 90% ثم 50%.

*المستوى الثالث : (Unipare) Archive database

وهو مستوى من القاعدة يحوي كافة المعلومات العامة المتوفرة حول تواليات البروتينات الموجودة في القاعدة .

• قاعدة (MIPS) Munich information center for protein sequences

هذه القاعدة توفر إمكانيات أكثر لدراسة البروتينات ، اذ يمكن بواسطتها اختيار الكائنات الحية المرتبة وفق النظام الهجائي ، وكل مدخل منها يرتبط بـ NCBI وقواعد تصنيفية . وتوفر القاعدة البروتينات المراد التحري عنها ووصف لأفضل تطابقات BLAST المتوفرة ، وفي حالة الجينومات غير المكتملة تظهر القاعدة المناطق المجاورة . وتكون الجينات المشفرة للبروتينات مفصولة عن تلك الخاصة بـ RNA والتي تكون مفصولة بدورها الى tRNA , rRNA والجزيئات الأخرى فضلا" عن إدراج تراكيب العروة Stem – loop structure لهذه الجزيئات . وتوفر القاعدة بيانات أخرى كثيرة منها إدراج البحوث والنشرات حول الموضوع .

* بنك معلومات البروتينات PDB يعد البنك مركزا " عالميا" للمعلومات التركيبية الخاصة بالجزيئات العملاقة (وهي البروتينات على وجه الخصوص) ويدار من قبل مؤسسة خاصة

(RCSB) Research collaboratory for structural bioinformatics . وموقع البنك على شبكة الانترنت يوفر العديد من الخدمات مثل إمكانية إيداع البروتينات الجديدة وكذلك البحث عن البيانات وإمكانية استردادها . وبذا يمكن للمستخدم ان يحصل على ملخص عن البروتين وتركيبه ، فضلا" عن إمكانية الحصول على صور للبروتين باستعمال الروابط الخاصة ببرامج الرسوم مثل RasMol وغيرها بدرجات وضوح مختلفة وتوفر مخرجات الموقع الناتجة الحصول على إحداثيات الذرات في البروتين ، والبحوث الخاصة بالبروتين .

والجدير بالذكر ان أكثر مداخل القاعدة مخصصة للبروتينات التي تم دراستها باستعمال الأشعة السينية والقليل منها يعتمد على الرنين النووي المغناطيسي

كل بروتين له مدخل خاص PDB Id يتكون من رموز أما حرفية (Z – A) يتبعه رقم (9 – 0) مثل ILYZ , 4RCR ويمكن البحث في هذه البنك او القاعدة باستعمال الرقم

التعريفى او استعمال كلمات مفتاحية ذات علاقة بتهميش البروتين . ومن الجدير بالذكر ان بعض البرامج الخاصة بدراسة البروتينات التابعة للبنك يكون أساسها الرقم التعريفى لانجازها المهام المنوطة بها .

وصيغ المخرجات للبنك فضلا" عما ذكر أعلاه من توفير إحداثيات الذرات فهي توفر المراجع العلمية الخاصة بالبروتين ، وطرق تحديد تركيبه والمؤشرات البلورية وتوالي البروتين والعوامل المرافقة والتي تقع ضمن HETAM أي المجاميع المتباينة Heteroatom group وتشمل فضلا" عن العوامل المساعدة مواد الأساس ان كان إنزيما" ، ووصف لأنواع التركيب الذي يمكن ان يكونه فضلا" عن وصف موقع البروتين في الخلية .

ولكن ما يؤثر بشكل سلبي في استعمال قاعدة البنك هو عدم سعتها لاستيعاب عدد كبير من التواليات (مثلا" أكثر من 26 توالي) وهذا يعني ان المعقدات البروتينية الكبيرة مثل الريبوزوم لا يمكن ان تمثل بملف PDB واحد وانما يجب ان تقسم الى عدة ملفات . وهذه التقديرات أدت الى تطوير صيغ جديدة للتعامل مع الأعداد المتزايدة والمعلومات الهائلة حول تركيب البروتينات وتظهر مخرجاتها بشكل ASN. file وغيرها وتحتوي على معلومات إضافية تساعد في رسم شكل البروتين .

تصنيف البروتينات والمجاميع الفعالة

يعد تصنيف البروتينات من ضمن المؤشرات المهمة في المعلوماتية الحيوية وبرامجها . فالبروتينات تصنف بعض الأحيان اعتمادا" على التشابه في القطيفات او الدومينات فقط وتودع في قواعد البيانات مثل قاعدة بيانات عوائل البروتينات Pfam و PRINTS ، وهناك بعض القواعد تشمل النوعين من العوائل (أي المعتمدة على القطيفات والدومينات) . وفي جميع الأحوال فان علاقة البروتين مع العائلة هي ليست علاقة واحد الى واحد فهي يمكن ان تكون واحد الى عدة عوائل وهذه تودع في بعض قواعد البيانات .

والحقيقة ان عملية تصنيف البروتينات الى عوائل كبيرة (Superfamily) وهي مجموعة من البروتينات التي لها تماثل جذري) تشمل عوائل معينة (Family) التي تمثل مجموعة من البروتينات التي تكون علاقتها مع بعضها معتمدة على تماثل التواليات لتكون مجموعة خاصة (ضمن العائلة الكبيرة) يساعد في تحديد وظائف البروتينات غير المعروفة والتي يطلق عليها البروتينات الافتراضية . وعلى العموم فان تحديد وظيفة البروتين يحتم توفر نتائج اصطفاة شامل للعوائل وكذلك اصطفاة موضعي للقطيفات والدومينات التي تحويها . وفي مجال التقسيمات الحديثة التي تلازم التعامل معها في مجالات المعلوماتية الحيوية تقسم البروتينات الى مجاميع واسعة بالاعتماد على الجينات المتناظرة في الأحياء وأهمها (COGs) والذي له جزء مخصص في الموقع الأمريكي NCBI. ويعتمد التصنيف على قيم E- Values بعد إجراء عمليات الاصطفاة باستعمال BLAST في عملية العنقدة المعتمدة على درجات إحصائية ، والعوائل الناتجة من هذا الطريق لا تعتمد على وجود توالي دليل خاص Sequence signature لذلك تكون أكثر شمولية ، ولكن المعوقات في هذه الطريقة هو وجود الكواذب في تصنيف البروتينات . وتوجد قاعدة خاصة بـ COG التي تحوي عوائل من البروتينات على أساس العلاقات التطورية فمثلا" تقسم البروتينات البكتيرية الى أكثر من 23 مجموعة فعالة خاصة تتوزع على ثلاث أصناف رئيسة كما موضح في الجدول 7.5 :

جدول 7.5 : عوائل ومجاميع البروتينات والجينات المتناظرة COGs

الصفة	الدومينات	COGs	شفرات القاعده
معلومات حول الخزن والمعالجة			
J	10,572	245	Translation, ribosomal structure and biogenesis
A	137	25	RNA processing and modification
K	11,271	231	Transcription
L	10,338	238	Replication, recombination and repair
B	228	19	Chromatin structure and dynamics
المعلومات الخلوية ونقل الاشارات			
D	1,678	72	Cell cycle control, cell division, chromosome partitioning
Y	-	-	Nuclear structure
V	2,380	46	Defense mechanisms
T	7,683	152	Signal transduction mechanisms
M	7,853	188	Cell wall/membrane/envelope biogenesis
N	2,747	96	Cell motility
Z	128	12	Cytoskeleton
W	25	1	Extra-cellular structures
U	3,743	159	Intracellular trafficking, secretion, vesicular transport
O	6,206	203	Post-translational modification, protein turnover, chaperones
الايض			
C	9,830	258	Energy production and conversion
G	10,816	230	Carbohydrate transport and metabolism
E	14,939	270	Amino acid transport and metabolism
F	3,922	95	Nucleotide transport and metabolism
H	6,582	179	Coenzyme transport and metabolism
I	5,201	94	Lipid transport and metabolism
P	9,232	212	Inorganic ion transport and metabolism
Q	4,055	88	Secondary metabolites biosynth., transport and catabolism
مجاميع غير موصوفة بشكل جيد			
R	22,721	702	General function prediction only
S	13,883	1346	Function unknown

وهذه المجاميع هي المستعملة لتصنيف كل الجينات المهمشة في الجينوم للتعرف على الوظيفة ، فالبروتينات التي لا تظهر تشابهاً واضحاً مع الموجود في قواعد NCBI توضع في المجموعة R و S أي تكون غير موصوفة بشكل جيد وهي التي تهتمش على أنها بروتينات افتراضية .

طوي البروتينات

عمليات طوي البروتينات وفك طويها تعد من الأحداث المهمة في الخلية (الفصل الثالث) ويكون من الصعوبة وصفها بالتفصيل خارج الأنظمة الحية ، وللعملية تأثير كبير في وظيفة البروتين فمثلاً البريونات التي تعمل كعوامل إصابة بروتينية تنتج من طوي البروتينات بشكل غير صحيح والتي تؤدي إلى الأمراض المعروفة مثل جنون البقر وغيرها ، ولكن هذه البروتينات عندما تكون مطوية بشكل صحيح تكون جزءاً من مكونات الخلية . كما ان بعض البروتينات تحتاج ان تخرج من الخلية لتؤدي وظيفتها وهذه تحوي في اغلب الاحيان على انها اشارات خاصة (كما سيأتي ذكره لاحقاً) والتي عند قطعها يمكن للبروتين ان يطوى بالشكل الصحيح .

وقد ظهرت وسائل ضمن مجال المعلوماتية الحيوية لتحاكي طوي البروتينات وأحياناً أخرى صممت تجارب باستعمال طفرات مهندسة وراثياً لتحليل قيم زاوية ϕ التي تحدد التراكيب الوسطية والحالات الانتقالية في مسارات الطوي وفك الطوي على مستوى الذرات . ولكن من الجدير بالذكر انه يصعب في بعض الحالات تصوير طوي البروتينات باستعمال وسائل المعلوماتية الحيوية خاصة عند إضافة السكريات ، كما ان هناك العديد من البروتينات التي تساهم في إعطاء البروتين الشكل النهائي ، لذلك كانت عمليات حدس طوي البروتينات والتعامل معها من التطلعات المستقبلية في مجال المعلوماتية الحيوية . وهناك مواقع خاصة بدراسة عمليات طوي البروتينات وتوضح طرق الطوي المختلفة الطبيعية وغير الطبيعية التي تؤدي إلى العديد من الأمراض ومنها Folding @Home .

التنبؤ بوظائف البروتينات

تحديد وظائف البروتينات بالتجربة يكون متخلفاً مقابل التطورات الكبيرة المستعملة في الطرق السريعة . وقد ذكرت بعض جوانب الموضوع آنفاً . فعند عدم إمكانية الكشف عن وظيفة البروتين بالطرق التجريبية يكون من الأجدي البحث عنها بإيجاد التشابه مع البروتينات المهمشة والمعلنة وظائفها في قواعد البيانات ولكن يجب توخي الحذر في هذه الحالة وذلك لان هناك عدداً من إعلانات وتهميشات الجينومات هي خاطئة ، وخطأ بعضها تضخم بإعداد قواعد البيانات المعتمدة عليها . وهناك عدة أسباب لهذه الأخطاء ، منها ان بعض البروتينات تقوم بأكثر من وظيفة ولذلك فيإيجاد وظيفة واحدة سيكون خاطئاً . ولكن يمكن إعلان الوظيفة عن طريق إجراء المقارنة واعتماد التشابه فيما اذا تم استعمال برامج مختلفة وقواعد بيانات مختلفة فضلاً عن إجراء التدقيق والتشذيب العملي واعتماد المصفوفات الملائمة الخاصة بتركيب وترتيب الحوامض الامينية في البروتينات مثل استخدام مصفوفات Dayhoff (PAM matrices).

وإضافة إلى ذلك فإن الوصول إلى هدف تحديد الوظيفة يمكن ان يتم بمساعدة معرفة موقع البروتين في الخلية والتحويلات التي تجري عليه بعد الترجمة فمثلاً هذه المعلومات تساعد في تحديد الوظائف العامة أو المجموعة التي تمتاز بوظيفة عامة بالرغم من عدم وجود تطابق

وتشابه عالي بين توالي الاستعلام والتواليات الموجودة في قواعد البيانات ، وهناك العديد من الوسائل التي تستعمل للتنبؤ بهذه الجوانب والأفضل البحث عنها في المراجع العلمية الخاصة بها . وإضافة الى ما ذكر أعلاه فأن هناك وسائل او برامج للتنبؤ بالتغيير الحاصل في وظائف البروتينات عند استبدال احد الحوامض الامينية AAS تكون مهمة في حالة البحث عن تغييرات النيوكليوتيد المفرد (SNP) التي تؤدي الى تغيير البروتينات ومن ورائها تغيير النمط المظهري ، او تحديد الطفرات غير المحسوسة (Miss sense) .

البرامج المستعملة في اصطفاف البروتينات

من المهام التي تقوم بها المعلوماتية الحيوية هو إجراء اصطفاف تواليات متعددة للجزيئات الحيوية ومنها البروتينات والكشف عن الثمالات الثابتة ونمطها لمساعدة الباحث على استنتاج تركيب ووظيفة البروتينات او عائلة البروتينات . ومن البرامج المهمة في هذا المجال هي مجموعة Clustal بتغييراتها وإصداراتها المختلفة وكذلك BLASTP . ونتائج مثل هذه البرامج تكون ملائمة ما لم تتخفف نسبة التشابه الى منطقة الغبش التي تكون تحت 30% ، وعندها تكون نتائج عمليات الاصطفاف ضعيفة ، ولكن هناك بعض الجوانب الأخرى التي تساعد في تقوية نتائج الاصطفاف ، ومن هذه الجوانب او الخيارات هو التركيب الثانوي للبروتينات الذي كان الأساس في وضع البرامج وتطوير الموجود منها . وقد وجد ان البرامج التي اعتمدت التركيب الثانوي قد وصلت دقتها الى 77% في حالة حدس التركيب الثلاثي ، اذ تعد الأخيرة مقياساً لكيفية التنبؤ بالحلزونات والصفائح والالتفافات التي يعبر عنها كنسبة مئوية للتركيب الثانوي . ولذلك فأن كان من الممكن الجمع بين توالي البروتين (التركيب الأولي) والتركيب الثانوي عندها يمكن التغلب على الشذوذ الصادر عن بعض البرامج وذلك بالتغطية من مصادر المعلومات من الجهة الثانية .

طرق التنبؤ بتركيب البروتينات

البروتينات المتوفرة في بنك بيانات البروتينات PDB تم تحديد تركيب أكثر من 90% منها باستعمال الأشعة السينية وذلك لان الطريقة تسمح بقياس توزيع كثافة الالكترونات في الأبعاد الثلاثية في البروتين في حالته البلورية وبالتالي تمكن من استنتاج الإحداثيات الثلاثية 3D- Coordinates لكل الذرات التي يجب ان تحدد بدرجة عالية من الدقة . أما الباقي من تراكيب البروتينات المودعة في البنك فقد تم تحديدها بواسطة الرنين النووي المغناطيسي الذي يستعمل لتحديد التراكيب الثانوية .

والمعروف ان معرفة تركيب البروتين يعطى وضوحاً أكثر لوظيفة البروتين أكثر مما يعطيه معرفة تواليه . وبصورة عامة فأن للبروتينات أربع مستويات من التراكيب كما مر ذكره في فصل سابق . وحقيقة الأمر ان هناك أكثر من طريق للتنبؤ بتركيب البروتينات والمذكورة آنفاً (ضمن التوجهات المستعملة) ومنها طرق المقارنة المعتمدة على تشابه البروتينات والتي تقاس بالنسبة المئوية للثمالات المتماثلة في كل موقع . والتشابه بين البروتينات المتقاربة يكون عالياً جداً في المناطق المركزية التي تتكون من عناصر التركيب الثانوي مثل α - helixes , β - sheets ومناطق العروات مثل استدارات البيتا β - turns التي تربط هذه التراكيب الثانوية . ويتم بداية بناء نموذج المقارنة وذلك بإجراء عمليات اصطفاف بين التواليات لكل بروتين تم تحديده بالطرق التجريبية ويسمى هذا البروتين الأب مع بروتين الاستعلام وهذه

تؤدي الى بناء النموذج الأولي والذي يستعمل في الخطوات اللاحقة ، وبذلك فأن هذه تقع ضمن توجه طرق البداية المعتمدة على التوالي .

أما البروتينات التي لا تخضع للطرق أعلاه فيتم تحديد تراكيبها بطريقة قياس اللياقة Threading وهي قياس لياقة البروتين وثبوته وتستعمل لذلك عمليات مقارنة مع قواعد بيانات البروتينات المعروفة التركيب الثلاثي وملاحظة موافقتها مع البروتينات غير المعروفة التركيب او الطوي ، وتحدد الموائمة بتسجيل الدرجات ، أي ان الطريقة تعتمد الى مقارنة التوالي الاستعلام مقابل مكتبات تحوي القوالب التركيبية لإعطائه قائمة من الدرجات ثم يتم ترتيب البروتينات لأفضل الدرجات . أما الطريقة الأخرى فهي Homology modeling تستعمل لبناء نموذج ثلاثي الأبعاد للبروتين غير معروف التركيب وذلك بالاستعانة بواحد او أكثر من البروتينات معروفة التركيب .

والطرق العامة المذكورة أعلاه تستعمل في تحديد مستويات التركيب للبروتينات وذلك لان حدس تركيب البروتينات يؤدي عدة أغراض ولها تطبيقات مهمة منها :

- تصنيف البروتينات
- فصل دومينات البروتين والقطيقات الوظيفية
- يوفر البديل النظري عن العمل التجريبي عندما تكون الطرق التجريبية غير ناجحة .

مستويات تركيب البروتينات والمؤشرات المستعملة

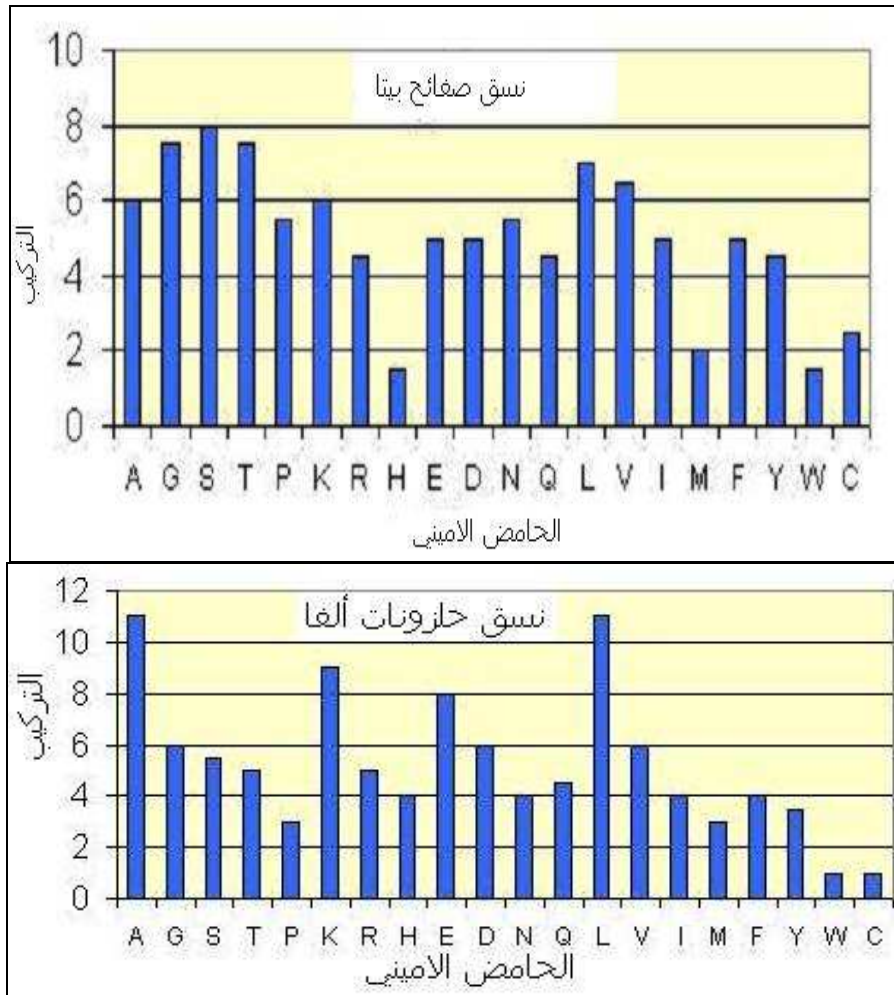
للبروتينات أربعة مستويات من التراكيب ، التركيب الأولي ويتمثل بتوالي الأحماض الامينية فيه ، والتركيب الثانوي ويشمل تكوين حلزونات ألفا وصفائح بيتا ، ثم ينتج التركيب الثلاثي من تجميع عناصر التركيب الثانوي ، أما التركيب الرباعي فيتمثل بتجميع الوحدات (او الوحدات الفرعية) الناتجة عن التركيب الثلاثي لتكوين البروتينات الكاملة وفي هذه الحالة تكون الإنزيمات أحسن الأمثلة لها . وملخص هذه التراكيب موضح في الشكل (78.3) الفصل الثالث.

وللحوامض الامينية بعض الميل او التفضيل في ان تكون ضمن حلزونات ألفا او في اشرطة بتا . كما موضح في الجدول (8.5) وهي مكملة للبيانات الموضحة في الجدول (9.3) الفصل الثالث) ولكن وفق قياسات مختلفة بعض الشيء .

جدول 8.5 : تفضيل الحوامض الامينية العام للتراكيب الثانوية

α -helix	β -strand	الحامض الاميني
1.59	0.52	Glu
1.41	0.72	Ala
1.34	1.22	Leu
1.30	1.14	Met
1.27	0.98	Gln
1.23	0.69	Lys
1.21	0.84	Arg
1.05	0.80	His
0.90	1.87	Val
1.09	1.67	Ile
0.74	1.45	Tyr
0.66	1.40	Cys
1.02	1.35	Trp
1.16	1.33	Phe
0.76	1.17	Thr
0.43	0.58	Gly
0.76	0.48	Asn
0.34	0.31	Pro
0.57	0.96	Ser
0.99	0.39	Asp

وقد حسب نسق احتمالات وجود كل حامض في عدد كبير من البروتينات ويمكن ملاحظة ذلك في الشكل 115.5 :



شكل 115.5 : نسق وجود الحوامض الامينية في صفائح بيتا وحلزونيات ألفا

و"فضلا" عن هذه المؤشرات هناك صفات أخرى يمكن ان تدخل في البرامج المستعملة لحدس التركيب المختلفة وهي موضح في الجدول 9.5 :

جدول 9.5 : مجمل صفات الحوامض الامينية المستعملة في وضع برامج حدس البروتينات

Amino Acid			Molecular Formula	Hydropathy Index	log P	Acidic/Neutral/Basic	Polar/Hydrophobic	pI
Arg	R	Arginine	$\text{H}_2\text{N}(=\text{NH})\text{NHC}_3\text{H}_6\text{CCH}(\text{NH}_2)\text{COOH}$	-4.5	-4.20	Acidic	Polar (Basic)	10.76
	K	Lysine	$\text{H}_2\text{N}(\text{CH}_2)_4\text{CH}(\text{NH}_2)\text{COOH}$	-3.9	-3.05	Basic	Polar (Basic)	9.74
Asp	D	Aspartic Acid	$\text{HOOCCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-3.5	-3.89	Acidic	Polar (Acidic)	2.77
Asn	N	Asparagine	$\text{H}_2\text{NCOCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-3.5	-3.82	Neutral	Polar (Neutral)	5.41
Glu	E	Glutamic Acid	$\text{HOOCCH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-3.5	-3.69	Acidic	Polar (Acidic)	3.22
Gln	Q	Glutamine	$\text{H}_2\text{NCO-CH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-3.5	-3.64	Neutral	Polar (Neutral)	5.65
His	H	Histidine	$\text{H}_2\text{NCH}[\text{CH}_2(\text{C}_3\text{H}_3\text{N}_2)]\text{COOH}$	-3.2	-3.32	Basic	Polar (Basic)	7.59
Pro	P	Proline	$(\text{C}_4\text{H}_8\text{N})\text{-COOH}$	-1.6	-2.54	Neutral	Nonpolar (Hydrophobic)	6.30
Tyr	Y	Tyrosine	$\text{HO}(\text{C}_6\text{H}_4)\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-1.3	-2.26	Neutral	Polar (Neutral)	5.66
Trp	W	Tryptophan	$(\text{C}_8\text{H}_6\text{N})\text{-CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-0.9	-1.05	Neutral	Nonpolar (Hydrophobic)	5.89
Ser	S	Serine	$\text{HOCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	-0.8	-3.07	Neutral	Polar (Neutral)	5.68
Thr	T	Threonine	$\text{CH}_3\text{CH}(\text{OH})\text{CH}(\text{NH}_2)\text{COOH}$	-0.7	-2.94	Neutral	Polar (Neutral)	6.16
Gly	G	Glycine	$\text{H}_2\text{NCH}_2\text{COOH}$	-0.4	-3.21	Neutral	Nonpolar (Hydrophobic)	5.97
Ala	A	Alanine	$\text{CH}_3\text{CH}(\text{NH}_2)\text{COOH}$	1.8	-2.85	Neutral	Nonpolar (Hydrophobic)	6.00
Met	M	Methionine	$\text{CH}_3\text{SCH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	1.9	-1.87	Neutral	Nonpolar (Hydrophobic)	5.74
Cys	C	Cysteine	$\text{HSCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	2.5	-2.49	Neutral	Polar (Neutral)	5.07
Phe	F	Phenylalanine	$(\text{C}_6\text{H}_5)\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	2.8	-1.38	Neutral	Nonpolar (Hydrophobic)	5.48
Leu	L	Leucine	$(\text{CH}_3)_2\text{CHCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	3.8	-1.5	Neutral	Nonpolar (Hydrophobic)	5.98

					2			
Val	V	Valine	$(CH_3)_2CHCH(NH_2)COOH$	4.2	- 2.2 6	Neutral	Nonpolar (Hydrophobic)	5.96
Ile	I	Isoleucine	$C_2H_5CH(CH_3)CH(NH_2)COOH$	4.5	- 1.7 0	Neutral	Nonpolar (Hydrophobic)	6.02

حدس التركيب الثانوي للبروتينات

تحديد التركيب الثانوي للبروتينات يهدف الى جسر الهوة بين التوالي في التركيب الأولي وباقي مستويات تركيب البروتينات ، والتركيب الثانوي للبروتين يكون ثابت موضعياً للبتيدات المتعددة المهمة في المحافظة على التركيب ثلاثي الأبعاد ويبدأ تكوين التركيب الثانوي بعد نمو التوالي أثناء عملية الترجمة وذلك لان التركيب الثانوي يمثل الحالة الأفضل من ناحية صرف الطاقة . وتتكون حلزونات ألفا وصفائح بيتا التي يتم التحكم بها بالصفات الفيزيائية للحوامض الامينية وكذلك ترتيبها في التوالي وصفات البيئة المحيطة ، وهناك بعض الحقائق التي لا بد من ذكرها لانها تؤثر في التركيب الثانوي وطرق تحديده ، منها ان الحوامض الامينية الليوسين والالنين وحمض الكلوتاميك توجد عادة في حلزونات ألفا ، في حين ان الفالين والايزوليوسين والفنيل - النين تكون شائعة في صفائح بتا (كما موضح بعضها في الجدول 8.5) ، أما البرولين فلا يمكن ان يوجد في وسط حلزونات ألفا لانه يؤدي الى كسر الحلزونات ، ولكن يمكن ان يوجد في نهاية التركيب الحلزوني ، والملاحظ ان الحوامض الامينية المشحونة تستبعد من صفائح بيتا ، والأخيرة يمكن ان توجد بشكل متوازي او غير متوازي اعتماداً على الاتجاه النسبي للأواصر الببتيدية التي تربط الحوامض الامينية وعند السطح تظهر تناوباً بين الثمالات المحبة للماء والأخرى الكارهة للماء ، أما أشرطتها المدفونة او الداخلية فتكون في الغالب من الثمالات الكارهة للماء . والتراكيب البروتينية تثبت بالأواصر الهيدروجينية والقوى الأيونية بين الحوامض الامينية المشحونة وكذلك تشارك الأواصر التساهمية ثنائية الكبريتيد . وحلزونات ألفا تضم 3.6 حامض أميني ألفة والتركيب يثبت بالأواصر الهيدروجينية بين ثمانية معينة والموقع 4 + الثمالة معينة . (وقد مر مناقشة بعض هذه الجوانب في الفصل الثالث)

وباعتماد الصفات المذكورة آنفاً ظهرت طرق كثيرة لتحديد التركيب الثانوي منها استعمال الرنين النووي المغناطيسي وكذلك طريقة Circular dichroism وطريقة الاستقطاب المزدوج Dual polarization واستعمال مقياس التداخل Interferometry وطريقة المجهر الالكتروني Cryo-electron microscopy التي توفر وضوحاً عالياً يصل الى اقل من 5 انكستروم ويمكن ان تتوزع هذه التقنيات على طرق متفرقة لحدس التراكيب الثانوية .

والتنبؤ بالتركيب الثانوي يعتمد الى حدس حالة التوزيع المجسم لكل من مكونات البروتين مثل الحلزونات (H) والأشرطة (E) واللفات (C) ، وبالاعتماد على ان التراكيب الثانوية التي لها ترتيب معين من الحوامض الامينية يثبت بنمط خاص من الأواصر الهيدروجينية وهذه النمطية تساعد في وضع الخوارزميات المستعملة في الحدس . وتحتاج الى معرفة الثمالات الموجودة على السطح وكذلك الثمالات الداخلية ، فضلاً عن معرفة الطاقة الحرة اللازمة لنقل الجزيئات

من الطور الزيتي الى الطور المائي وبالعكس وتحديد المساحة السطحية المتاحة للبروتين وكل هذه لها برامج خاصة تحددتها.

والملاحظ ان التركيب الثانوي أكثر ثبوتا من التواليات أثناء عمليات التطور لذلك فإن تحديد مكونات التركيب الثانوي بشكل صحيح تساعد في صف التواليات او تحسين عملية الاصطفاف للتواليات ذات العلاقة البعيدة (كما ذكر سابقا) ، فضلا عن ان تحديد التركيب الثانوي يعد خطوة وسطية لتحديد التركيب الثلاثي .

ونظرا للاختلاف الكبير بين البروتينات الخيطية او الليفية والكروية لذلك هناك توجهات مختلفة في البرامج الخاصة بتحديد التركيب الثانوي . ومن هذه التوجهات ، استعمال طرق البداية والتي تعتمد الى حدس التركيب الثانوي اعتمادا على توالي الاستعلام ، وفيها يتم قياس ميل كل حامض أميني يعود الى تركيب ثانوي خاص ، ودرجة النزوع او الميل تشتق من التراكيب البلورية للبروتينات المعروفة والطرق المعتمدة على هذه وضعت في السبعينات عندما كانت المعلومات حول تراكيب البروتينات نادرة .

والطرق تحدد مدى ميل كل ثمالة ان تكون ضمن الحلزون او الشريط او استدارات بتا - β turn ، اذ تميل بعض الحوامض - كما ذكر أعلاه - ان تكون ضمن حلزونات ألفا التي يندر وجود البرولين والكلايسين فيها. والجدول التالي (10.5) يمثل القيم لكل حامض أميني والمستعملة في بعض طرق حدس التراكيب الثانوية التي تستعمل نوافذ مختلفة الحجم لكل من حلزونات ألفا وصفائح بيتا .

جدول 10.5 : ميل الحوامض الامينية للدخول في التراكيب الثانوية المختلفة وفق طريقة Chou-Fasman

طريقة Chou-Fasman			
P (Turn)	$(\alpha$ -Helix)	P (β -Strand)	Amino Acid
0.66	1.42	0.83	Alanine
0.95	0.98	0.93	Arginine
1.56	0.67	0.89	Asparagine
1.46	1.01	0.54	Aspartic acid
1.19	0.70	1.19	Cysteine
0.74	1.51	0.37	Glutamic acid
0.98	1.11	1.11	Glutamine
1.56	0.57	0.75	Glycine
0.95	1.00	0.87	Histidine
0.47	1.08	1.60	Isoleucine
0.59	1.21	1.30	Leucine
1.01	1.14	0.74	Lysine
0.60	1.45	1.05	Methionine
0.60	1.13	1.38	Phenylalanine
1.52	0.57	0.55	Proline
1.43	0.77	0.75	Serine
0.96	0.83	1.19	Threonine
0.96	0.83	1.19	Tryptophan
1.14	0.69	1.47	Tyrosine
0.50	1.06	1.70	Valine

وتختلف الطرق في دقة حدسها للتراكيب الثانوية وكذلك تختلف من جيل الى جيل وكانت التحسينات التي أجريت على الطرق القديمة هو الأخذ بنظر الاعتبار العلاقات الخاصة بنظريات طوي البروتينات والتداخلات الحاصلة بين البروتينات ، وبذلك أعيدت الإحصائيات الخاصة بكل ثمالة اعتمادا" على عدد اكبر من البروتينات المحددة التركيب فضلا" عن إدخال اكبر للتداخلات الممكنة للثمالات (جدول 8.5) . ولهذا ظهرت أجيال جديدة من الطرق التي طورت في الثمانيات وبداية التسعينات من القرن المنصرم وقد تحسنت الدقة بنسبة 10% عن سابقتها المعتمدة على الحدس العشوائي .

وإضافة الى الطرق المذكورة هناك طرق تعتمد على التماثل الجذري للبروتينات . ولكن في حقيقة الأمر ان الطرق لا تزال غير قادرة على الأداء بشكل روتيني وان الخطر يحدث من فقدان أشرطة بيتا β - strands والتركيب الثانوية القصيرة لكل من الحلزونات والأشرطة ولحد الآن لا تزال عمليات حدس أشرطة بيتا عشوائيا" نوعا" ما .

دقة حدس التراكيب الثانوية

تحديد التركيب الثانوي يعني تميز مناطق الطوي ولذلك فان النوعية Quality هي العامل المهم في النجاح او الفشل في تميز مناطق الطوي ، وبالمقابل فان زيادة الدقة لتحديد التركيب الثانوي تزيد في دقة عمليات الاصطفاف عندما تتعثر الأخيرة . وتقل دقة تحديد التراكيب

الثانوية في حالة الاعتماد على قواعد بيانات لتراكيب صغيرة وتتنخفض دقتها الى 50% وهذه غير ملائمة وذلك لان الحدس العشوائي يمكن ان تصل دقته الى 40% في البروتينات الكروية ويكون التوزيع فيه 30% للحلزونات ، 20% لأشرطة بيتا و 50% للفتات ، ودقة الحدس تقاس بما يسمى Q_3 score أي الاعتماد على الحالات الثلاثة وهي : الحلزونات (H) ، والأشرطة (E) والفتات (C) ، والدرجة هي نسبة مئوية للثلاثيات في البروتين التي تكون قد حددت بدقة ، وهي عادة تستخلص من معدل النتائج التي يتم الحصول عليها من فحص عدد من البروتينات المعروفة التراكيب . وتتراوح دقة عدد من البرامج المستعملة (Q_3) بين 74.3 – 79% ، والملاحظ ان أعلى نسبة هي 79% ، والأخطاء الشائعة التي تخفض من دقة الحدس هي :

- الالتباس الحاصل بين الحلزونات والأشرطة
- عدم الدقة في تحديد بداية ونهاية المواقع الصحيحة للحلزونات والأشرطة
- عدم الدقة والخطأ في تحديد عناصر التركيب الثانوي .

وفي حالة تسجيل أعلى نسبة من الدقة المسجلة فهذا يعني ان 21% من الثمالات تحس بشكل خاطئ ، ولتجاوز مثل هذه الحالات يوصى بعدة طرق جيدة اذ ان هنالك بعض البرامج تكون مفيدة لأداء هذه المهمة وتتفق في الخطوط العريضة ومنها انها تستعمل توالي الاستعلام للبحث في قواعد البيانات باستعمال PSI-BLAST لثلاث مرات ، ثم يتم طرح التطابقات المتكررة Redundant sequence hits والنتائج يستخلص معه النسق الذي يخضع لستة برامج مختلفة أخرى (يراجع الفصل الرابع) ، وإذا كان هنالك تطابقا كبيرا بين نتائج البرامج الستة فيعلن على ان التراكيب هي التراكيب الثانوية المطلوبة ، ثم استنتاج التوافق والاعتماد على أكثر من قاعدة من قواعد البيانات العامة ، وفي هذه الحالة يمكن الوصول الى 80% من الدقة وهذه النسبة تكون مهمة جدا لأنها تكون مكافئة للوضوح الذي يمكن الحصول عليه من التجارب المختبرية لتحديد التراكيب الثانوية . ولتحقيق الهدف من استعمال عدة طرق لحدس التراكيب الثانوية بدقة عالية ظهرت بعض مزودات الخدمة التي تقوم باستلام توالي الاستعلام وإرساله الى مزودات خدمة أخرى لتجرى عليه الدراسات ، ولعل مثل هذه العملية تأخذ وقت طويل لذلك ترسل النتائج بالبريد الالكتروني ، والحقيقة ان بعض مفردات الخدمة او المواقع ترسل الاستنتاجات ولكن البعض الآخر تترك للمستخدم إجراء الاستنتاج والتوافق الذي يحتاجه ، وقد ساعدت مزودات الخدمة هذه والبرامج العاملة فيها الى تحسين النتائج في هذا المجال .

حدس التركيب الثلاثي للبروتينات

تحديد التركيب الثلاثي المجسم للبروتينات يكون ضروريا لفهم الوظيفة ، ولا تزال هناك العديد من البروتينات التي تتوفر لها المعلومات على مستوى توالي الثمالات ولكن تراكيبها ثلاثية الأبعاد غير معروفة ، فمثل هذه البروتينات لا تزال وظائفها غير معروفة وكذلك إمكانية ربطها الى الربائط ، ولا تعرف الآليات التي تقوم بها في المساعدة في التفاعلات الحيوية في حالة كونها إنزيمات ، وعليه كان من الضروري الوصول الى تركيب البروتين (وان كان بشكل تقريبي) باستعمال الحاسوب وهذا سيتم تناوله في هذا الجزء من الكتاب .

التركيب الثلاثي يصف ترتيب حلزونات ألفا وصفائح بيتا وغيرها من التراكيب في الفراغ وقد تم تطوير وسائل لحدس هذه التوزيعات الهندسية التي يتم مقارنة نتائجها مع المعلومات الناتجة

من الدراسات البلورية لتحديد مدى صلاحية الحس بالحاسوب . وبما ان الدراسات البلورية تكون ملائمة بشكل جيد مع البروتينات النقية تماما" والذائبة في الماء او التي يحدث تغيير في توزيعها المجسم عند سحب الدهون منها وهذه كلها ممثلة في بنك بيانات البروتين PDB . أما البروتينات غير معروفة التركيب فيمكن ان يتم التنبؤ بها باستعمال نماذج كوبنهايكن Copenhagen models المتوفرة في موقع القاعدة الدانمركية CBS . وهناك عدة طرق لتحديد التراكيب الثلاثية منها الفيزيائية والاخرى المعتمدة على برامج الحاسوب منها

• **الطرق الفيزيائية :** مثل استعمال الأشعة السينية والرنين النووي المغناطيسي وهذه الطرق تكون ذات معدلات سرعة واطئة مما أدى الى زيادة الهوة في السرعة بين المعلومات حول توالي البروتينات والمعلومات حول تركيب البروتينات . فضلا" عن ان بعض البروتينات مثل بروتينات الأغشية يكون من الصعب تحديدها بهذه الطرق .

• **الطرق التجريبية** وهي طرق مسرفة في الوقت ومحددة الإمكانيات .

• **طرق الحاسوب:** الطرق التي يهتم بها مجال المعلوماتية الحيوية ويمكن ان توفر نتائج ملائمة لتكون انطلقا" لإجراء التجارب الحيوية مثل إجراء التطوير الموضعي الموجه ، ودراسة ثبوت البروتين وتحليل وظائفه ، فضلا" عن استخدامها كدليل نظري لتصميم التجارب الخاصة بالبروتينات وتوصيفها ، وهذه كلها تصب في النهاية الى معرفة التركيب الثلاثي للبروتينات في مجال المعلوماتية الحيوية وهي :

• **دراسة التشابه** وإيجاد النماذج الخاصة به Homology modeling .

• **الحس باستعمال قياس اللياقة للبروتينات الموجودة** Threading prediction

• **الطرق المعتمدة على مواصفات التواليات الأصلية** (استعمال طرق البداية)

طريق التشابه Homology modeling : ويشير اسم الطريقة الى ان عملية الحس او التنبؤ تعتمد على تماثل التواليات للتركيب المعروفة وتعرف أيضا" بطريقة النمذجة المقارنة ، وتعتمد الطريقة على توفر المعلومات عن التراكيب الموجودة في قواعد البيانات . فبدية يتم إنشاء موديل للذرات اعتمادا" على التراكيب المحددة بالتجارب التي تكون ذات علاقة وثيقة بالتوالي . والافتراض وراء هذه الطريقة هو ان البروتينان اللذان يشتركان بنسبة عالية من تماثل التوالي تشير الى احتمال كبير ان يكون لهما تركيب ثلاثي الأبعاد متشابه بشكل كبير ، واذا كان احدهما معروف التركيب فأن هذا سينسحب على البروتين الثاني وبدرجة عالية من الثقة . والطريقة تتم بعدة خطوات للوصول الى الاستنتاج الأخير ومن هذه الخطوات :

• **اختيار القالب المحدد** التوالي من قواعد البيانات الخاصة بتركيب البروتينات والذي يجب ان يكون بمواصفات خاصة منها ان يكون له تشابه لا يقل 30 % مع بروتين الاستعلام وبعض الأحيان يمكن استعمال 20 % كحد حرج مع مواصفات أخرى ، ويجب العناية بهذه الخطوة لانها تمثل الأساس لبقية الخطوات . واختيار القالب يستلزم البحث في بنك بيانات البروتين لاختيار البروتينات المشابه ذات التركيب المعروف ، وبما ان البحث سيكون كبير لذلك تستعمل معه الطرق السريعة مثل خوارزميات BLAST او FASTA وان كان البحث باستعمال البرمجة الدائميكية يؤدي الى نتائج أفضل ، وفي جميع الأحوال يعتمد الوقت والسرعة على حجم قاعدة البيانات ، وهذه الخطوة يمكن ان تعطي خيارات متعددة وعندها يتم اختيار القالب

الذي له نسبة عالية من التماثل وأعلى وضوحاً وغيرها من المؤشرات ، ولكن بعض الأحيان قد لا يتوفر القالب في قواعد البيانات ويصبح اختيار القالب صعباً لذلك يتم اللجوء الى استعمال طرق أخرى مثل حدس وقياس اللياقة لتحديد مسافة التشابه ، وفي حالة عدم إمكانية الوصول الى قالب ملائم للبروتين ككل يمكن العمل مع الدومينات ذات الاصطفا الملائم .

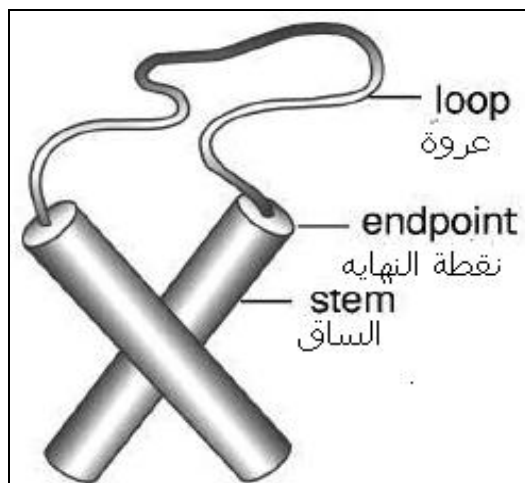
• اصطفا التواليات وتمثل الخطوة الثانية ، فبعد تحديد البروتين القالب يلزم إعادة اصطفا بروتين الاستعلام مع البروتين القالب وإعادة عملية الاصطفا لأكثر من مرة وذلك لأهمية الخطوة في التأثير مباشرة في عملية تشكيل النموذج النهائي ، لان الخطأ في عمليات الاصطفا في هذه المرحلة يقود الى إنتاج موديلات تركيبية خاطئة ، كما ان الخطأ في هذه الخطوة لا يمكن تلافيه في الخطوة القادمة ، ولذا وجب استعمال خوارزميات جيدة مثل T- COFFEE او غيرها ولكن يجب التذكر ان استعمال الخوارزميات الجيدة لا يعني انتقاء الأخطاء ، وعليه وجب التأكد منها خاصة الثمالات المهمة Key residues التي يجب ان تكون مصطفة بشكل صحيح .

• بناء الهيكل النهائي وتمثل الخطوة الثالثة ، فبعد الحصول على عمليات اصطفا جيدة وان الثمالات في بروتين الاستعلام تكون مشابهة للبروتين القالب فأن هذا يعني ان إحداثيات الثمالات المتقابلة في البروتين القالب تكون مشابهة للثمالات في بروتين الاستعلام . وفي حالة الثمالات المتماثلة فأن إحداثيات السلاسل الجانبية تكون مستنسخة على طول الذرات الرئيسية ، وإذا كانت الثمالات مختلفة فأن إحداثيات البروتين القالب تكون مشابهة لإحداثيات الذرات الرئيسية في بروتين الاستعلام ويعاد بناءها لاحقاً .

وفي حالة تشكيل الهيكل الرئيس يفضل استعمال تركيب قالب واحد الذي يكون قد تم اختياره على أساس النوعية الجيدة والوضوح العالي من بين خيارات متعددة (وان كانت نادرة)، ومثل هذا القالب يكون قليل الأخطاء ، وهذا السبب وراء إعادة عمليات الاصطفا قبل اختيار القالب .

• تشكيل العروات Loop modeling ، في عمليات الاصطفا التي تجري لوضع النموذج المجسم للبروتين تظهر بعض الفجوات نتيجة لعمليات الحذف والإقحام ، وفي هذه الحالة لا يمكن إقامة النموذج لانها ستؤدي الى وجود ثقب وتكون مصدراً كبيراً للخطأ ، وعندها يمكن عد العروات كبروتينات صغيرة وحل مشكلة الفجوات بطريقة البحث عن ما يشابه القطع في قواعد البيانات او استعمال طريقة البداية لإيجاد القطع الاحتياطية الملائمة .

أما اختيار الساق Stem والذي يمثل الذرات التي تسبق وتلحق العروة فبنائها يتم بدراسة ما يسمى منطقة رسو الساق Anchor region والبحث عن قطع في بنك بيانات البروتينات لإيجاد قطع بالطول نفسه ومتوائمة ، وفي الحقيقة توجد قطع بديلة متوفرة تلاؤم السياق . وأفضل عروة مع ساقها يتم اختيارها هي المعتمدة على تشابه التوالي بشرط ان لا تؤدي الى تصادم مع أجزاء التركيب الباقية كما موضح في الشكل 116.5 :



شكل 116.5 : العروات الممكن استعمالها عند إيجاد التركيب الثلاثي للبروتينات

أما طريقة البداية المستعملة لتشكيل العروات فهي تنتج العديد من العروات العشوائية ، وعندها يجب اختيار عروة واحدة لا تؤدي إلى التصادم مع أجزاء البروتين الأخرى وكذلك لها مستويات واطئة من الطاقة وقيم زوايا ϕ و ψ على مخطط راماندران ملائمة . ومن الجدير بالذكر ان بناء الشكل المجسم للعروات الصغيرة (3-5 ثمانية) يكون أسهل من العروات الطويلة . وتوجد عدة برامج تتعامل مع بناء العروات على اختلاف أطوالها والبعض منها يتعامل مع عروة تصل إلى 80 ثمانية .

- تشكيل السلاسل الجانبية ، بعد بناء السلاسل الأساسية لابد من إضافة السلاسل الجانبية وهذه تحتاج إلى تحديد الأبعاد المجسمة وذلك لأنها ستكون مهمة جدا في تداخلات البروتين مع الربائط عند المواقع الفعالة وكذلك التداخل مع البروتينات الأخرى . ولاختيار هذه السلاسل تعقيدات كبيرة . وخالصة هذه التعقيدات والاختيارات المبنية عليها هو اختيار سلاسل تكون لها قيم طاقة واطئة لعمليات التداخل مع الذرات المجاورة ، وتوجد برامج حاسوب تعنى بهذه الناحية التي تدخل في حسابها أمثلة التداخلات مع باقي مجسم البروتين ووظائف السلاسل الجانبية وعدم التصادم مع باقي أجزاء البروتين .

وفي العموم فإن البرامج المعنية بتصميم هياكل العروات والسلاسل الجانبية تهدف إلى تقليل الطاقة لتجنب حدوث تغييرات في التركيب النهائي ، ويتم بتحريك الذرات (للموديل على الحاسبة) بطريقة يجعل التوزيع الكامل الملائم هو الذي يكون له أقل مستوى من الطاقة ، وعندها يجب الحذر من تقليل الطاقة بشكل كبير جدا" لأنه سيؤدي إلى تحريك الشمالات من مواقعها الصحيحة ، أي ان تقليل الطاقة يجب ان يكون ضمن مدى معين ويكون فقط لإزالة الأخطاء .

- تقييم النموذج النهائي وتمثل الخطوة الأخيرة ويجب ان يقيم لملاحظة ملائمتها مع المؤشرات الفيزيو- كيميائية . وتشمل عمليات التقييم :

- * قياس الشذوذ في قيم الزوايا ϕ , ψ
- * طول الأواصر

* التصادم مع محاور التناظر (Close contacts)

وغيرها من المؤشرات ، ولذلك تستعمل طرق أخرى في التقييم مثل اخذ المواصفات التجسيمية لتحديد الأخطاء . وبالمقارنة مع المؤشرات الإحصائية العامة مع الموديل الذي تم بناؤه ستظهر فيما اذا كانت بعض مناطق التواليات قد طويت بشكل صحيح وأياها طويت بشكل غير صحيح . وقد أنشأت قواعد بيانات خاصة بالطريقة .

طرق قياس اللياقة والتعرف على الطويات Threading and fold recognition

تكشف الطريقة وتعرف البروتينات المتشابهة من حيث التركيب بغض النظر عن وجود او غياب تشابه التواليات . والطريقة في أساسها تعتمد على انه يوجد حوالي 1000 من أنواع البروتينات المطوية مقارنة بملايين التواليات البروتينية ، وهذا يعني ان العديد من البروتينات يمكن ان تشترك في طريقة الطوي ومن ورائها التراكيب الثلاثية الأبعاد حتى عند غياب تشابه التواليات مما يشير الى ان التراكيب الثانوية وغيرها من المستويات تكون أكثر ثبوتا أثناء عمليات التطور ، ويساعد هذا في إيجاد طرق حاسوب للتعرف على تركيب البروتين بعيدا عن تشابه التواليات التي يمكن ان تستعمل في تحديد الطويات التركيبية والتركيب النهائي للبروتينات غير المعروفة وذلك بالمقارنة مع قواعد البيانات الخاصة بتراكيب البروتينات واختيار الأفضل .

والطرق المستعملة في هذا التوجه تعتمد على إجراء الاصطفاف لزوج من التواليات المعتمدة على الطاقة (Pairwise energy based) وهي التي تمثل طريقة حدس مقياس اللياقة والاعتماد على النسق Profile based والتي تعرف على انها تمييز الطوي Fold recognition والمصطلحات تستعمل بشكل متبادل في المراجع الحديثة .

• الطريقة الأولى وهي حدس التركيب المتماثل وعند استعمال الطريقة يتم البحث عن توالي الاستعلام في قواعد البيانات الخاصة بالتراكيب Structural fold database لإيجاد أفضل تلاؤم باستعمال مؤشر الطاقة وتشمل العملية صف توالي الاستعلام مع كل طوية تركيبية في قاعدة بيانات الطويات باستعمال البرمجة الديناميكية الدقيقة او الطرق السريعة الموجهة. وعملية الصف الموضوعية هذه تعدل للوصول الى أوطأ مستوى للطاقة وبالتالي أفضل تلاؤم وتتم عمليات التعديل باستعمال برامج خاصة .

وتلي ذلك خطوة بناء نموذج أولي Crude model لتوالي الاستعلام وذلك بصف الثمالات في تركيب القالب مع ما يقابلها في توالي الاستعلام . وبعدها يتم حساب مؤشرات الطاقة للنموذج الأولي والتي تشمل طاقة تداخل الثمالات والطاقة الخاصة بكراهية الماء وطاقة الذوبان Solvation energy ، وبعدها يتم ترتيب الموديلات او النماذج اعتمادا على مؤشرات الطاقة لإيجاد الطويات ذات الطاقة الأقل التي تعنى أفضل تلاؤم للطيات

• طريقة النسق وفيها يتم بناء نسق لمجموعة من البروتينات ذات العلاقة التركيبية والمعلومات الإحصائية من الثمالات المصطفة وتستعمل لإنشاء النسق الذي يحوي عددا من الدرجات التي تصف النزعة لكل 20 ثمالة . ودرجات النسق Profile score تحوي على معلومات تصف أنواع التركيب الثانوي ودرجة التعرض للمذيبات وكذلك تصف القطبية وكراهية الماء . ونظرا لكون الطريقتان أعلاه لا تعتمد على التواليات لذلك تكون أكثر حساسية

في أيجاد العلاقات التطورية من غيرها . ولكن الدقة والحساسية الكبيرة قد تكون نقطة ضعف لهذه الطرق وذلك لان الحساسية العالية ترتبط بانخفاض التخصص ، والحدس بهذه الطرق تظهر فيه نسبة عالية من النتائج الموجبة الكاذبة ، لذا تؤخذ نتائجها على جانب من الحذر ، وفضلا" عن ذلك فإن من سلبيات الطريقة انها تكون ممكنة التطبيق فقط للبروتينات ذات التركيب المعروف الموجودة في قواعد البيانات وتفشل مع غيرها ، فضلا" عن انها تكون غير دقيقة على مستوى الذرات وتعطى طوبوغرافية عامة لتركيب البروتين . وللطريقة برامجها الخاصة التي يعمل بعضها مع بروتينات معينة ويفشل في الأخرى

الطرق المعتمدة على المعلومات الأولية (طرق البداية)

اتضح من المناقشة أعلاه ان كل من طريقة التشابه وطريقة حدس التركيب المتماثل تعتمد على وجود تركيب قالب في قواعد البيانات لغرض الوصول الى تحديد تركيب البروتين وفي حالة عدم وجود القالب تفشل هذه الطرق .

ومن المعروف ان البروتينات في الطبيعة تنطوي اعتمادا" على معلومات موجودة فيها وليس اعتمادا" فيما اذا كان التركيب موجود في قاعدة البيانات ام لا . وهذا واضحا" من ان بعض المعلومات في توالي البروتين تضيف عليه التركيب الخاص ، ومن وجهة نظر الدراسات الفيزيائية - الحيوية فإن معظم البروتينات تنطوي بشكل تلقائي الى تركيب ثابت عند اقل ما يمكن من الطاقة وهي التي تسمى الحالة الأصلية وان عمليات الطوي هذه لا تكون عشوائية ولكن تتم بآليات قد يكون بعضها غير معروف لحد الآن ، وقلة المعلومات هذه تمثل الأساس لطرق البداية للحدس . وكما يشير اسم الطريقة ان عملية الحدس او التنبؤ تعتمد على المعلومات المتوفرة في التوالي بدون مساعدة المعلومات المتوفرة عن التركيب ، والفائدة من الطريقة ان عملية الحدس لا تنقيد بما معروف من الطويات . ونظرا" لكون القوانين الفيزيوكيميائية التي تتحكم بطوي البروتينات لا تزال غير كاملة الوضوح لذلك فإن العلاقات المستعملة في الطريقة تعرفها الدقة .

اما البرامج الخاصة بالطريقة فاعليتها غير قادر على محاكاة عملية طوي البروتينات والبعض منها يستعمل الطرق السريعة الموجهة ، ولكن نظرا" لكون البروتينات الطبيعية تكون اقرب الى مستويات الطاقة المنخفضة لذلك تصمم معظم البرامج على أساس تقليل الطاقة ، وهي تعمل لإيجاد كل الحالات الممكنة للتوزيع الجزيئي الممكن لإيجاد التوزيع الذي له أوطأ طاقة . ولكن مع هذا فإن البحث عن حالة الطوي بأقل طاقة مطلقة قد لا يكون ملائما" لحقيقة الواقع وهذا ما يشكل أهم المآخذ على هذا التوجه في تحديد طوي البروتينات . كما ان البحث عن كل التوزيعات التركيبية لا يكون ملائما" للعمل بالحاسوب فقد وجد انه عند استعمال واحد من أسرع الحواسيب في العالم الذي يقوم بتريليون عملية ١ ثانية لإيجاد التوزيعات والتشكيلات لبروتين مكون من 40 ثمالة يحتاج الى 10 - 20 سنة ولذلك تستعمل البرامج الطرق الموجهة والسريعة لتقليل البحث ، وبعض برامج هذه الطرق تعتمد الى البحث عن قطع من التواليات واستعمال طريقة حدس التركيب المتماثل للتسريع في إعطاء موديل لبروتين غير معروف . ويجب التذكر دائما" ان كل برامج طرق البداية هي بعيدة عن النضج ودقة الحدس ولا تعد مفيدة من الناحية العملية ولذلك يتوقع ان تنقلص البرامج والتوجهات الخاصة بها ولكن هي لا تزال قيد الاستعمال في عدد من المجالات . ولكن لحصول هذه الحالات لابد من إجراء عمليات تقييم للطرق والبرامج التي تعتمد عليها .

حس التركيب الرباعي وتداخلات البروتينات

البروتينات تكون غير متناظرة وينتج التركيب الرباعي من تداخل عدد من الوحدات الفرعية الناتجة من التركيب الثلاثي بقوى لا تساهمية وبعض الأحيان الجسور الكبريتيدية الثنائية وقد تكون هذه الوحدات متماثلة أو متباينة وتختلف في أعدادها لتكون التركيب الرباعي الذي يكون بروتين قليل التكوثر Oligomeric protein وقد يكون دائريا أو هرميا أو غيرها من التوزيعات . وهذا التجمع يكون مهما في أداء وظائف البروتينات إذ تشارك هذه البروتينات في العمليات الأيضية ونقل الإشارات وتضاعف DNA وغيرها من الفعاليات الحيوية .

وتمتاز البروتينات التي لها تركيب رباعي فضلا عن أهميتها في أداء الوظائف الضرورية فهي تتميز بأنها تجري عمليات الإصلاح فيها وذلك باستبدال أي من الوحدات الفرعية المعطوبة بأخرى سليمة . وبصورة عامة فإن تداخل البروتينات مع بعضها وتكوين التراكيب الرباعية تكون مرتبطة مع بعضها ودراسة التداخلات تكون مهمة لفهم الوظائف . وقد وجد أن البروتين الواحد يمكن أن يتداخل مع 10 من البروتينات الأخرى مؤديا إلى إنتاج الآلاف من البروتينات النهائية . والبروتينات المودعة في بنك بيانات البروتين لها بيانات ومعلومات تشير إلى أن البروتينات التي لها درجة من التماثل 30% تميل للتداخل مع بعضها وتكوين تراكيب ثابتة ولكن بتفاصيل تداخل مختلفة ، أما إذا كان التماثل أقل من 30% فيكون التداخل ضعيفا أو لا يحصل تداخل . ولذا فإن التداخلات القوية تؤدي إلى تكوين معقدات ثابتة ، أما إذا كانت التداخلات ضعيفة فهي تحدث بشكل عابر أو وقتي ، وتحليل تداخلات البروتينات يساعد في التعرف على وظائف البروتينات غير الموصوفة سابقا . كما أن تحليل التعبير البروتيني يكون أفضل من تحديد الوظائف مقابل تحليلات الجينوم الذي ليس بالضرورة أن يقود إلى حس وظائف البروتينات . وفي هذا المجال يحتاج الجانب العملي إلى التعريف بالبروتينات على مقياس كبير وتحتاج إلى المعلوماتية الحيوية ووسائلها للبحث في قواعد البيانات عن تطابق البروتينات والبيبتيدات . وتفيد عمليات حس التداخلات في تهميش الجينومات وفهم وظائف الجينات التي تشفر لها

ونتيجة للدراسات المستفيضة لكيمياء البروتينات والبايولوجي الجزيئي جمعت كميات كبيرة من البيانات التي لها علاقة بتداخل البروتينات ، وجمعها في قواعد البيانات يساعد في استخلاص واسترداد هذه البيانات مما أدى إلى تطوير العديد من البرامج .

إن التركيب الرباعي ينتج من ترابط الوحدات الفرعية ذات التركيب الثلاثي كما ذكر أعلاه . والترابط يعتمد على وجود البقع المتوافقة أو المتكاملة Complementary patches على السطوح لذلك كانت هذه البقع المدفونة في مناطق التلاقي مهمة في تحديد التركيب الثلاثي والرباعي ، ومثل هذه البقع يمكن تحديدها من التوالي الأولى للبروتينات بالتجربة ولكنها تكون مكلفة وبطيئة كما أن تحديد التركيب الرباعي باستعمال الأشعة السينية أو الرنين النووي المغناطيسي يكون مكلفا وبطيئا ، لذلك انبرت المعلوماتية الحيوية لتقديم الحلول البديلة ، لذا فهناك العديد من الطرق التي تستعمل أسس مختلفة لحس التداخلات والتركيب الرباعي ، وأغلب البرامج تعتمد على الدراسات التطورية لتواليات الجينات Gene linkage patterns أو نمط اندماج الجينات والتوجهات العامة تكون أما :

دراسة التماثل الجذري (المارة الذكر عند مناقشة التركيب الثلاثي)

- **طريقة حدس مقياس اللياقة** ولذلك فالبروتينات ذات التشابه الأقل من 30% تظهر تراكيب رباعية مختلفة .

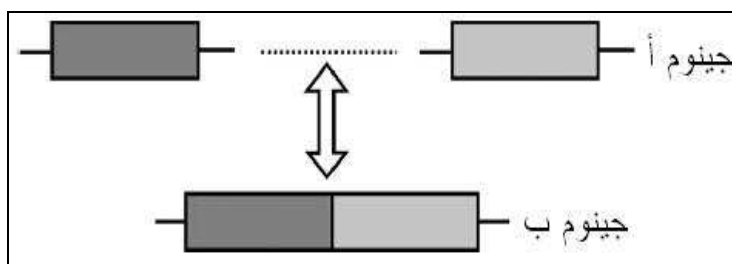
وهذه الطرق وكما ذكر سابقا" تحتاج الى أيجاد قالب للعمل وإيجاد موديل ثلاثي الأبعاد يساعد في حل مشكلة حدس التركيب الرباعي . ولكن المشكلة في هذه الطرق ان النماذج لا تستوعب التغييرات الكبيرة في الفراغ او ترتيب العروات عند الارتباط مما يؤدي الى التصادم عند مناطق الالتقاء ولا يمكن ان يعدل الترتيب النسبي الى درجات عالية من الدقة في حدس تلاقي سطوح البروتينات .

وهناك طرق مختصة ضمن هذه التوجهات ومنها

- الطرق المعتمدة على التواليات وفي هذه الحالة اذا كان هناك 60% من التماثل بين البروتينات المستعملة يمكن تحديد التركيب في حين اذا انخفضت النسبة 30-60% يمكن الحصول على حدس للتراكيب العامة الثابتة ولكن باختلاف التفاصيل . ومثل هذه الطرق تظهر فيها نسبة عالية من النتائج الموجبة الكاذبة نظرا" لعدم الأخذ بالمظاهر التركيبية اللازمة لتحويل البروتينات وهذه يمكن ان تقلل باستعمال بعض الوسائل مثل SVM .

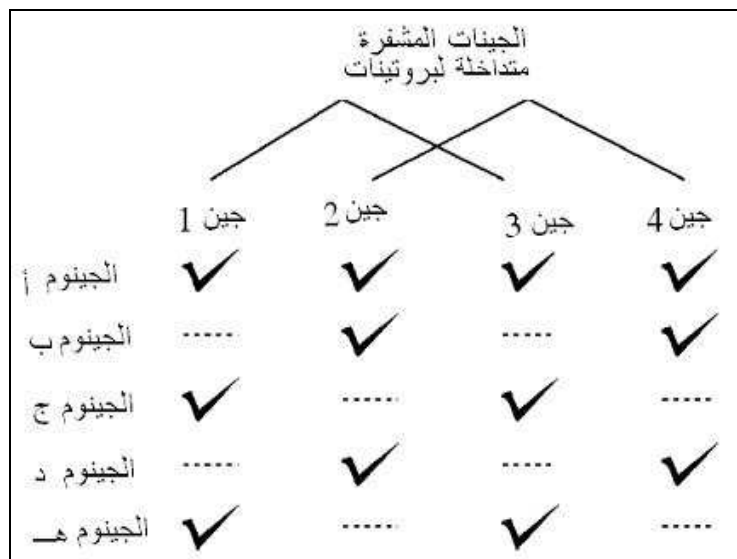
- الطرق المعتمدة على الدومينات الفعالة وهذه تكون مهمة في حدس التركيب الرباعي نظرا" لانها مهمة في وظائف البروتينات . وفي هذه الحالة فأن استعمال البرامج الخاصة بها تعتمد الى اختيار البروتينات من قواعد البيانات التي لا تكون فيها ثملالات غير معروفة (مثلا" معلمة بـ X او Z) فضلا" عن اختيارها بأطوال معينة ، اذ تستبعد البروتينات المكونة من اقل من 50 حامض أميني لانها لا يمكن ان تشكل دومين وكذلك تستبعد البروتينات ذات الأطوال الأكثر من 6000 حامض أميني فضلا" عن إزالة المكررات ويكون الحد الفاصل لتشابه تواليات البروتينات 25% . ويمكن استعمال قاعدة بيانات عوائل البروتينات Pfam لانها تضم العديد من دومينات البروتينات التي حددت باستعمال نماذج ماركوف الخفية وتحمل أرقام تسجيل خاصة بها متوفرة في قاعدة بيانات البروتينات السويسرية لتسهيل الحصول على التواليات . والبرامج المعتمدة على الدومينات يتم إخضاعها لعدد من الفحوص الإحصائية وأهمها Jackknife test الذي يستعمل عند وجود تماثل 25%.

- استعمال الدومينات ضمن ظاهرة دمج الدومينات وهذه تعتمد على تواليات الدومينات في DNA والطريقة تسمى Rosetta stone protein الموضحة في الشكل 117.5 :



شكل 117.5 : طريقة دمج الجينات للبروتينات المتداخلة

- طريقة إرساء البروتينات Protein- protein docking ، الطريقة مستعملة منذ زمن بعيد لبناء معقدات التراكيب من مكوناتها ، وتصلح الطريقة فيما اذا كان التركيب الثلاثي للبروتينات معروفاً . وتتم الطريقة بجعل احد البروتينات ثابتاً (راسي) والآخر يدور عليه وتترجم المعلومات نسبة للبروتين الأول ، والطريقة تعاني من ارتفاع نسب النتائج الموجبة الكاذبة وتكون غير قادرة على وصف التغيرات التوزيعية التي ترافق تكوين المعقدات ، أي ان النتائج تكون غير صحيحة نتيجة لحصول تغيرات أثناء تجميع الوحدات ، والمشكلة في هذا تكمن عند تلاقي سطوح البروتينات ، وقد تمت محاولة إدخال مؤشرات مرونة البروتين ولكنها لم تقلح في تحسين النتائج .
- حدس التداخلات باستعمال الجينات المتجاورة وهذه تعتمد على ان ترتيب الجينات يكون غير ثابت بين جينومات الأحياء بدائية النواة ، ولكن اذا كان هناك ترابطاً بين الجينات فهذا يشير بقوة الى وجود الاوبرونات التي تشفر لبروتينات تكون لها علاقة وظيفية وتكون مترافقة فيزيائياً . وطريقة الحدس هذه للتداخل بين البروتينات ممكنة في الأحياء بدائية النواة . اما بالنسبة للأحياء حقيقية النواة فإن ترتيب الجينات يكون اقل قوة في حدس تداخلات البروتينية .
- طريقة إضافة الحوامض الامينية الكاذبة Pseudo amino acids ، الطريقة طورت أصلاً لإيجاد مواقع البروتينات في الخلية . فاعتماد التوالي الأولي في البروتينات يؤدي الى عدد غير محدد من الأنماط نتيجة الاختلافات في الطول وترتيب الحوامض فيه ، لذا يمكن اللجوء الى تصنيف البروتينات وفقاً للحوامض الامينية المكونة لها . ففي الحالة الاعتيادية هناك (20) حامض أميني ، اما باستعمال هذه الطريقة فهناك صيغة كاذبة هي $(20 + \lambda)$ أول جزء في λ يمثل حامض أميني اعتيادي ثم تضاف حوامض أمينية أخرى (نظرياً) ثم تستعمل هذه القيم في معادلات خاصة لإجراء الحسابات عليها واستنتاج التركيب الرباعي نظرياً وتستعمل فيها عدة فحوص إحصائية وأفضلها Jackknife test . وفي العموم فإن هذه الطريقة النظرية تكون قاصرة ، ويمكن استعمال خوارزمية الجوار الأقرب لتحسين الحدس لان احتمالات الخطأ فيها قليلة وتطبق خاصة للبروتينات غير معروفة التوزيع وبعض المجاميع الموجودة في قواعد البيانات .
- طرق الحدس المعتمدة على العلاقات التطورية وتعتمد الطريقة على تشابه اثنين من البروتينات من الناحية الطبوغرافية وانها كانت ذات علاقة قوية من ناحية معدل التطور فإن احتمال تداخلهم يكون كبيراً ، أي بالاعتماد على حالة التناظر Orthology وذلك لان حدوث طفرات في مواقع سطوح التداخل لأحد البروتينات فإن طفرات مشابهة تكون قد حصلت للبروتين المقابل لضمان عملية التداخل وبالتالي فإن البروتينين تكون لهم شجرة التطور نفسها مما يعكس تأريخ تطوري متشابه .
- والمنطق وراء هذا التواجد المترافق هو ان البروتينات تقع ضمن معقدات فإذا فقد احد مكونات هذه المعقدات فإنه يؤدي الى فقدان المعقد كله . وتحت الضغط الانتخابي فإن البروتينات المتبقية (غير الفعالة) من المعقد سوف تفقد أثناء عملية التطور لانها تصبح غير ضرورية من الناحية الوظيفية . وهذه الطريقة اقل دقة من طرق دمج الجينات المتجاورة . وهي موضحة في الشكل 118.5 :



شكل 118.5 : طريقة التواجد المترادف لحديث البروتينات

ولتحليل التطور يستعمل معامل الارتباط (r) بين مسافات التطور وقد وجد ان قيم r الأكبر من (0.8) تشير الى تداخل قوي للبروتينات.

• دراسة التداخل المترافق باستعمال طرق هجينة ، ان الطرق المذكورة أعلاه يعتمد كل منها على افتراض محدد ويمكن ان تظهر نوعاً من الانحياز او الانحراف المرتبط بالافتراض المعتمد . ونظراً لصعوبة تقييم أي من الطرق المستعملة ، لذلك ينصح باستخدام أكثر من برنامج حدس واعتماد أكثر من طريقة لتقليل الانحياز ومعدل الخطأ للحصول على مستوى عال من الثقة في حدس البروتينات المتداخلة ، وتوجد بعض البرامج التي تستعمل هذا النهج .

قواعد البيانات

هناك عدة قواعد بيانات للدومينات البروتينية المسجلة بالاعتماد على تراكيب البروتينات ثلاثية الأبعاد ، ومسجلة لها ثملالات التلامس على السطوح.

ومن أهم المواقع الخاصة بالتداخلات والتراكيب الرباعية للبروتينات هو STRING (Search tool for the retrieval of interacting genes / proteins) التابع

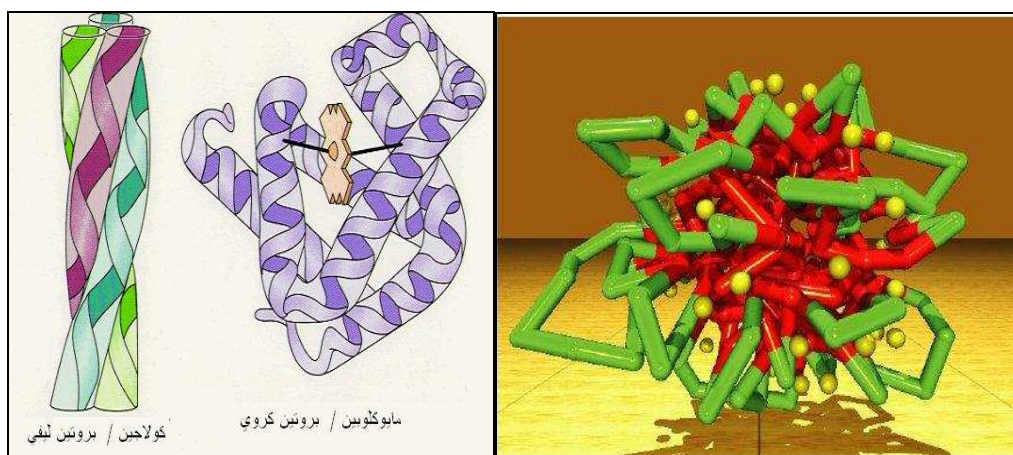
للاتحاد الأوروبي ومزود الخدمة هذا يحوي على مجاميع متناظرة Orthologous groups من العديد من الأحياء التي تم تحديد توالياتها المصنفة على أساس COG (المارة الذكر) ، ويظهر الموقع قائمة من البروتينات والمعلومات المتوفرة عنها فضلاً عن مصورات تمثلها . ويحتوي الموقع على برامج لحديث التداخلات التي تشمل التداخلات المباشرة (الفيزيائية) وغير المباشرة (الوظيفية) . وتوجد مواقع أخرى يمكن ان تؤدي العديد من هذه الأغراض.

التنبؤ بتراكيب البروتينات الكروية

التنبؤ بالتراكيب الثانوية للبروتينات الكروية مهمة ليست سهلة وبقيت كذلك على مرور عقود من الزمن وذلك لان التركيب يعتمد على البيئة ، فتكوين الحلزون ألفا يعتمد على

مدى قصير من التداخلات في حين ان صفائح بيتا تعتمد على مدى طويل من التداخلات والذي يكون من الصعوبة التنبؤ بها ، وبعد عقود من العمل تم إحراز تقدم لا يتجاوز 50-75% . والطرق المستعملة اما تكون معتمدة على الصفات الأصلية للبروتين وتواليه أي استعمال طرق البداية او معتمدة على التشابه الذي يعتمد بدوره على نتائج صف العديد من التواليات . والطريقة الأولى التي استخدمت في وقت مبكر تعتمد على حسابات إحصائية للثملات في التوالي .

وتعود المعوقات المذكورة أعلاه الى طبيعة تركيب البروتينات الكروية اذ تميل فيها الثملات الكارهة للماء باتخاذ المواقع الداخلية مشكلة لب كاره للماء كما موضح في الشكل 119.5 الذي يوضح فرق التركيب عن البروتينات الليفية .



شكل 119.5 : تراكيب البروتينات الكروية والليفية

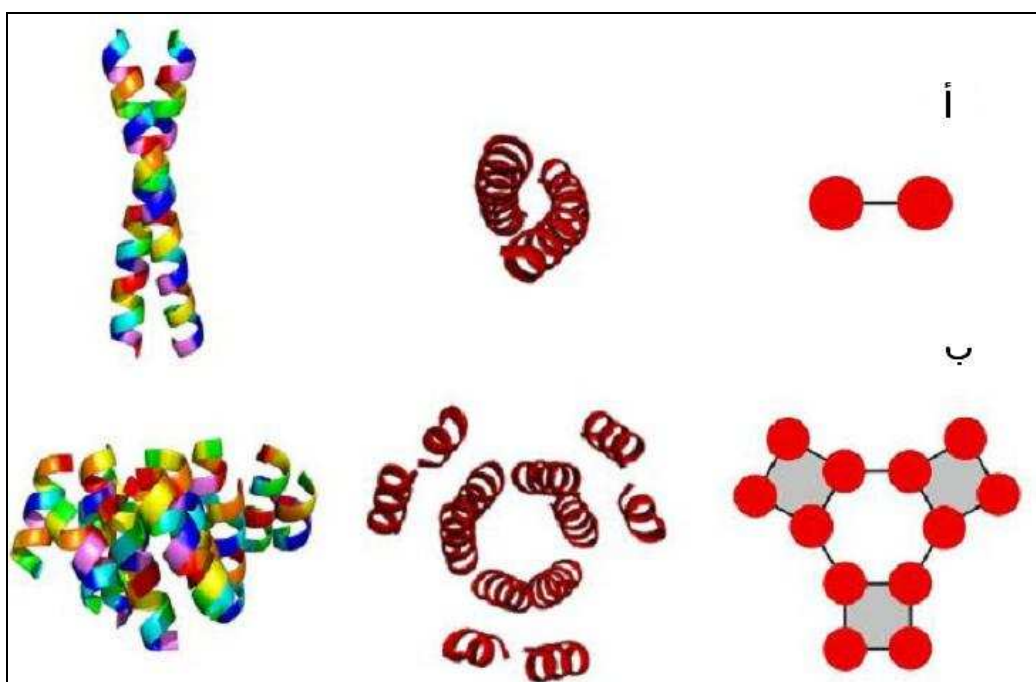
واللب الكاره للماء له تأثير كبير في تركيبة ووظيفة هذه البروتينات فالثملات اللاقطبية تكون بعيدة عن الماء وهذا الحال يؤدي الى جعل الثملات في وضعين اما معرضة او مدفونة ، وهناك مساحة مخصصة لكل ثمالة من الثملات المدفونة خاصة الكارهة للماء، وهناك عدة طرق لقياس كراهية الماء من قبل الحوامض الامينية وتختلف مقاييس كره الماء Hydropathy index اعتمادا" على الطريقة المستعملة لإيجاد هذا المقياس وتكاد تتفق الطرق على ان السستيين هو الأكثر كراهية للماء ولكن هذا قد لا يكون صحيحا" ، اذ يميل السستيين الى التواجد في داخل لب البروتين وهذا قد يعود الى قابليته على تكوين الجسور الكبريتيدية S=S التي تكون الى الداخل . ويستعمل المقياس بكثرة في حدس حلزونات ألفا في الأغشية عند غياب إمكانية تحديد الاصطفااف وذلك بتطبيق عددا" من الطرق أهمها استعمال طريقة الموجيات Wavelet ، التي تعدل فيها النتائج وفقا" لمقياس الكراهية لثملات الحوامض الامينية بعد تحديد أعلى وأوطأ تردد من القيم والإشارات المعطاة .

اما الطرق الأخرى فتعتمد على تحديد عدد الحوامض الامينية التي هي في حالة تلامس ١ عدد الثملات الكلي وهي طريقة بديلة لحدس التركيب الثانوي ومكنت من الوصول الى دقة 70% . فذرات التلامس تمثل عدد ذرات الكربون في دائرة قطرها 8 انكستروم المتمركزة عند الذرات

$C\alpha$ للشمال ، وبرسم عدد ذرات التلامس مقابل عدد الشمالات لتوالي البروتين يمكن ان يعطى البديل للتركييب .

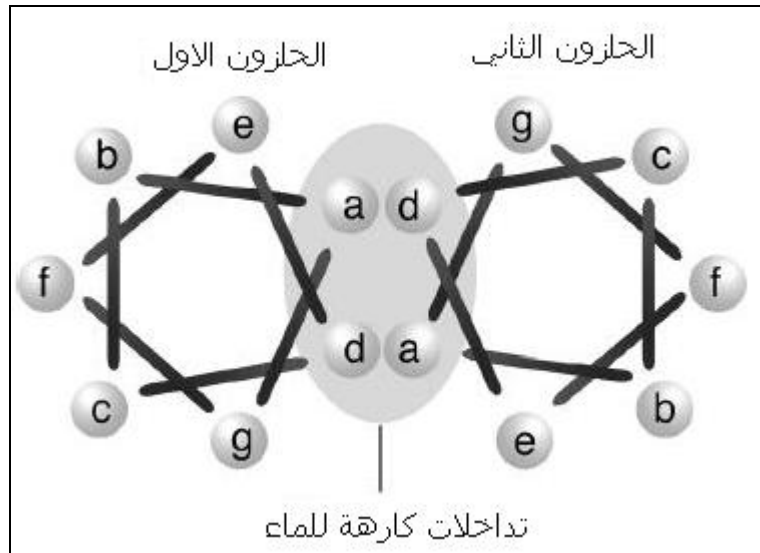
حده البروتينات الملتفة - الملتوية Coiled coil proteins

تكون تراكييب هذه البروتينات بمثابة حلزونات ملتفة جدا"وتضم اثنين او أكثر من حلزونات ألفا الملتفة على بعضها بطرق مختلفة مكونة تراكييب بشكل حزم ، لذلك فهي مرحلة من التراكييب الثانوية الفائقة أي تكون بين التركيب الثانوي والثلاثي كما موضح في الشكل 120.5 :



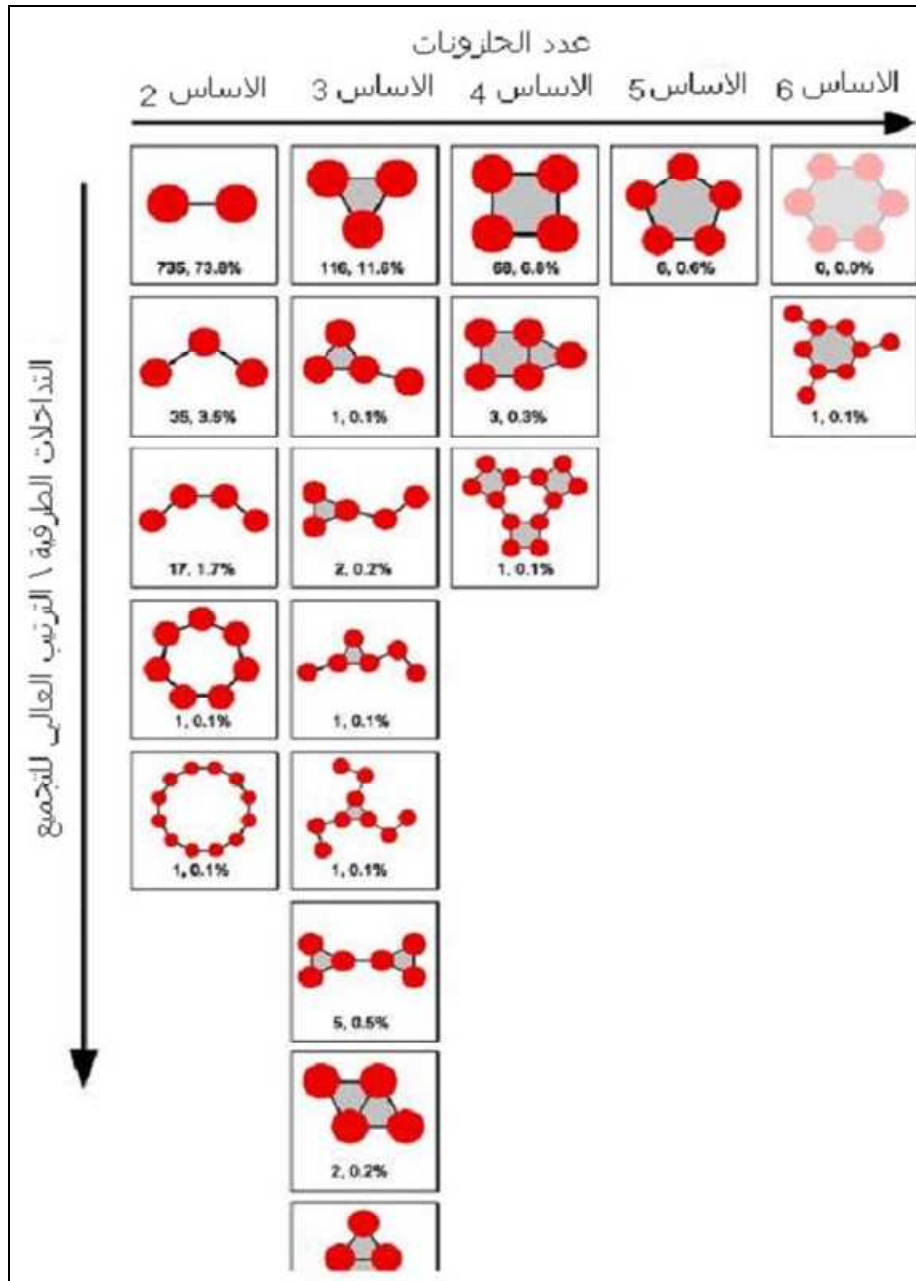
شكل 120.5 : التراكييب البروتينية فائقة الالتفاف

فالتراكييب في الصنف (أ) هي التقليدية المكونة من اثنين من الحلزونات المتوازية بمناظر مختلفة ، اما الجزء (ب) يمثل البروتينات الملتفة المعقدة والمكونة من أكثر من تركيب . ان توزيع هذه التراكييب وأشكالها المجسمة يكون ضروريا" في تداخل البروتينات مع بعضها وكذلك مع الجزيئات الأخرى . والبروتينات الحاوية على هذه الدومينات تعمل في العادة في تنظيم عمليات الانتساخ او المحافظة على الهيكل الخلوي . ويعتمد تداخل والتفاف هذه البروتينات على وجود توالي مكون من سبع شمالات مختلفة في كراهيتها للماء ، فالأولى والرابعة تكون كارهة للماء التي تسمح للبروتينات بالتداخل والباقي تكون محبة للماء وتواجه المذيب كما موضح في الشكل 121.5 :



شكل 121.5 : تداخلات كارهة للماء في البروتينات الملتفة - الملتوية

وتوجد هناك جداول لعدد الحلزونات المشتركة في تكوين التراكيب الملتفة الملتوية كما موضح في الشكل التالي (122.5) الذي يوضح عدد الحلزونات ونسبها للبروتينات المعروفة المسجلة في بنك بيانات البروتينات PDB .



شكل 122.5 : جدول دوري لعدد الحلزونات في التراكيب الملتفة - الملتوية

ان تكرار التماثل في التواليات يشكل الأساس في وضع البرامج الخاصة لحس هذه البروتينات وعند تحديد الالتواءات فيمكن عندها تحديد التركيب الثلاثي الأبعاد . وهناك العديد من البرامج المستعملة للكشف عن اللفات الملتوية فمثلا "البرنامج Coils يقوم بالعمل بنافذة 14-21 او 28 ثمانية ويقارن احتمالية التوالي مقابل مصفوفات مكونة ومجموعة من لفات مكونة من اثنين من الأشرطة المتوازية الملتفة ويسجل درجات التشابه وترتيب الاحتمالات .

والبرنامج دقيق بالنسبة لأجزاء اللغات المواجه للمذيب خاصة في التراكيب الملفوفة الى اليسار واقل حساسية للأنواع الأخرى من تراكيب اللغات الملتوية مثل الملتقة الى اليمين او المدفونة الى الداخل . وهناك برامج أخرى تستعمل مصفوفات مختلفة .

التنبؤ بالقطيفات والدومينات

من أهم جوانب تطبيق التواليات البايولوجية هو تحديد القطيفات والدومينات في البروتينات او غيرها من الجزيئات الحيوية . والقطيفة قطع متشابهة لها الطول نفسه موجودة في تواليات مختلفة وقد تكون متعددة أي تتكرر في التواليات ، والقطيفات بالذات تكون تواليات قصيرة مرتبطة بأداء وظيفة معينة في البروتين او DNA ويكون ارتباطها بموقع تركيبي معين لأداء وظيفة معينة ومن أمثلتها المثالية قطيفة أصبح الزنك بطول 10 - 20 حامض أميني . اما الدومين فيمثل نمط من التوالي الثابت ، ويعرف على انه وحدة وظيفية وتركيبية مستقلة وتكون أعداد الثمالات فيه من 40 - 700 وبمعدل طول 100 ثمالة والدومين قد يحتوي على قطيفة او لا يحتوي ، ومن أمثلته الدومينات التي تستعرض الأغشية الخلوية . وكل من القطيفة والدومين تمثل أجزاء ثابتة في البروتينات من الناحية التطورية ولذا تميل للظهور كوحدة واحدة ويمكن ان تفقد او تكتسب او تدلف كوحدة واحدة .

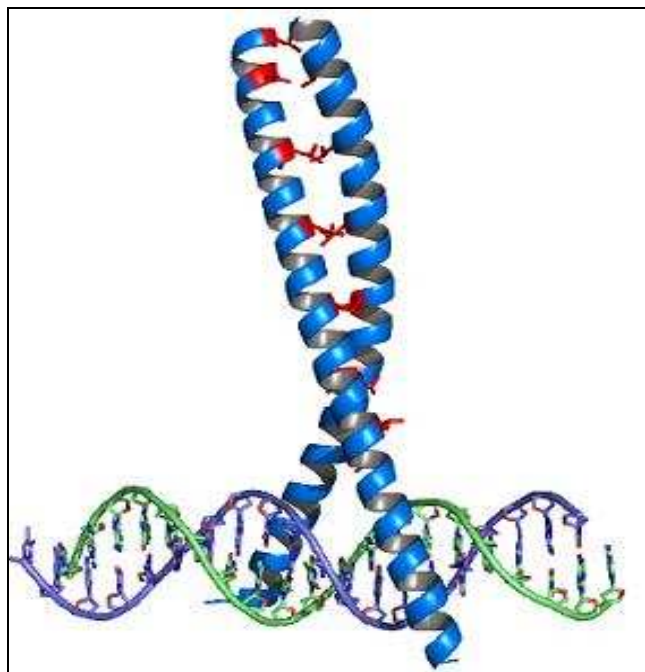
أهمية وأنواع القطيفات والدومينات

التعرف على هذه المكونات يكون على جانب كبير من الأهمية لان لها علاقة وثيقة بالخواص التركيبية والوظيفية للتوالي الذي توجد فيه وعليه تكون مهمة في تصنيف البروتينات . هناك أنواع مختلفة من القطيفات منها القصيرة وهي التي لا يكون لها تركيب ثانوي ، وتستعمل لوسم البروتين لتسليمه الى أجزاء أخرى من الخلية او توسم البروتين لإجراء الفسفرة عليه كما في القطيفة الخماسية EPYIA الموجودة في سموم Cag A للبكتريا *Helicobacter pylori* التي تجعل من السموم ضارية بعد فسفرتها . ويمكن ان توجد القطيفة في داخل اكسون ألجين وتشفر لقطعة تخص تركيب البروتين الذي سينتج عنه . اما عند وجودها خارج الاكسونات فهي تؤدي مهام التنظيم ، وعند وجودها في توابع ومكررات DNA فهي تؤثر في شكل جزيئات الحامض النووي .

وتستعمل في الكشف عن تداخلات البروتينات مع بعضها التي تتم بواسطة اللغات الملتوية المارة الذكر . ومن أفضل الأمثلة على هذه التراكيب هو زمام الليوسين المنزلق Leucine zipper ويحوي هذا الزمام على اثنين من حلزونات ألفا المتعكسة التوازي ترتبط مع بعضها بواسطة التداخلات الكارهة للماء لثمالات الليوسين والتوالي السباعي لها

$$L - X(6) - L - X(6) - L - X(6) - L$$

كما موضح في الشكل 123.5 :



شكل 123.5 : تركيب زمام الليوسين المنزلق Leucine zipper

ان اعتماد التواليات لوحدها يمكن ان يعطى نتائج غير صحيحة ولذلك طورت برامج جيدة تأخذ بنظر الاعتبار كل مكررات الليوسين والتوزيع المجسم التركيب للوصول الى حدس دقيق .

طرق التحري عن القطيفات والدومينات

نظرا" للافتراق التطوري الكبير بين الأحياء فأن العلاقات الوظيفية بين البروتينات لا يمكن تمييزها بطرق بسيطة مثل الاعتماد الكلي على عمليات الاصطفاف التي تجرى باستعمال BLAST و FASTA ومشتقاتها ، فضلا" عن ذلك فأن بعض الإنزيمات والبروتينات يمكن ان تؤدي أكثر من وظيفة لذلك لا يمكن الإفصاح عن وظائفها بنهميش واحد او أكثر عند البحث في قواعد البيانات ، لذلك كان الحل الأنسب هو تحديد القطيفات والدومينات والذي يعتمد بدوره على عمليات الاصطفاف المتعدد وبمساعدة موديلات ماركوف الخفية . وهناك توجهان في هذا المجال الأول استعمال نموذج Gibbs الذي يعتمد طريقة نمذجة مونتي كارلو - Monte Carlo sampling . والثاني تضخيم الانتروبي كما في البرنامج المعتمد MEME ، وهذه الطرق المعقدة يمكنها الكشف عن القطيفات في تواليات متباعدة . وأكثر الطرق تنتهج الخطوات الآتية :

- تحليل النسق وذلك بعمل اصطفاف شامل للتواليات المطلوب تحليلها لإيجاد المناطق الثابتة التي تمثل تواليات القطيفات من بين مجموع تواليات الاستعلام وتكون بمثابة مظاهر معروفة لعوائل البروتينات
- عزل المناطق الثابتة التي تم الحصول عليها لإنشاء مصفوفات النسق .

• ترتيب هذه المصفوفات اعتماداً على الدرجات ، ويتم حساب التكرار لكل حامض أميني أو نيوكليوتيد في كل موقع لمعرفة توزيع المناطق الثابتة وعدم الاعتماد على التوزيع التجريبي . Empirical distribution .

• استعمال مصفوفات النسق الناتجة للبحث عن تواليات أخرى للقطيفات . وفي حالة التواليات عالية العلاقة أو عند استعمال عدد قليل من التواليات في الخطوة الأولى يتم تعديل النتائج بإضافة أعداد كاذبة Pseudocounts لتعديل صفة التوزيع الممثل للقطيفات .

• الطرق الأخرى تستعمل نهج آخر إذ يتم تقطيع توالي الاستعلام إلى عدد من القطع المتداخلة لتشكيل ما يشبه العقدة وتدخل إلى البرنامج المستعمل الذي درب على تمييز القطيفات في البروتينات المعروفة ليقوم بفصل القطع ، والقطع الجديدة الناتجة تتم مقارنتها وذلك بصفها مع قطيفات معروفة والتي تكون معظمها ذات علاقة .

• وفضلاً عن ما ذكر أعلاه فهناك الطرق الإحصائية وتعتمد على استعمال النماذج إحصائية المذكورة أعلاه مثل استعمال النسق أو موديلات ماركوف الخفية لتشمل المعلومات الناتجة لكل الاحتمالات المشتقة من عملية اصطفاف التواليات المتعددة التي تساعد في التعرف على القطيفات والدومينات التي تكون سريعة وصالحة فقط للتواليات القصيرة عند البحث في قواعد البيانات العامة وفي هذا المجال هناك أكثر من توجه :

◊ تقليل عدد التواليات المستعملة للاصطفاف المتعدد للبحث عن القطيفات والدومينات وهو ما يعرف بالتعبير المنتظم وتكون بعض نتائجه ممثلة في القطيفة مثل القطيفة المسئلة عن فسفرة البروتين $[RK] - X - [ST]$. وهذه الطريقة تكون مختصرة ولكنها مفيدة عند استعمال بعض الضوابط التي تصف نمط التوالي فيها

◊ عندما يكون هناك حامض أميني ثابت فيذكر كما هو باستعمال شفرة الحرف الواحد للحوامض الأمينية

◊ عندما تكون هناك ثملالات متعددة (أي أكثر من حامض أميني) فتوضع بين قوسين مربعة [] ، ولكن بعض الأحيان هناك استثناء لبعض الثملالات وعندها توضح داخل أقواس المجموعة { }

◊ اما الثملالات غير المتخصصة الموجودة في موقع ما يشار إليها بـ X

◊ عند التكرار توضع العناصر داخل الأقواس الخاصة بها ويذكر بعد القوس رقم التكرار

◊ يفصل كل مقطع بشارحة كما مبين في المثال أدناه

E - X (2) - [FHM] - X (4) - { P } - L

ثم بعد ذلك يتم مطابقة القطيفات الناتجة مع توالي الاستعلام ويتم ذلك :

❖ إجراء التطابق الصارم أو الكامل Exact matching أي انه توائم مضبوط وبدون وجود أي تغاير أو تراخي في تطابقات القطيفات مع قواعد البيانات ، ونتيجة لهذه تنتج حالة وجود تلاؤم أو تناظر أو عدم وجود تلاؤم أو تطابق ، والطريقة لذلك تعطي فرصة لفقدان عدد من القطيفات ذات العلاقة والتي بها قليل من التغاير ، والنهج هذا يؤدي إلى العديد من النتائج السالبة الكاذبة ، والعيب الآخر في هذا التوجه ان التواليات

الجديدة سوف لا تجد لها مكان ويصبح مهجورا" ما لم يتم تحديثه باستمرار لمواكبة التغييرات .

❖ الطريقة الثانية في إجراء التطبيقات هي استعمال الموائمة المشوشة Fuzzy matches وتسمى أيضا "بالموائمة التقريبية" ، والطريقة أكثر مرونة وتعطي إمكانية أكبر للتوائم في مطابقة الثمالات ذات الصفات الكيموحيوية المتشابهة ، فمثلا "توالي القطيفة الحاوي على التايروسين تسمح بوجود الفنيل - النين والتربتوفان بهذه الطريقة باعتبارهم ضمن مجموعة الحوامض الامينية الحلقية ، وبذا فإن الطريقة تسمح بوجود أشكال متغايرة من القطيفات التي تكون وظائفها ثابتة .

قواعد البيانات

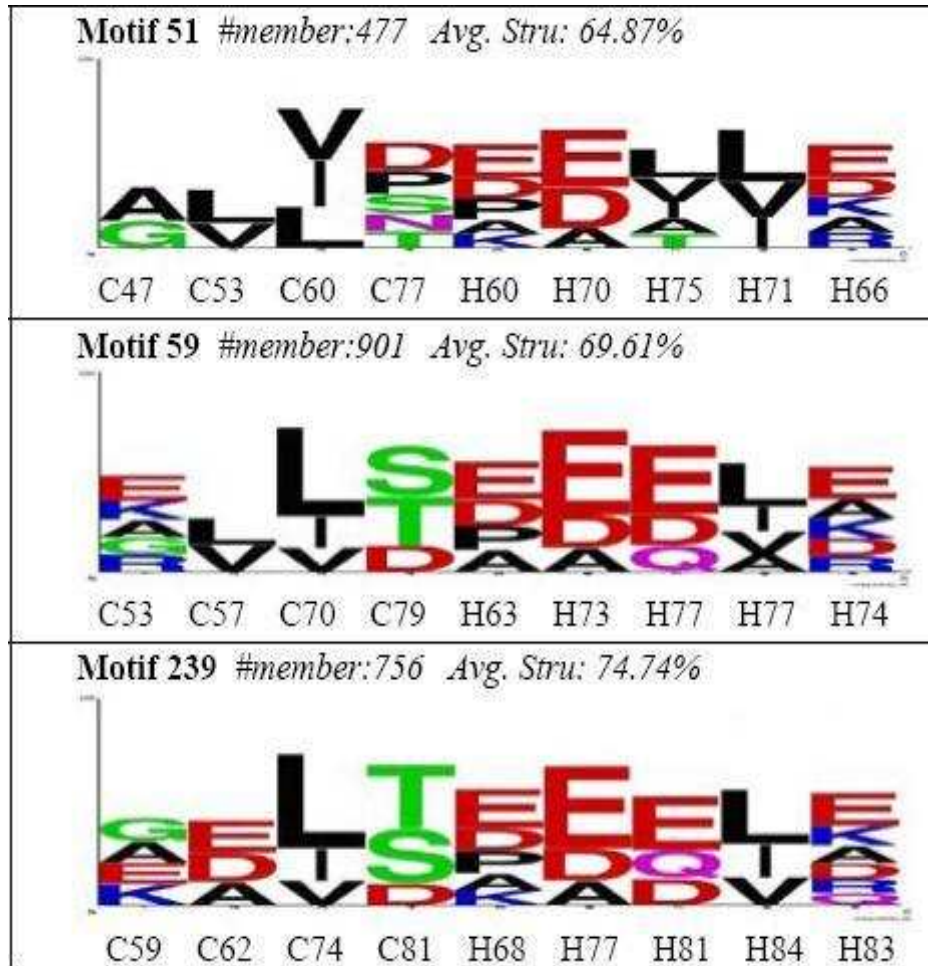
توجد عدة قواعد بيانات خاصة بالقطيفات والدومينات ، ونظرا" للاختلاف في الأسس المستعملة في إنشاء هذه القواعد وطرق إيجادها للتطبيقات فكل منها له نقاط قوة ونقاط ضعف ، ولذلك فإن قواعد البيانات هذه لا يمكن ان تغطي كل القطيفات او الدومينات الموجودة لذلك كان من الأفضل استعمال عدة قواعد للبحث للحصول على تغطية جيدة حول القطيفات والتوزيع وغيرها من المعلومات ، وفي حالة وجود تضارب في النتائج عند استعمال عدة قواعد لا بد من اللجوء الى وسائل أخرى لغرض التمييز .

بعض القواعد تستعمل لتصنيف البروتينات على أساس الوظائف المتماثلة وذلك بتحديد العلاقات التطورية والتركيبية ولعل أول وأهم قاعدة في هذا المجال هي PROSITE وتستعمل توجهات مختلفة لوصف البروتين وعائلته والبرامج الخاصة باستعمال التعبير المنتظم (المذكور أنفا") واعتمادا" على الوظائف الآتية من المراجع العلمية المتعلقة بالقطيفات ، ومن ميزات القاعدة انها تكون صغيرة ولا تعطي نتائج سالبة كاذبة اذ يكون معدل الأخطاء فيها لا يتعدى 20% ولذا فإن حدوث عملية التطابق او عملية عدم التطابق يجب ان يؤخذ بحذر ويجب إعادته .

وتوجد قواعد بيانات أخرى تعتمد على قواعد ثانوية وتكون أكبر من قاعدة البيانات السالفة الذكر وتستعمل طرق اصطفاة مختلفة وطرق مختلفة لإيجاد القطيفات ، لذلك فالبعض منها يحوي على قطيفات ودومينات غير معروفة الوظيفة . وتختلف درجات الثقة بهذه القواعد بشكل كبير وأكثر القواعد موثوقية هي التي يجري تدقيقها وتشذيبها عمليا" . وتختلف في نوعية ملفات المخرجات للنتائج ، ونظرا" لكثرة هذه القواعد تم الإشارة الى بعض مواصفاتها وعدم إدراجها يكون لضيق المجال فضلا عن اختفاء او تبديل إصدارات القواعد .

مخرجات برامج إيجاد القطيفات والدومينات

تختلف مخرجات قواعد البيانات والمواقع على الانترنت وكما ذكر في أكثر من موقع ان الرسوم البيانية تكون هي الأفضل .واهم البرامج المستعملة في هذا المجال هو برنامج الشعار Logo ، كما موضح في الشكل 124.5 :

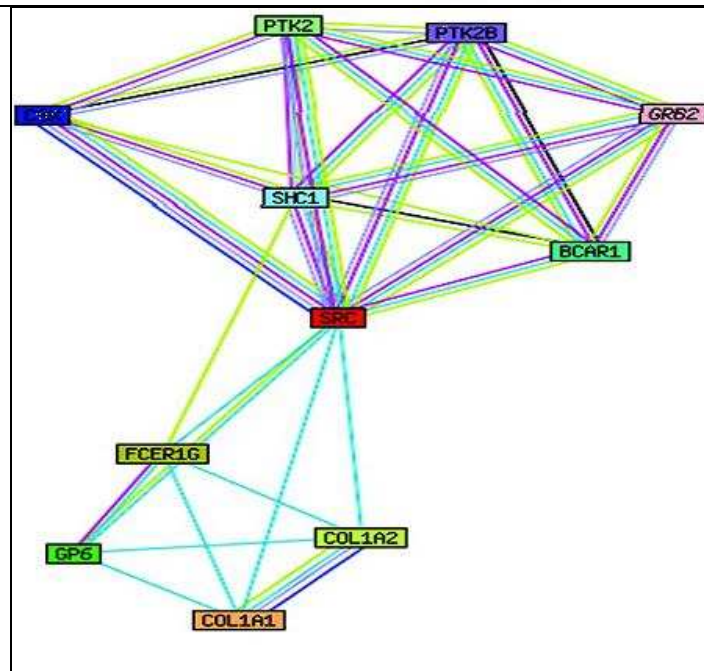


شكل 124.5 : نتائج برنامج الشعار لبعض قطيقات البروتينات

وهذه البرامج يمكن ان تدمج مع برامج أخرى للحفاظ على القيم الإحصائية والمؤشرات الأخرى للبرامج التي تكون في النتيجة مخرجات تراكمية للبرامج المدمجة . وفي الشكل أعلاه فإن الارتفاع الكلي (أي المحور الصادي) يمثل ثبوت ذلك الموقع ، اما ارتفاع كل حرف في الموقع فيشير الى تردد تلك الثمالة في عملية الاصطفاف للتواليات المدخلة والممثلة بالمحور السيني .

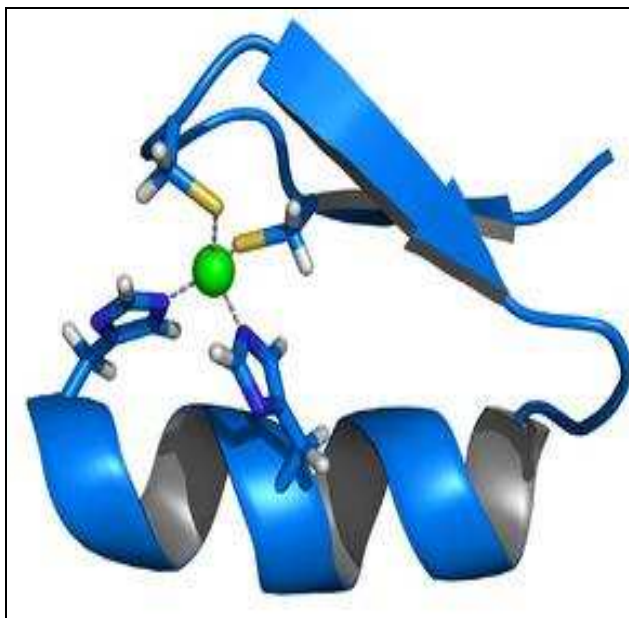
فالمواقع الثابتة تكون لها حروف كبيرة اما المواقع غير الثابتة فيكون لها خليط من الحروف الصغيرة مكدسة الواحدة على الأخرى، وهذا النظام يعطى صورة أوضح عن التواليات المتوافقة (المتطابقة) .

وتنتهج بعض القواعد المهمة في هذا المجال مثل SMART في إظهار مخرجات النتائج بالصور ، فمثلا" الشكل التالي (125.5) يوضح الدومينات وتفاصيلها في احد البروتينات :



شكل 125.5 : تمثيل قاعدة البيانات SMART للدومينات البروتينية

فضلاً" عن إمكانية إظهار الدومين بما يحتويه من تراكيب ثانوية كما موضح في الشكل التالي (126.5) الذي يظهر دومين أصبع الزنك



شكل 126.5 : التراكيب الثانوية لدومين أصبع الزنك

تحديد المناطق السطحية المتاحة للبروتين

تحديد المناطق السطحية المتاحة ASA تساعد في تحديد التراكيب البروتينية خاصة الثانوية . وتتم باستعمال خوارزمية الكرة المتدرجة Rolling ball algorithm وفيها يتم استعمال كريات من المذيب بنصف قطر محدد لقياس المساحة السطحية للجزيئات وهناك طرق أخرى لتحديد هذه الصفة ، وتختلف الطرق في دقتها وتعد القيمة 14 انكستروم هي القيمة المثالية وهي تقارب قطر او مساحة جزيئة الماء . وتستعمل هذه صفة قياس المساحة السطحية المتاحة لقياس انتقال الطاقة الحرة اللازمة لتحريك الجزيئة الحيوية من المذيب المائي الى المذيب اللاقطبي مثل البيئات الدهنية كما في الأغشية الخلوية . والمساحة السطحية لكل ثمانية معرصة للمذيب المحيط بالبروتين تكون مهمة في حدس المواقع الفعالة للبروتين وكذلك تحديد حواتم الأجسام المضادة ومعرفة التحويرات التي تجري على البروتينات بعد الترجمة. ان قيمة المساحة السطحية المتاحة تعتمد بشكل كبير على عدد من المؤشرات المستعملة في قياسها في التركيب ثلاثي الأبعاد وتتأثر بقوة فاندرفال وأهمها حجم الكرية . وقد صمم عدداً من البرامج لقياس الصفة منها يعتمد على الشبكات العصبية الصناعية ، وتدخل في البرامج المؤشرات الإحصائية بقوة لتحديدتها .

التحويرات الجارية على البروتينات

تجري على البروتينات عدداً من التحويرات مثل عمليات الطوي والتعلق وإضافة السكريات والفسفرة والاستئلة وغيرها، وبعض التحويرات التي تحصل بعد الترجمة مثل تكوين الجسور الكبريتيدية الثنائية التي تكون مهمة في إدانة ثبوت البروتينات والفسفرة او إضافة السكريات تعد جزءاً من التركيب الأولي للبروتينات ولا يمكن ان تقرأ من توالي الجينات المشفرة . ومن هذه التحويرات ينتج الإصدار الخام للبروتينات ، والإنزيمات التي تقوم بهذه العمليات يمكن ان

تحدد بسهولة ولكن الصعوبة تكمن في تحديد مواد الأساس التي تعمل عليها اذ لا توجد وسائل لهذه الحالة . وتعد البرامج التي تعالج هذه المسألة هي أكثر تقدماً بالنسبة للخلايا حقيقية النواة مقارنة ببداية النواة ولذا لا تستعمل في التحري عن الأخيرة ولذلك نشأ ما يسمى بالمعلوماتية الحيوية الميكروبية Microbial bioinformatics لمعالجة القضايا غير المشتركة بين نوعي الخلايا .

ان التعريف بالتحويرات يكون ضرورياً "لانه يؤثر مباشرة في تداخلات البروتينات مع بعضها وتوزيعها في الخلية في مواقع مختلفة وبالتالي التأثير في وظائفها فضلاً عن مساعدتها في التعرف على التركيب الثانوي والثلاثي للبروتينات . لذلك كان من الضروري استعمال وسائل المعلوماتية الحيوية للتعرف والتنبؤ بمواقع التحويرات على مناطق محددة من البروتينات .

ان التعريف بالتحويرات يكون ضرورياً "لانه يؤثر مباشرة في تداخلات البروتينات مع بعضها وتوزيعها في الخلية في مواقع مختلفة وبالتالي التأثير في وظائفها ، فضلاً عن مساعدتها في التعرف على التراكيب الثانوية والثلاثية ، ولذلك كان من الضروري استعمال وسائل المعلوماتية الحيوية للتعرف والتنبؤ بمواقع التحويرات في مناطق محددة من البروتينات .

وفي العموم يكون من الصعوبة تحديد مواقع التحوير وذلك لان التواليات القطيقات المعنية بالتحوير تكون قصيرة والذي يؤدي الى حصول العديد من النتائج الموجبة الكاذبة ، فمثلاً في قطيفة الفسفرة [RK] - X - [ST] المارة الذكر هي قطيفة قصيرة ويمكن ان توجد في مواقع متعددة في كل البروتينات تقريباً ، لذلك كان الاعتماد في عملية الحدس على هذه القطيفة لوحدها تكون خاطئة مؤدياً الى ارتفاع معدل النتائج الموجبة الكاذبة ، وحال مماثلة يمكن ان تحصل لأنواع أخرى من التحويرات ، واحد أهم عمليات الحدس الفاشلة تعود الى عدم الاكتراث بالبيئة المجاورة او التواليات المجاورة لموقع التحوير . ولأجل النقص من النتائج الكاذبة تستعمل الطرق الإحصائية مثل Support vector machine (SVM) التي يمكن ان تزيد من تخصص عمليات الحدس وهي عملية تصنيف للبيانات التي تشابه بعض العمليات الإحصائية الأخرى . وفي هذه الطريقة يتم تعريض البيانات الى ما يسمى بـ 3D- Space او الفراغ المتعدد الاتجاهات للحصول على دالة خطية او غير خطية تستعمل لإيجاد الإشارات الحقيقية من بين العديد من النقاط (Noise) . والبرامج المستعملة لهذه الخاصية يتم تدريبها على متغيرات بيئية مختلفة ونماذج تركيبية مختلفة وكافية لتصبح قادرة على تمييز عدداً من أنماط التحوير بعد الترجمة وأفضل عمليات التدريب تكون عند استعمال صنف محدد من القطيقات ووظائفها.

اما عمليات حدس الجسور الكبريتيدية الثنائية فتعتمد على معرفة قابلية الازدواج Pairing potential لحالات الارتباط لاثنتين من ثمالات السستئين في البروتين ، وتساعد في التعرف على التركيب الثانوي والثلاثي للبروتينات قيد الدراسة . ويمكن تحديد هذه القابلية من استعمال نسق يتم إنشاؤه من عمليات اصطافاف متعددة او من حساب قابليات الثمالات على التلامس Residue contact potentials بالاعتماد على البيئة الموضعية للتوالي ، وفي هذا المجال توجد برامج خاصة بالسستئين وارتباطاته اذ تحدد مواقع السستئين وبناء نسق محدد بالاعتماد على المعلومات الناتجة من عمليات اصطافاف متعددة ، ثم تقوم بترتيب ثمالات السستئين المؤهلة لتكوين الجسور الكبريتيدية المزدوجة .

وإضافة الى ما ذكر أعلاه فهناك برامج متخصصة للتعرف على إضافة السكريات التي تعتمد على الاختلاف في الكتلة بين ما يحدد للبيتيدات أي بصمة البيتيد مع القيم النظرية . ان الاختلاف في الكتل والذي يطلق عليه مبدأ الكتلة الإضافية يمكن ان يستعمل لحدس تحويلات أخرى وليس إضافة السكريات فقط ، وفيها يتم مقارنة كتلة البيتيد بالاستعلام مع القيم النظرية للبيتيدات المماثلة فإذا كان هناك اختلاف بالاعتماد على قياسات محددة مسبقاً فإنه يشير الى إحدى التحويلات وباستعمال طريقة المقارنة هذه أمكن التعرف على أكثر من 28 نوع من التحويلات مثل المثيلة ، الفسفرة ، إضافة الدهون Lipidation وإضافة الكبريتات Sulfation وغيرها .

وتوجد قواعد بيانات خاصة بتحويلات البروتينات بعد ترجمتها وتحتوي العديد من البروتينات المحورة ويمكن البحث فيها اما باستعمال الكلمات المفتاحية او الوزن الجزيئي او اسم البروتين او غيرها من الإمكانيات .

التنبؤ بالتحويلات الجارية على البروتينات الخاصة بالتوزيع

بعد تخليق البروتينات تذهب الى موقعها لأداء وظيفة معينة التي لا يمكنها أدائها في موقع آخر لذلك كان التعرف على مواقع البروتينات مهماً لتحديد الوظيفة وكذلك في تهميشها . فالبروتينات المخلفة يمكن ان يبقى بعضها في الساييتوبلازم والآخر يتوجه الى الأغشية ليكون في وضع خاص يلاءم وظيفته مثلاً استعراضه للغشاء او بروز جزءاً منه الى داخل الغشاء الخلوي او خارجه ، وبعض البروتينات تفرز الى الخارج .

وفي البكتريا السالبة لصبغة كرام الحاوية على غشائين فإن بعض البروتينات تتجمع في الفسحة المحيطة والبروتينات الغشائية التي لها نتؤات الى الخارج تمثل أهدافاً مهمة لتوليد الأجسام المضادة لها . وكل من هذه المواقع لها برامج خاصة لحدسها ، واغلبها يعتمد على توالي الحوامض الامينية في البروتين . وتتخصص البرامج أكثر في البحث عن أطراف البروتينات وخاصة النهاية الامينية ، أن غياب المناطق الكارهة للماء في البروتين يزيد من احتمالية بقاؤه في الساييتوبلازم أي انه من النوع الذائب . والعديد من البرامج المتوفرة لمعرفة موقع البروتين تنتهج توجهات مختلفة فضلاً عن الطرق المستعملة في الحدس منها استعمال الشبكة العصبية او غيرها من التوجهات .

حدس الإشارة والبروتينات المفروزة

الموضوع مكملًا لما ذكر أعلاه ويكون وثيقة العلاقة بدراسة مكنون الإفراز Secretome والذي يعني بالتعرف على كل البروتينات التي توجد على السطوح او المفروز من الكائن الحي .

وحدس الإشارات للبروتينات المفروزة يعتمد بشكل كبير على الدراسة المستفيضة للطرف الأميني للبروتين الناضج . فالبروتينات المرشحة للإفراز خارج الخلايا تكون حاوية على بيتيدات او قطيقات إشارة تزال من البروتين بعد إيصاله الى وجهته الأخيرة وبعض الأحيان يمكن ان تبقى ملازمة للبروتين . وتوجد أنواع من الإشارات التي تميز بواسطة البرامج ، ومن هذه الإشارات :

- بيتيد الإفراز Sec signal peptide

- الإشارات الحاوية على زوج من ثملات الارجنين Twin arginine signal peptide
- الإشارات الحاوية على الدهون Lipoprotein signal peptide .
- إشارات البكتريوسينات Bacteriocin signal peptide
- إشارات عامة غير تقليدية .

وهذه الإشارات لا يمكن تغطيتها ببرنامج واحد لذلك اختلفت البرامج في الأسس المستعملة ، فالبعض يهدف الى التحري عن طريقة الإفراز او مسار الإفراز او الطرق التقليدية للإفراز في الخلايا حقيقية النواة ، وهناك بعض البرامج التي طورت لحدس البروتينات التي تفرز بطرق غير تقليدية ، فضلا عن ان بعض البرامج تهدف الى تحديد موقع قطع ببتيدي الإشارة . ومثلما ذكر أعلاه فإن بعض طرق الحدس تحتاج الى وجود توالي مضبوط للطرف الأميني في البروتين السابق ومن ثم البروتين الناضج . وهذه يجب التأكد منها لان بعض تهميشات ألجين للنهاية 5' في بعض قواعد البيانات تكون غير صحيحة وبالتالي تؤدي الى تنبؤ غير صحيح. البرامج المستعملة في هذا المجال يمكن ان تحدد موقع الإشارة داخل الخلية باستعمال برامج الشبكات العصبية او موديلات ماركوف الخفية . وبرامج الشبكة العصبية تجمع اثنين من الدرجات الأولى لتمييز ببتيدي الإشارة والثانية لمواقع فلق الإشارة من البروتين بواسطة البروتينات ، وتتأثر هذه بالتغيرات الكبيرة في الإشارات من حيث الطول ومكوناتها من الحوامض الامينية مما يجعل حدسها مهمة صعبة .

اما المعتمدة على موديلات ماركوف فتميز ببتيديات الإشارة يعتمد على الطرف الأميني اللازم لإقحام او إدخال البروتين الى الغشاء ، وهذه البرامج مدربة على ثلاثة أنواع من مجاميع المعلومات وهي الخاصة بالبكتريا الموجبة لصبغة كرام والبكتريا السالبة لصبغة كرام والخلايا حقيقية النواة ، وهذا التمييز ضروري بوجود اختلافات كبيرة في ببتيديات الإشارة بين هذه المجاميع من الأحياء ، والبرامج تحدد توالي الإشارة وموقع الفلق في توالي الاستعلام .

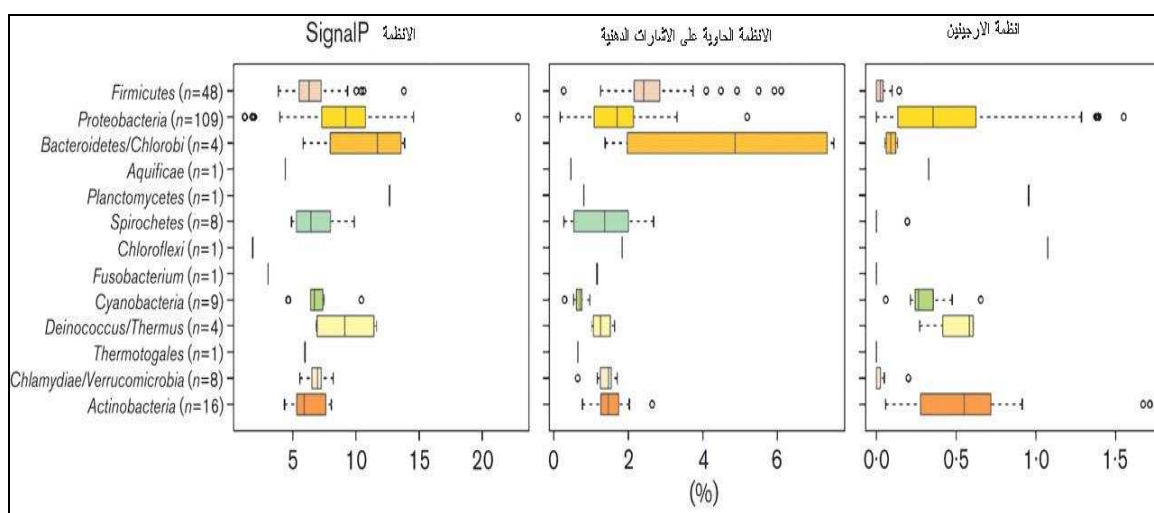
- ولكن مع التطورات العديدة في البرامج وتطور أنظمة الحاسوب ظهرت برامج جيدة منها
- برامج تعتمد على الإشارات وذلك من معرفة الشحنة وكرائية البروتين للماء او وجود القطيقات المتوافقة

- المجموعة الثانية من البرامج هي التي تعتمد على النواحي الإحصائية للتواليات مثل مكوناتها من الحوامض الامينية

- المجموعة الأخرى التي تعتمد الى التهجين بين الاثنين وهي البرامج الانجح كما في البرامج المعتمدة الشبكات العصبية وموديلات ماركوف الخفية . وتصل دقة هذه البرامج الى 65 - 70 % . وتعتمد إحصائيات خاصة بكل برنامج . اما مخرجاتها فيمكن ان تكون ملفات نصية او تمثيل بالرسوم . وفي اغلب الأحيان لا تكون النتيجة قاطعة بنعم ا و لا لذلك تحتاج الى معلومات سائدة للتقرير النهائي ويستحسن في هذه الحالة استعمال أكثر من برنامج وإجراء المقارنة واستخلاص النتائج

البروتينات المفروزة : وهي التي تترك الخلية تماما الى البيئة المحيطة وتكون لها أهمية طبية مثل السموم وفي مجالات الصيدلة وصناعات التقنيات الحيوية وتحديدها يقع ضمن عمليات حدس البروتينات المفروزة .

وفي مجال دراسات البكتيريا تكون عمليات الإفراز في الأجناس الموجبة لصبغة كرام أسهل من تلك الخاصة بالبكتيريا السالبة لصبغة كرام ويوجد العديد من مواقع الانتزيت التي تتعامل مع هذا الموضوع . وقد أمكن باستعمال وسائل المعلوماتية الحيوية تحديد نسب البروتينات المفروزة من قبل الأحياء المختلفة وخاصة بالنسبة للبروتينات التي تحمل ببتيدات الإشارة (Sec-) المذكورة سابقا" وتستعمل بكثرة من قبل الأحياء (البكتيريا) وتليها البروتينات الحاوية على الدهون (Lip-) ، ثم تليها إشارات توائم الارجنين (Tat) أي الحاوية على اثنين من ثمالات الارجنين كما تم حدها بالبرامج المختلفة والموضحة في الشكل 127.5 :



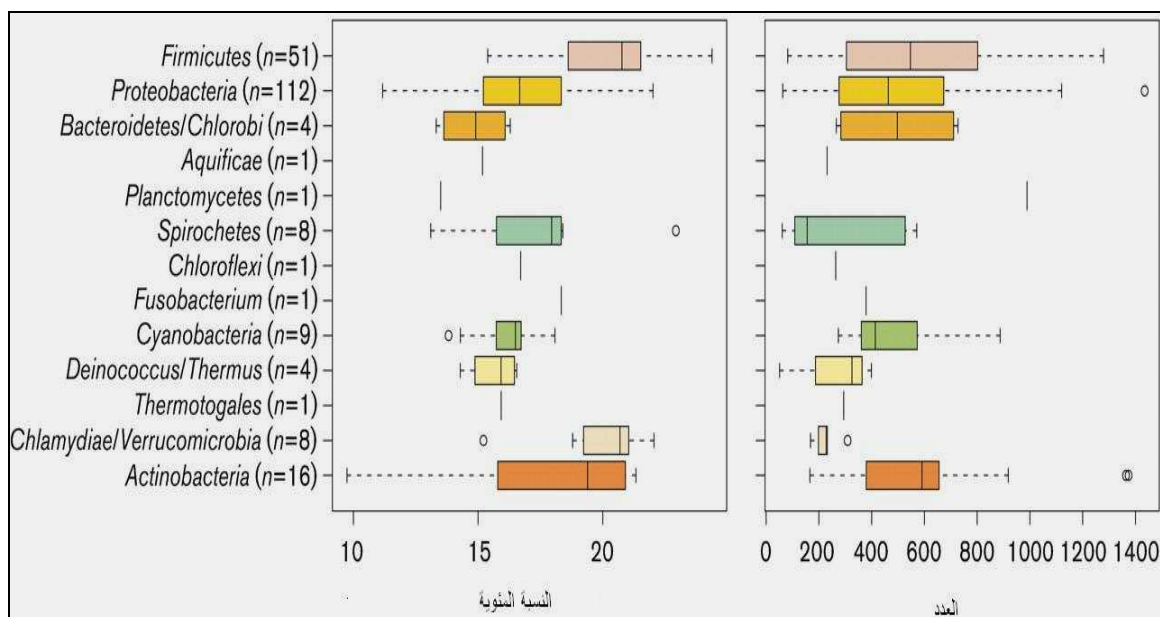
شكل 127.5 : أنواع الإشارات الببتيدية الموجودة في الخلايا بدائية النواة

والملاحظ من الشكل ان كل من Bacteroides و Proteobacteria تفرز قدرا كبيرا من مكونات البروتينات وان كان هناك تباين داخل المجموعة الواحدة ، وكما موضح الشكل ان Bacteroides تستعمل نظام Lip للبروتينات الدهنية بكثرة ، اذ تنتج كميات من البروتينات الدهنية التي هي في اغلب الأحيان من اللاصقات Adhesins . ونسبة إفراز البروتينات العالية يكون لها علاقة وثيقة بالبيئة التي تعيش فيها البكتيريا او غيرها من الأحياء وكذلك لها علاقة بحجم الجينوم والبيئة المحيطة . فالأحياء التي تنطف على المضيف تحتاج الى إفراز البروتينات لتحليل المواد الرابطة في أنسجة المضيف وتصل النسبة الى 2 % من مكوناتها البروتينية ، في حين ان الأحياء التي تنطف داخل الخلايا فهي تحتاج الى نسبة قليلة من البروتينات المفروزة .

اما في البكتيريا المفترسة *Bdellovibrio bacteriovorus* السالبة لصبغة كرام والتي تفرس البكتيريا الأخرى فهي تحتاج عدد كبير وكميات كبيرة لتفكيك والتهام البروتينات الغريبة . وقد وجد ان 23% من بروتينات هذه البكتيريا يحوي على إشارة الإفراز Sec signal وتحوي على 59% إشارات الدهون . وفي العموم فإن البروتينات الحاوية على إشارات الارجنين تكون اقل من البروتينات الحاوية على إشارة الإفراز او الإشارات الدهنية .

التعرف على البروتينات الغشائية

تشكل البروتينات الغشائية حوالي 30 % من بروتينات الخلايا وتكون مسئولة عن أداء عدد من الوظائف ، منها نقل الإشارات ونقل المواد عبر الأغشية وأداء تحولات الطاقة الحيوية وتختلف نسب البروتينات في الأحياء بدائية النواة بشكل كبير كما موضح من الشكل 128.5 :



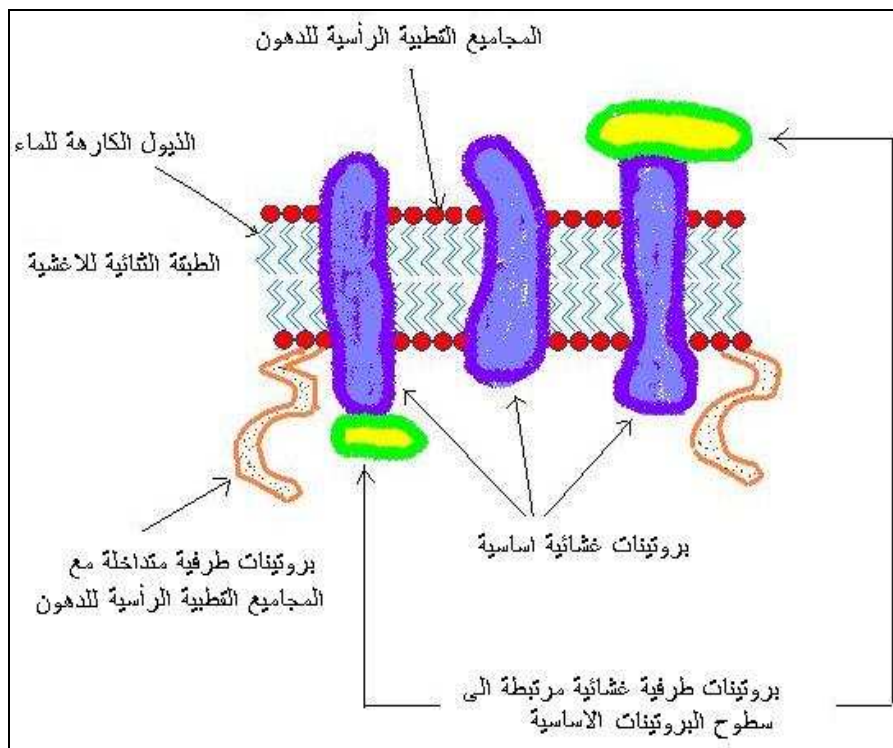
شكل 128.5 : اختلاف عدد وكمية البروتينات في مجاميع من الأحياء بدائية النواة

وتصل نسبة البروتينات في معظم الشعب من 15 - 20 % ، والأحياء الحرة المعيشة تحوي على أكثر من 400 - 500 بروتين يستعرض الأغشية ، وتشكل هذه البروتينات أهدافاً جيدة للأدوية ، وهناك بعض التراكييب البروتينية التي يمكن أن توجد في الأغشية منها

• حلزونات ألفا

• اسطوانات بيتا β - barrels

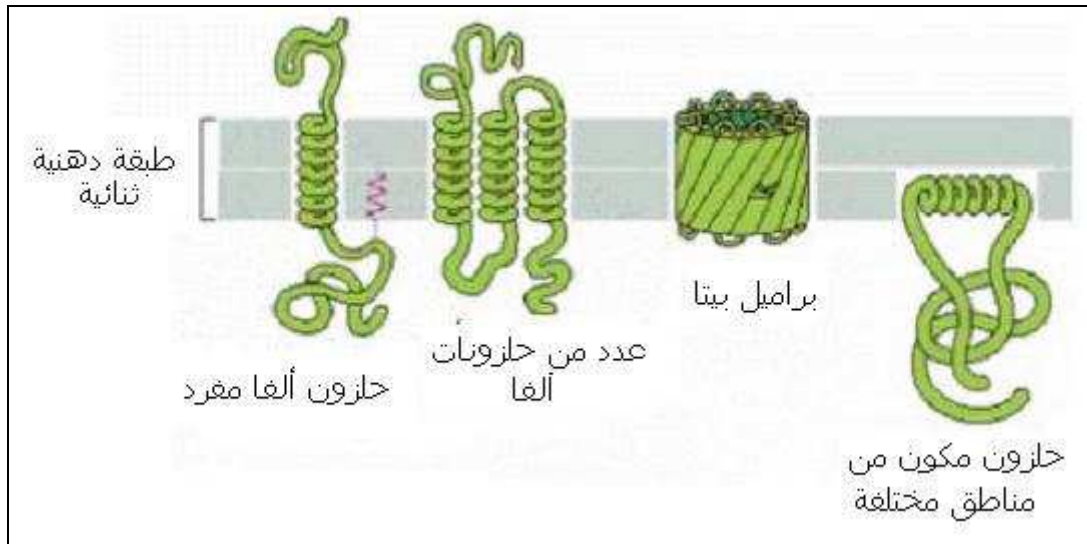
ومعظم البروتينات الغشائية تحوي على حلزون ألفا واحد يستعرض الغشاء والقليل منها تحوي على أشربة بيتا مكونة اسطوانات بيتا المكونة من صفائح بيتا متعكسة التوازي مثل تلك الموجودة في الأغشية الخارجية للبكتريا السالبة لصبغة كرام . وتختلف البروتينات في وضعيتها داخل الأغشية فالبعض يستعرضها تماماً وفيها تكون الطبقة الخارجية كارهة للماء لضمان بقاءها داخل الغشاء والأخرى تتخذ وضعيات أخرى لعلاقتها بالأغشية . والشكل التالي (129.5) يوضح بعض البروتينات الأساسية Integral المستعرضة للأغشية .



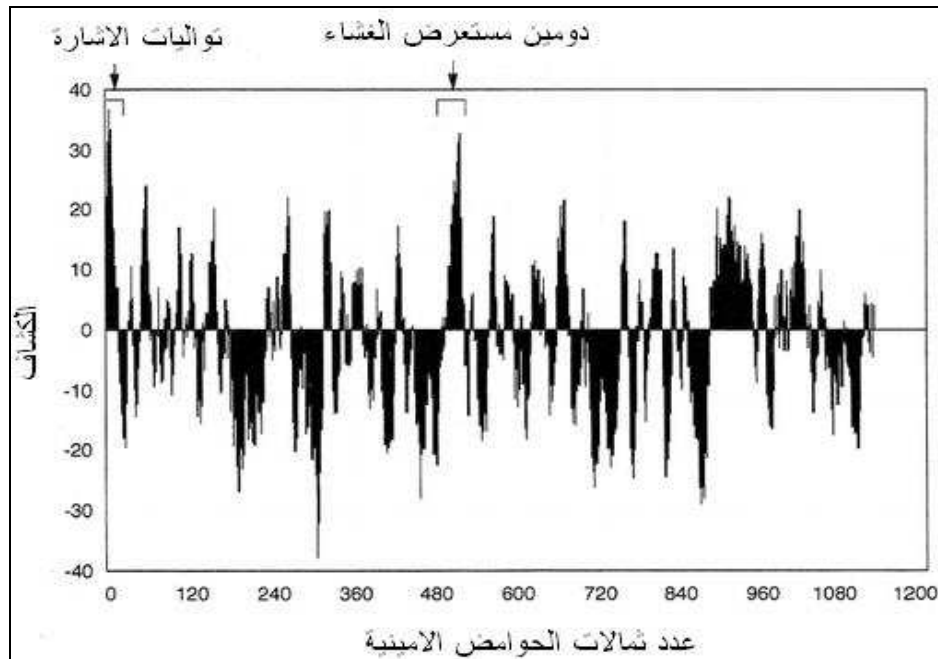
شكل 129.5 : بعض البروتينات الأساسية المستعرضة للأغشية

طرق حدس البروتينات الغشائية

ان التعرف على تركيب البروتينات الغشائية باستعمال الأشعة السينية او الرنين النووي المغناطيسي يكون صعباً لذلك يتم اللجوء الى طرق المعلوماتية الحيوية وتختلف الطرق المستعملة لحدس البروتينات الغشائية ، اذ ان البعض يهدف للتعرف الى حلزونات ألفا والأخرى تتعامل مع تراكيب بيتا وأخرى تتعامل مع قطع التواصل بين عنصري التركيب الثانوي الأساسية . وتتعامل البرامج مع طوبوغرافية البروتين في الغشاء ، ومن أقدم الطرق والتي لا تزال قائمة هو الحدس بالاعتماد على كراهية البروتينات وثمالاته للماء التي تعد مقياساً لإمكانية بقاء البروتين منظماً في الأغشية ، لذلك تربط البرامج مع مقياس الكراهية ويجري توزيعها اعتماداً على هذا المقياس والشكلان التاليان (130.5 و 131.5) يوضحان توزيع البروتين ونسق الكراهية في بروتين مكون من 1125 حامض أميني .



شكل 130.5 : انطمار بعض التراكيب البروتينية في الأغشية الخلوية



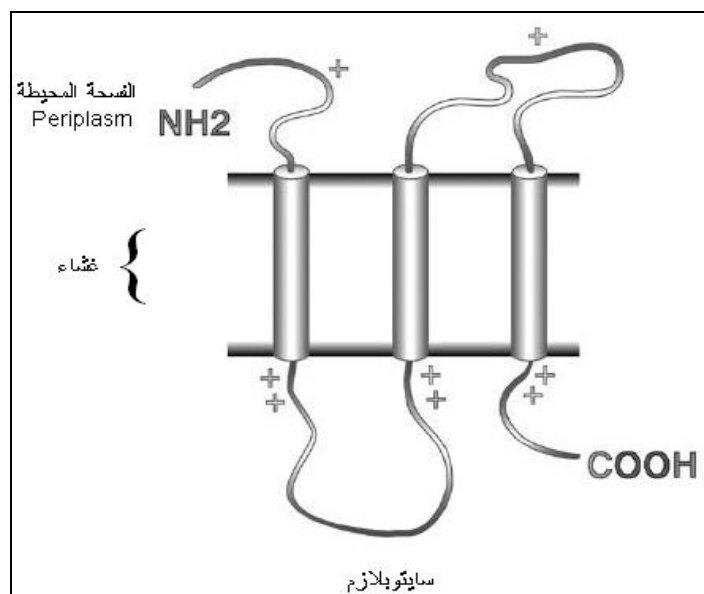
شكل 131.5 : نسق الكراهية للماء لبروتين مكون من 1125 حامض اميني

فالقيم فوق المستوى صفر تمثل الثمالات الكارهة للماء والأخرى تمثل الثمالات المحبة للماء. وتصيح مسألة تحديد التركيب أسهل عند اعتماد صفات الكراهية للماء بعد الأخذ بنظر الاعتبار البيئة الكارهة للماء داخل الأغشية والتي تسمح بوجود قطع البروتين كارهة للماء فقط. والأساسيات المطورة لحس تراكيب البروتينات الذائبة يمكن ان يطبق بعضها على البروتينات الغشائية ولكن البعض لا يعمل بشكل صحيح نظرا لارتفاع او قوة صفة كراهية الماء لذلك فان

الصفة الأخيرة تدخل في البرامج الخاصة بالبروتينات المستعرضة للغشاء فقط ، إذ ان هذه القطع تتمثل بالثملات غير القطبية والحوامض الامينية الحلقية وتكون صفتها العامة محبة الدهون. ويظهر الشكل أعلاه (131.5) توزيع بعض البروتينات في الأغشية:

حدس الحلزونات

حلزونات ألفا المستعرضة للأغشية هي عادة كارهة للماء وفيها توزيع خاص للثملات الموجبة الشحنة ، وتكون الحلزونات عمودية على المستوى الأفقي للغشاء ويتراوح طولها بين 17 - 25 ثمالة ، وتفصل هذه بمناطق او عروات محبة للماء يمكن ان يصل طولها الى 60 ثمالة وهي مناطق موجبة الشحنة بشكل اكبر، ومستوى الشحنات هذا يختلف ففي المناطق القريبة من الجانب السايئوبلازمي قرب الجزء الكارهة للماء تكون مشحونة بشحنة موجبة مقارنة بالجانب الذي يظهر في الفسحة المحيطة كما في البكتريا السالبة لصبغة كرام (شكل 132.5)

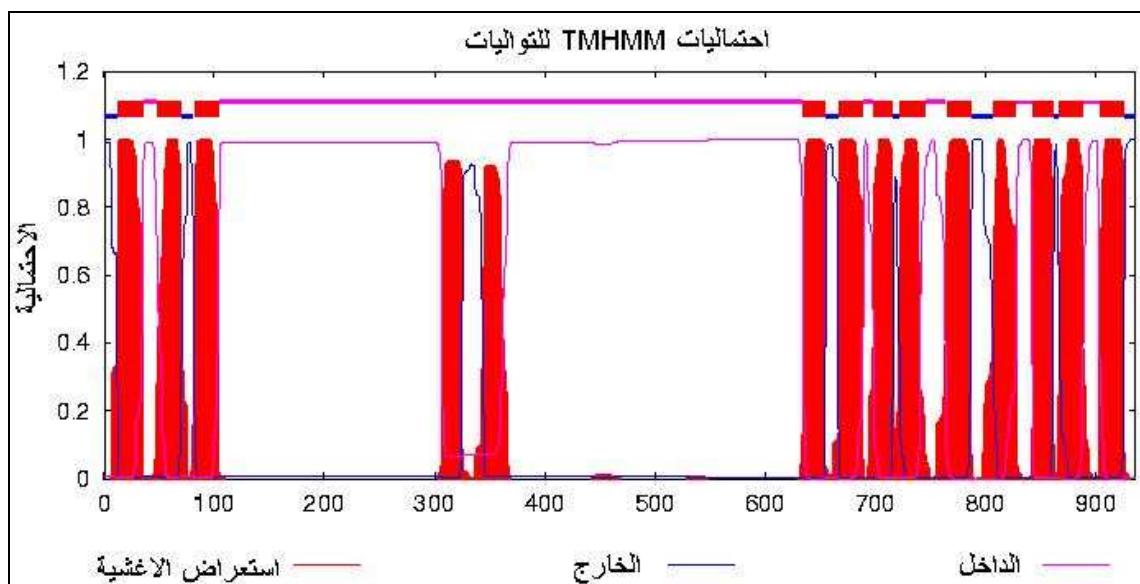


شكل 132.5 : توزيع الشحنات على البروتينات الغشائية في البكتريا السالبة لصبغة كرام

وهذا الوضع يسمح بالتنبؤ بعناصر تركيب البروتين اذ ان مثل هذه الصفات المذكورة تشكل الأساسيات المستعملة في حدس البروتينات الغشائية . وهناك عدد من الخوارزميات الموجودة لمعرفة الحلزونات المستعرضة للأغشية والتي تعتمد على كراهية الماء وتستعمل نافذة مساوية لعدد الثملات الكارهة للماء (مثلا 17 - 25) وتحدد فيما إذا كان البروتين يستعرض الغشاء ام لا اعتمادا" على درجات الكراهية للماء المسجلة ، في حين ان برامج أخرى تعتمد على مناطق الشحنات الموجبة المذكورة أنفا" .

وعلى العموم فإن الحدس المعتمد على كراهية الماء تكون فيه معدلات الخطأ عالية وذلك لان ببتيديات الإشارة الكارهة للماء يمكن ان تشوش على دقة الحدس ولذا تعتمد بعض البرامج الى التمييز بين الحلزونات الغشائية وببتيديات الإشارة الكارهة للماء وتستبعد الأخيرة من عملية

الحدس . وبأقران هذه العمليات مع المعلومات الخاصة بالتطور وبمساعدة الشبكات العصبية وموديلات ماركوف الخفية أمكن تحسين دقة البرامج لتصل الى 70% .
ومن أهم البرامج المستعملة في هذا المضمار Transmembrane Hidden Markov Model (TMHMM) والبرنامج مدرب لتمييز التركيب الحلزوني للبروتينات التي تستعرض الأغشية باستعمال عدد كبير من الحلزونات المستعرضة للأغشية المعروفة . وعند عمل البرنامج فإنه يقوم بمسح توالي الاستعلام ويعطى احتمالية وجود دومينات حلزونات ألفا ، وترتيب البروتين يعتمد على أساس توزيع الشحن الموجبة ، ومخرجات البرنامج تمثل حلزونات ألفا وحدودها فضلا عن إعطاء تمثيل صوري له كما موضح في الشكل 133.5 :

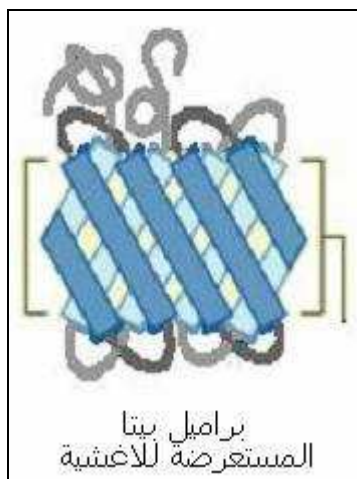


شكل 133.5 : التنبؤ بأحد بروتينات *Helicobacter pylori* الغشائية الأساسية

والبرامج الأخرى تأخذ بنظر الاعتبار وجود ببتيديات الإشارة الكارهة للماء والبروتينات الغشائية وتميز بينها للتقليل من النتائج الموجبة الكاذبة ثم بعد ذلك تعمل على التوالي او الجزء الغشائي وهي تعطي نتائج أفضل 94% مقارنة بالبرنامج أعلاه TMHMM الذي تصل دقته الى 70% .

حدس اسطوانات بيتا في الأغشية

هذه الأجزاء اغلبها مزدوجة الصفات Amphipathic وتحتوي على حوالي 10-22 ثمانية ويكون ترتيب الثمالات الكارهة للماء بشكل خاص ، فالثمالة الثانية كارهة للماء وكذا مع الثمالات زوجية الموقع وتواجه الطبقة الدهنية من الغشاء والباقي تقابل تقوب الاسطوانات وتكون محبة للماء ، وتعطي بعض البرامج الحدس بشكل مصور كما موضح في الشكل 134.5 :



شكل 134.5 : توزيع الثملات في اسطوانات بيتا المستعرضة للأغشية

والبحث عن كراهية الماء في تحديد أجزاء البروتين في الأغشية لهذه الأجزاء (اسطوانات بيتا) لا يعطي نتيجة ، لذا لا تستعمل معها برامج حدس حلزونات ألفا . وطور لذلك عدد قليل من البرامج تعتمد الشبكات العصبية والتقنيات المماثلة بعد تدريبها على تراكيب مشابهة .

حدس المكونات البروتينية الأخرى

هناك العديد من الجوانب والبروتينات الخلوية التي كان للمعلوماتية الحيوية الحضور الواضح في التعامل معها ، وهي جوانب واسعة لا يمكن إدراجها في هذا الكتاب العام ، ولكن لا بأس من ذكر بعضها . المجال المناعي ويدخل ضمن Immunological bioinformatics يعطى جوانب واسعة من الجوانب المناعية التي ترفد الجوانب الصحية بالمعلومات وهناك العديد من مواقع الانترنت الخاصة بهذا ، وتوجد العديد من البرامج التي تهتم بتداخلات الأجسام المضادة مع المستضدات وتحديد الحواتم في المستضدات والجوانب الخاصة بجزيئات MHC وصور جيناتها الكثيرة والمختلفة فضلا عن التعامل مع الخلايا المكونة للجهاز المناعي التي سيمر ذكرها لاحقا . ومن الجوانب الأخرى التي لم تغط في هذا الكتاب هو تدخل المعلوماتية الحيوية مع الأنظمة الثنائية المكونات التي تنتشر في مجاميع البكتيريا وبعضها يوجد في الخلايا حقيقية النواة . واهم وظائف هذه الأنظمة نقل الإشارات والتي تختلف أعدادها في الأحياء المختلفة اعتمادا على البيئة .

اهتمت المعلوماتية الحيوية بهذه الأنظمة وطلورت قواعد البيانات وبرامجها الملحقة للتعامل مع هذه الأنظمة .

تصوير وتمثيل تركيب البروتينات

توفر العديد من البرامج التمثيل الصوري للبروتينات الكاملة التي تحدد بالطرق العامة مثل الترحيل الكهربائي أو غيرها من الطرق والتي توثق بالصور . وهناك العديد من البرامج لتحليل الصور ومعالجتها وإجراء المقارنات عليها بالاعتماد على الفحوص الإحصائية . لذلك يمكن تحديد نوعية البروتينات في الأنظمة الحية في حالة الراحة التي تعد ظروف سيطرة ، فضلا عن مقارنة المكون البروتيني لحالة الصحة والمرض من الصور الناتجة . وفي حالات

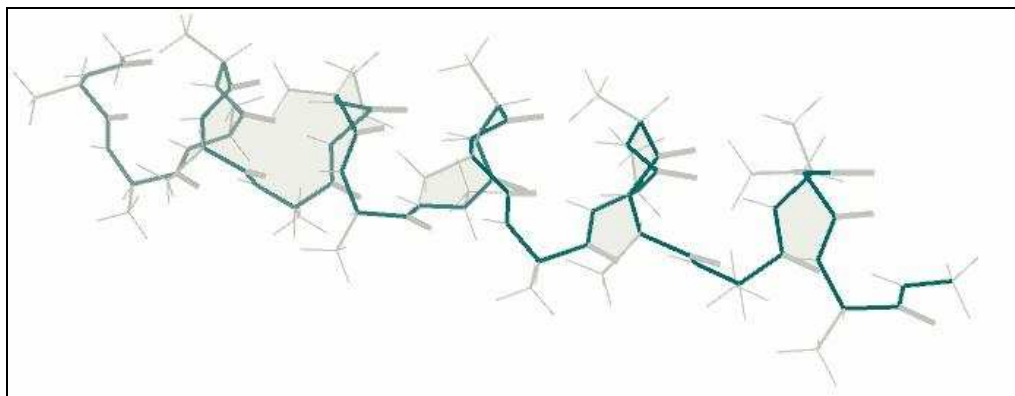
قواعد البيانات التي تعتمد هذه البرامج يمكن البحث فيها باستعمال الكلمات المفتاحية أو كاشف البقع Spot identifier .

أما على المستوى الجزيئي للبروتينات فعندما يتم تحليل ومعرفة تركيب البروتينات يمثل تركيبها بصور ثلاثية الأبعاد وقد كان التمثيل قبل ظهور برامج الحاسوب يمثل بموديلات فيزيائية مثل وجود الأسلاك والعصي والكرات ولكن بتطور برامج الحاسوب أصبح بالإمكان تصوير البروتين والتلاعب بالتركيب ثلاثية الأبعاد المعقدة وهذه الرسوم ساعدت في تحليل ومقارنة التركيب البروتينية لمعرفة الوظائف .

وأغلب برامج الحاسوب لها قدر من التفاعلية Interactivity التي تظهر على الواجهة ويستطيع المستخدم ملاحظتها وتحويلها والتلاعب بها .

وأكثر البروتينات المودعة في بنك بيانات البروتين PDB تكون للبروتينات فيها ثلاثة أبعاد , Z , Y X, للذرات وتظهر باستعمال برامج مختلفة بأشكال مختلفة منها :

- الشكل السلكي أو ما يسمى بالقالب Wireframe كما موضح في الشكل 135.5 :

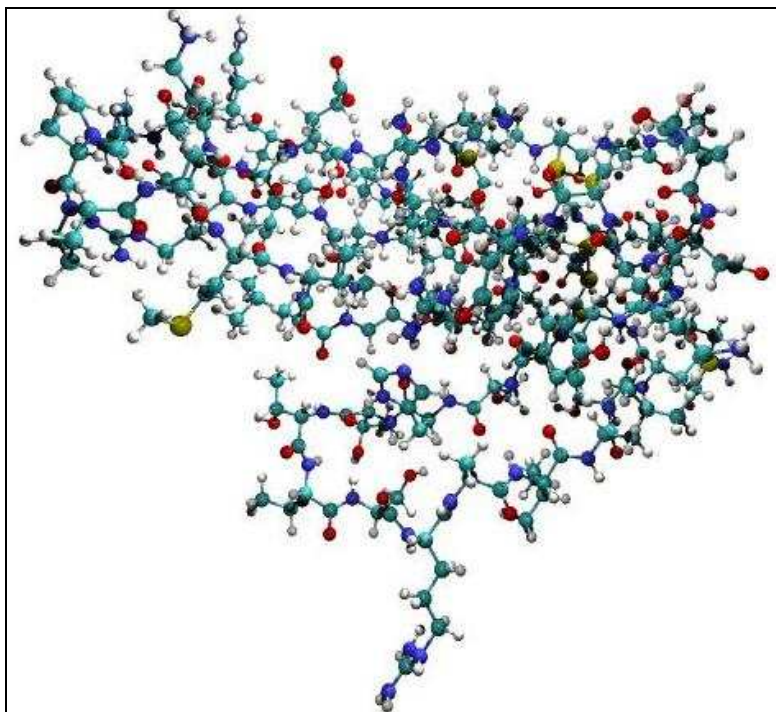


شكل 135.5 : التمثيل السلكي للبروتينات

يعتمد الرسم على تمثيل الخطوط التي توضح الأواصر بين مكونات البروتين ويعد أبسط طرق تمثيل التركيب ويساعد في إيجاد المواقع للثمالات في التركيب .

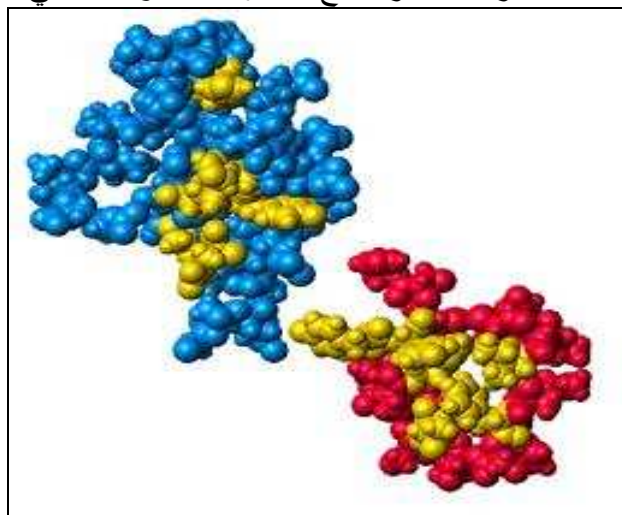
- تركيب العيدان والكرات Ball and stick structure

وفيه تظهر تراكيب البروتينات كما في الشكل 136.5 :



شكل 136.5 : تركيب العيدان والكرات لتمثيل البروتينات

- وتمثل العصي الأواصر بين الذرات ، اما الكرات فتمثل الذرات
- طريقة رسم ملئ الفراغ Space – filling structure وفيه تمثل البروتينات بشكل كرات متقاربة مع بعضها كما موضحة في الشكل 137.5 :

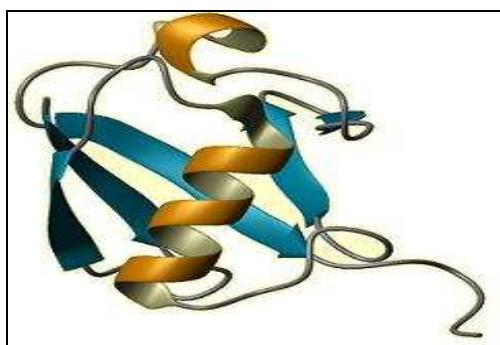


شكل 137.5 : طريقة ملئ الفراغ لإظهار تركيب البروتينات

وتمثل الذرات المختلفة بكرات مختلفة الألوان التي تكون أنصاف أقطارها تقابل أنصاف قوى فاندرفال للذرات . ويسمى التركيب أيضا "CPK" (من أسماء الواضعين للتركيب , Corey Pauling & Kohan)

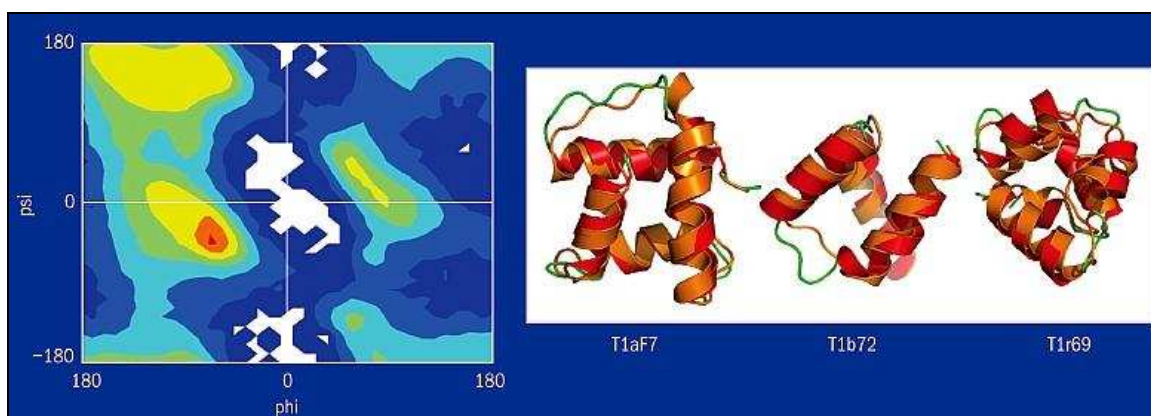
• التركيب الشريطي Ribbon Structure

وفيه تستعمل الأشرطة أو الخطوط لتمثيل الحلزونات واسهم عريضة لتمثيل صفائح بيتا كما موضح في الشكل 138.5 :



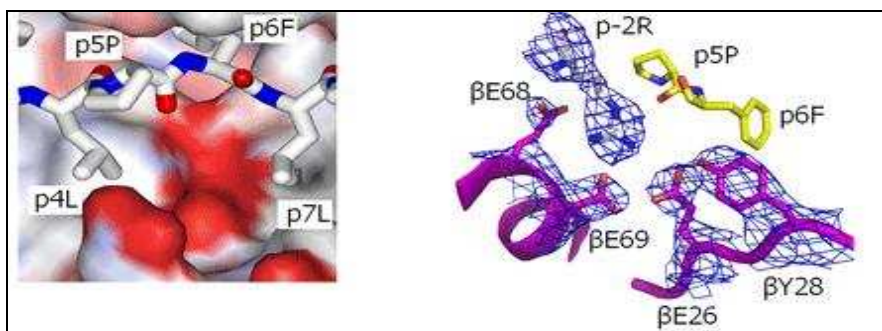
شكل 138.5 : الشكل الشريطي لتركيب البروتينات

ومثل هذه الرسوم تشد الانتباه لأنها تسمح بالكشف وتحديد عناصر التركيب الثانوي . وهناك أنظمة تلوين مختلفة تحدد من قبل المبرمجين . وفي حالة التعامل مع الحوامض الامينية يكون في الغالب التعامل معها اعتمادا" على الألوان المحددة بالصفات الفيزيوكيميائية للحوامض (والتي تم التطرق اليها في فصل سابق) . وتختلف البرامج في أدائها لرسم البروتينات فبعض البرامج يمكن ان ترسم كل أنواع التراكييب المذكورة ويمكن ان تقرأ كل الصيغ الموجودة في بنك بيانات البروتين وغيرها ، ويمكن ان تظهر البروتين كاملا" او بشكل جزئي وفق الحاجة . والبرامج تختلف في قوة تمثيلها للبروتينات ، فبعض البرامج تدخل في حسابها قياس الزوايا مثل ϕ و ψ التي تظهر مرافقة للرسوم كما موضح في الشكل التالي (139.5) الذي يمثل مخطط راماجاندران



شكل 139.5 : إظهار الزوايا ϕ و ψ ضمن مخطط راماجاندران عند التنبؤ بالبروتينات

كما ان بعض البرامج يمكن ان تقيس المساحة السطحية المتاحة للمذيبات فضلا عن إمكانية إظهار السحابات الالكترونية ومكوناتها من الكهربائية المستقرة كما موضح في الشكل 140.5 :



شكل 140.5 : إظهار قيم المساحات السطحية والسحابات الالكترونية عند التنبؤ بالبروتينات

تقييم مقارنة تراكيب البروتينات

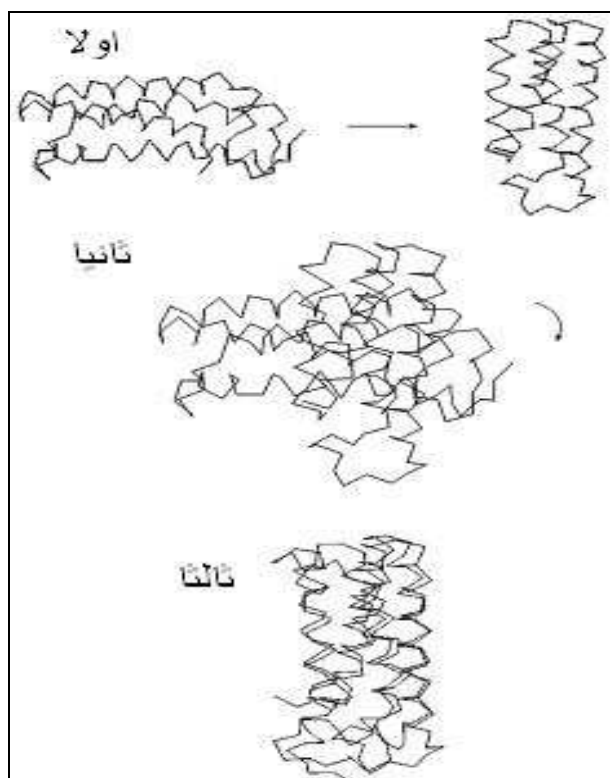
بوجود الوسائل والبرامج التي تظهر تراكيب البروتينات أصبح من الممكن مقارنة تركيب البروتينات لإظهار التشابه . والتحليل المقارن في العادة وليس دائما" يشمل الاصطفاف المباشر وإظهار التراكيب الثلاثية الأبعاد ، وإظهار الجزء الثابت من التركيب على مستوى التركيب الثلاثي . وهذه المقارنات تعد من التقنيات الرئيسة لتحليل تركيب البروتين لإيجاد التماثل الجذري البعيد Remote protein homology ، وذلك لان التراكيب البروتينية لها درجة عالية من الثبوت مقارنة بالتواليات وعليه فإن البروتينات تشترك في التراكيب حتى ان لم يكن لها تشابها" كبيرا" على مستوى التواليات . ولذلك فإن المقارنات يمكن ان تظهر المسافة التطورية او العلاقة التطورية بين البروتينات التي ليس من الملائم الاعتماد على مقارنة التواليات فقط .

واستعمال الرسوم للمقارنات تساعد لتقييم وتصنيف البروتينات الى مجاميع طوي Fold classes . كما ان الرسوم تساعد في تقييم طرق حدس البروتينات وذلك بمقارنة التوقعات النظرية مع تلك المحددة عن طريق التجارب . ومثل هذه المقارنات يمكن ان تتم عمليا" بالنظر وهي الطريقة المستعملة ولكن الأفضل استعمال الحاسوب لغرض الحصول على نتائج أكثر دقة ، وفي هذا المجال تتبنى البرامج المستعملة نظام احتساب الدرجات لقياس تشابه التراكيب بعد تضخيم نظام الدرجات باستعمال مؤشرات مختلفة والخوارزميات المستعملة يمكن ان تقسم الى عدة أقسام

- خوارزميات تقيس تطابق التراكيب البروتينية وذلك بتقليل المسافات بين الجزيئات
- خوارزميات تقيس المسافات داخل الجزيئة للتراكيب .
- خوارزميات تشمل التوجهات في استعمال الموصافات الداخلية للجزيئات وكذلك الموصافات الخارجية .

وفي هذا الموضع فإن الطرق المعتمدة على الموصافات الداخلية تستعمل عادة للتراكيب المتشابهة نسبيا" . وفيها يحرك احد التراكيب على الحاسبة مقابل تركيب ثاني الى حين

الحصول على أكبر تداخل للفراغ الثلاثي الأبعاد ، فبعد تحديد الشمالات أو الذرات المتكافئة وبعد الحصول على تكامل للشمالات عندها يزاح أو يحرك احد التراكيب جانبيا وعموديا" مقابل التركيب الآخر الى حين الحصول على الترتيبين في الموقع نفسه كما موضح في الشكل : 141.5



شكل 141.5: خطوات مطابقة تراكيب البروتينات

اي يستمر التدوير الى حين الوصول الى اقصر المسافات بين الجزيئات التي يصل عندها التطابق بين البروتينين الى أقصاها والتي بعدها تدخل في عمليات حسابية للتأكد منها ومقارنتها بقيم إحصائية معدة لهذا الغرض . أي ان الطريقة تعتمد التداخل بين صفات الجزيئات (Intermolecular method) .

وفي حالة مقارنة البروتينات البعيدة نسبيا" فتنتم اما بحذف التواليات التي تظهر تغيرات خارج عناصر التراكيب الثانوية أو تقسيم البروتينات الى قطع صغيرة مثل 6 - 9 ثمالات ، ثم تتم المطابقة للمناطق المتشابهة على المستوى الثلاثي الأبعاد قطعة قطعة ، ثم إيجاد القطع المتوائمة جدا" . الخوارزميات التي تستعمل المواصفات الداخلية بين الجزيئات تسمى ايضا" بالأمثلة المتكررة Iterative optimization وتشمل الخطوات الآتية

- صف التواليات باستعمال البرمجة الديناميكية
- تحديد الشمالات المتماثلة من الجولة الأولى من عملية الاصطفاف

• مطابقة الثمالات في المناطق القريبة (للثمالات الناتجة من الخطوة السابقة) على مستوى الأبعاد الثلاثية لتكون ما يعرف الثمالات المتكافئة .
واعتمادا" على الثمالات المتكافئة او المتطابقة الجديدة تبدأ جولة جديدة من التطابق لإجراء اصطفاف جديد ، وتعاد الطريقة الى حين الوصول الى حد اكبر من التطابق .
اما الطرق المعتمدة على الموصافات الداخلية Intramolecular للجزيئات فهي تعتمد على عمليات إحصائية للتركييب ولا تعتمد على تشابه التواليات بين البروتينات المراد مقارنتها ، ولا تظهر تمثيل فيزيائي للتركييب ولكن تزود بتقييم كمي لتشابه التراكيب بين أزواج الثمالات المعنية . والطريقة تعمد الى إيجاد مصفوفات للمسافات بين الثمالات لكل من البروتينين ثم تحرك هذه المصفوفات الواحدة بالنسبة للأخرى الى حين الوصول الى التداخل الأكبر او الأعظم وبذلك يمكن تحديد مجاميع مناطق التشابه ، ولتسهيل المهمة يتم تقسيم المصفوفات الى مصفوفات صغيرة مكونة من قطع من ستة بيتيدات .
وطرق المقارنة الحديثة تعتمد توجهات استعمال الطرق الهجينة إضافة الى استعمال موصافات تركيبية أخرى مثل نوع التركيب الثانوي وزاوية الانفتال Torsion angle والإتاحة للماء وأواصر الهيدروجينية الموضعية . وتستعمل البرمجة الديناميكية لزيادة التداخل لمؤشرات التوجهات الداخلية والخارجية .

مقارنة البروتينات المتعددة

بالإضافة الى إجراء عمليات الاصطفافات المزدوجة ، فإن عمليات المقارنة يمكن ان تستعمل لمقارنة اصطفافات لتركييب متعددة ، والاستراتيجيات المستعملة تماثل الطرق المستعملة في الاصطفافات المزدوجة ، وتتم مقارنة التراكيب كأزواج لإيجاد مصفوفات المسافات اعتمادا" على تشابه التركيب وتسجيل الدرجات لها ومن ثم تستعمل في إنشاء شجرة للعلاقات التطورية ثم ترتيب التراكيب في عناقيد ، والتراكيب الأكثر تشابها" يعاد صفها من جديد التي تؤدي الى الحصول على تراكيب وسطية تسمح بإضافة تراكيب أخرى بشكل متتالي للمقارنة ، وعند إضافة كل مجموعة من التراكيب فانه يؤدي في النهاية الى إيجاد عملية اصطفاف تراكيب متعددة . وهناك برامج تقوم بمقارنة التراكيب بالاعتماد على موصافات التواليات الداخلية (المذكورة أعلاه) وتضخيم التشابه لرسم البروتينات بالاعتماد على مصفوفات المسافات بين كل ذرات α لكل بروتين . وتوضع المصفوفتان لكل بروتين على بعضها وتحرك الواحدة بالنسبة للأخرى لإيجاد أكثر المناطق تشابها" . وتدخل القيم الإحصائية لتقييم نتائجها مثل استعمال Z- score التي وجد بالتجربة ان القيم لهذه الدرجات ان كانت اكبر من 4 فانه يشير الى مستوى عال من التشابه.

نتائج التمثيل والمطابقات

الطرق المذكورة أعلاه وما أسفرت عنه من مقارنة التراكيب المختلفة يمكن ان تستعمل في تصنيف البروتينات وإيضاح العلاقات بين تراكيب البروتينات . واستعمالها أدى الى وضع نظام هرمي لتصنيف البروتينات لإعطاء صورة واضحة عن عملية التطور للتركييب . والتسلسل الهرمي يسهل وضع البروتينات الجديدة في المجاميع او الأماكن الملائمة، وعليه وجدت أنظمة مختلفة منها (SCOP) Structural classification of proteins

و (CATH) Class architecture topology and homologous . وقد مرت هذه الأنظمة والقواعد بعدة خطوات لجعلها ذات ثقة عالية ومنها :

- إزالة التكرار والزيادة من قواعد البيانات العامة فمثلا " بنك بيانات البروتين PDB يحوي الكثير من المداخل للبروتين الواحد مع اختلافات بسيطة .
- إزالة التكرار والزيادة وذلك باختيار البروتينات الممثلة وذلك باستعمال التوجهات المعتمدة على اصطاف التواليات
- فصل الدومينات المختلفة التراكيب وذلك لان العديد من البروتينات مكونة من عدة دومينات والتي يجب ان تفصل قبل بدأ عملية مقارنة التراكيب . وعملية فصل الدومينات اما تتم يدويا" او باستعمال برامج خاصة ، ومن ثم يمكن مقارنة التراكيب على مستوى الدومين بطرق ملائمة . يلي ذلك تجميع الدومينات المتشابهة التراكيب وعندها باعتماد مستويات مختلفة من التشابه من مكونات التركيب الثانوي وترتيبها في الفراغ .

وقاعدة البيانات SCOP المذكورة آنفا" هي قاعدة بيانات تستعمل لتصنيف تراكيب البروتينات وتم إنشاؤها اعتمادا" على الفحص العملي لتراكيب البروتينات ورتبت بشكل هرمي كما يأتي

أصناف Classes

مجاميع Folds

عوائل كبيرة Superfamilies

عوائل Families

وتختلف اعدد المجاميع أعلاه وتزداد في الإصدارات الجديدة .

والأصناف المكونة من مجاميع لها تركيب مركزي عام متشابه الذي يميز مجاميع البروتينات بالاعتماد على مكونات التركيب الثانوي مثل كل تراكيب ألفا ، او كل تراكيب بيتا او الاثنين معا" وبعض الأصناف في هذا المستوى أنشأت اعتمادا" على المظاهر العامة مثل بروتينات الأغشية ، البروتينات غير المنتظمة وبروتينات صغيرة لها تراكيب ثانوية .

اما المجاميع Folds فهي تتكون من عوائل كثيرة لها تركيب أساسي مشترك الذي يحدد عمليا" . وهذا المستوى يصف التشابه العام للتراكيب الثانوية ، والأعضاء في هذا المستوى ليس من الضروري ان تكون ذات علاقات تطورية ، فبعض التراكيب العامة ربما نشأت من التناظر الوظيفي Analogy .

والعوائل الكبيرة تشمل عوائل لها تراكيب متشابهه ولكن تشابه التواليات فيها قليلا" . والعوائل الكبيرة مشتركة في أصولها .

اما العوائل فيها تضم البروتينات التي لها نسبة تشابه أكثر من 30% والذي يعني ان البروتينات في هذه الحالة لها علاقة تطورية وثيقة ولها وظائف متشابهه وتراكيبها متشابهة .

اما قاعدة البيانات CATH : تتكون من المستويات الآتية :

أصناف Classes

الأطر الخارجية Architecture

الطوبوغرافي Topologies المتماثلات الجذرية Homology

العوائل الكبيرة المتماثلة جذريا Homologous superfamilies

والأصناف تتشابه مع SCOP ، والقاعدة فيها درجة من التصنيف هو الأطر الخارجية التي تكون حالة وسط بين الأصناف والمجاميع Folds ، وهذا المستوى يصف ترتيب وتراص التراكيب الثانوية بشكل مستقل عن الترابط بين عناصر التركيب الثانوي ، وفيها أيضا "Topology مستوى يكون مكافئاً" للـ Folds في SCOP ويصف الترتيب العام للتركيب الثانوي آخذاً بنظر الاعتبار ارتباط التواليات بين عناصر التركيب الثانوي . وتوجد فروق بين القاعدتين يمكن للمستخدم ملاحظتها ، ناتجة عن الاختلاف في أسس بناء القواعد والطرق المستعملة في تحديد المستويات فضلاً عن الاختلاف مستويات التصنيف التي قد تلاؤم نوع معين من الأغراض دون الأخرى .

تهميش وظائف البروتينات

بعد المناقشة المارة الذكر التي ربما تناولت معظم الجوانب الخاصة بالبروتينات ، يفترض ان تقضي هذه الى تهميش وإثراء المعلومات حول البروتينات والجينات المشفرة لها . ومعظم وسائل المعلوماتية الحيوية التي تعتمد على تشابه التواليات لم تحقق المرجو منها ، فهناك 40% من الجينات التي تم تحديد توالياتها هي غير معروفة الوظيفة لذلك وضعت تحت مظلة الافتراضية وسميت بروتينات افتراضية . ولكن هذا قد يعود سببه الى ان الكشف التجريبي لوظائف هذه البروتينات وجيناتها يكون مسرفاً في الوقت والكلفة ، فضلاً عن الافتقار في تصميم التجارب . لذلك تتطلب الحالة حاجة ماسة لوسائل متطورة لحسد الوظيفة بالبحث عن التماثل الجذري البعيد Remote homology في كل قواعد البيانات والكشف عن التماثل البعيد الذي يتضمن عادة الجمع بين عمليات البحث عن قطيقات ودومينات البروتينات وحسد التركيب الثانوي والثلاثي والمواقع الفعالة الثابتة التي تكون بمساعدة استعمال النسق مثلما هو المتبع في بعض وسائل الحسد . والحسد في هذه الحالة يمكن باستعمال تحديد مقياس اللياقة او التعرف عن مواقع طوي البروتينات التي تشكل حجر الأساس في الصفات الخاصة بالبروتينات وكذلك التحري عن مواقع البروتينات في الخلية وتداخلات البروتينات مع بعضها . وإذا ما تم تحديد التماثل الجذري البعيد باستعمال الطرق المعتمدة على التركيب وربطها مع الوظائف الموصوفة بدقة عندها يمكن حسد صنف الوظيفة للبروتينات غير المعروفة وليس التوصل الى الوظيفة الدقيقة .

وعليه فإن كل ما ذكر أعلاه لا يعطى الضمان الكامل حول وظائف البروتينات الافتراضية ولكنها قد تزود ببعض الافتراضات المهمة لوظيفة البروتينات ليتم اختبارها مختبرياً . فضلاً عن ان استعمال قاعدة واحدة يكون غير مجدياً ويحتمل الخطأ الكبير . وعليه فإن الطرق الأوتوماتيكية يجب ان تشابه الطرق العملية التي تأخذ عدة خطوط من الحقائق في تحديد وظيفة البروتين او ألجين المشفر لتقليل الخطأ.

ومرة أخرى فإن إيجاد التماثل الجذري البعيد يساعد في إلقاء الضوء على الوظائف الممكنة للبروتينات التي لا توجد معلومات مسبقة عن وظائفها . ولذلك يلاحظ عجز المعلوماتية الحيوية عن حسد الوظائف واستقاء المعلومات الأخرى نظراً لان العديد من البروتينات التي تعد

افتراضية لها تراكيب وعمليات طوي خاصة بها لذلك بقيت هذه بمثابة تحديات للعمل التجريبي والحاسوبي على حد سواء .
ولكن وفقا للقاعدة « ما لا يدرك كله لا يترك جله » فقد أنشأت مزودات خدمة على الانترنت وبرامج وقواعد بيانات لتعمل في مجال تهيش البروتينات والبرامج المستعملة تعتمد جزءا من الوسائل المتكاملة لإظهار النتائج للوصول الى مستويات دقة عالية .

المعلوماتية الحيوية و RNA

يمثل RNA جزيئات وسطية بين DNA والبروتينات ماعدا في حالات خاصة من الفيروسات . اذ ان 90 % من جينوم الخلايا حقيقية النواة ينتسخ الى RNA وان الأخيرة يمكن ان تقوم بالمساعدة بالتفاعلات الحيوية التي كان معروفا انها حكرًا على الإنزيمات . ومثلما مر في فصول سابقة ان الجزيئات التقليدية مثل tRNA و rRNA تكون أساسية وتقع جيناتها قرب مناطق التضاعف في الخلايا بدائية النواة لزيادة عددها ، فضلا عن ان بعض جينات tRNA تقع ضمن اوبرونات rRNA ، لزيادة التعبير عنها .

وبزيادة المعرفة عن جزيئات RNA غير المشفرة ncRNA ووظائفها التي لم تعرف بعضها لحد الآن وإمكاناتها زادت من أهمية الجزيئات في عالم الأحياء حتى افترض البعض انها تمثل الوراثة الجديدة New genetics . وبذلك فأن دراسة RNA توفر فرصة لإعادة تقييم العلاقات التطورية ورسم خريطة جديدة للحياة . اذ تشارك في العديد من الفعاليات فضلا عن اشتراكها في العمليات التطورية وإنها المسؤولة عن تعقيد الخلايا حقيقية النواة ، ودراسة الجزيئات غير المشفرة وتحديد توالياتها ووظائفها سيكون مفيدا في دراسة تطور الأحياء ودراسة النواحي البايولوجية وكذلك في تصميم الأدوية ، فمن الدراسات في هذا المجال وجد ان miRNAs ودراسة أهدافها مثلا " جيدا " ، فهي تمثل احد ضروريات شبكات التنظيم التي تكون مهمة في تطور الخلايا وتمايزها فضلا عن أهميتها في عملية استجابة الخلايا وتكاثرها ، ووجد انها تشارك المسؤولية في حوالي 70 مرض في الإنسان في مقدمتها السرطان وأمراض القلب الوعائية .

والجزيئات غير المشفرة لها مناشئ متعددة فتمثل عوائل العناصر القافزة احد الروافد ، والأخرى تشتق من الأشرطة النقيضة وأخرى تكون ضمن تراكيب الجينات المشفرة للبروتينات وغيرها من المصادر (وقد مر تفصيل ذلك في فصل سابق).
ومن الجدير بالذكر ان الجينات المسؤولة عن الجزيئات غير المشفرة (وربما مناشئها المذكورة) تعاني من الطفرات بمستوى أعلى من الجينات المشفرة للبروتينات ، وقد أدت المعرفة بهذه الجزيئات وغيرها من جزيئات RNA الى فصل الاركيا الى دومين ثالث لشجرة الحياة العامة للأحياء .

تراكيب جزيئات RNA والتوجهات المستعملة للتنبؤ بها

يختلف تركيب الجزيئات تبعا للصنف الذي تنتمي إليه وقد مر في فصل سابق مناقشة التراكيب الخاصة بجزيئات RNA ونظرا لاختلاف تراكيب الجزيئات تختلف الوسائل والبرامج لحدها ولكن هناك توجهات واستراتيجيات يمكن ان تضم معظم البرامج .

التوجهات

التطورات السريعة في طرق حدس تراكيب جزيئات RNA المستعملة لدراسة ومرونة الأنماط المظهرية Phenotypic plasticity والتي تعني ان توالي واحد يمكن ان يعطي عدة تراكيب ثانوية أدت الى استعمال عدة طرق منها ما يختص بحدس التراكيب الثانوية والذي يؤدي بدوره الى تسهيل مهمة حدس التراكيب ثلاثية الأبعاد . فمثلا طرق الرنين النووي المغناطيسي تمثل احد الطرق الفيزيائية المستعملة في هذا المجال فهناك بصمة الرنين النووي المغناطيسي NMR fingerprint لعدد من القطيقات والعروات الخاصة بجزيئات RNA ، فضلا عن مشاركة الطرق الكيماوية في ذلك لتصنيع بصمة لخرائط الجزيئات Chemical mapping fingerprints وهذه التوجهات يمكن ان تساعد في إيجاد التراكيب الثلاثية في حالة وجود تناغم بين بياناتها . ولكن الطرق الفيزيائية تحتاج الى فهم عميق للتداخلات الجزيئية لتحديد النواحي الثرموداينمكية للتركيب ليتم التعرف منها على داينمكية جزيئات RNA التي تكون مهمة في أداء الوظائف . ولكن هذه الطرق قد تفشل في بعض الأحيان نظرا لتداخل الجزيئات مع بعضها او مع غيرها من الجزيئات الحيوية .

اما التوجهات الأخرى فهي الخاصة بالمعلوماتية الحيوية ، فجزيئات RNA تكون ملائمة للعمل الحاسوبي وبهذه الطرق يمكن الكشف عن العديد من العلاقات بين النمط الجيني والنمط المظهري ودور الطفرات التعويضية Compensatory mutants في بعض الجزيئات ، فضلا عن الكشف عن الانجراف المعتدل Neutral drift والانتخاب الذي يشمل تراكم الطفرات التركيبية الصامتة ومنها الطفرات التعويضية الأنفة الذكر والتي تسمح بتطور RNA وكذلك دراسة أفضل تطابق لأنواع الجزيئات المنظمة . والتوجهات في مجال المعلوماتية الحيوية وبرامجها يمكن ان تكون :

- البحث عن اقل مستوى للطاقة الذي يعطيه تركيب ما ، أي استخدام الثرموداينمكية باستعمال أفضل وسائل التعامل مع التواليات وهي البرمجة الداينمكية ويمكن ان يصل نجاح هذا التوجه ومعدل دقته الى 73 % .
- الاعتماد على تحليل التواليات المقارن لمجموعة من التواليات أي استعمال التغيرات المشتركة Covariation مثل جزيئات rRNA ذات الثبات العالي ، وهذه تكون أكثر الطرق دقة اذ يصل معدل دقتها الى 95 % . وهذه تظهر أهميتها القصوى عند الكشف عن أساسيات التطور .
- استعمال الطرق المشتركة منها الطرق أعلاه فضلا عن مشاركة نتائج التجارب للحصول على دقة عالية .

تداخلات جزيئات RNA

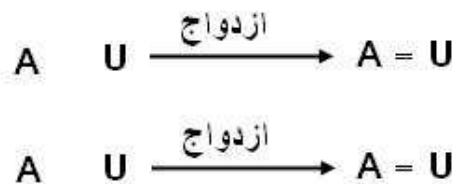
تعد الطرق أعلاها هي الأساس ولكن بعض الأحيان يمكن ان تدخل فيها بعض المؤشرات مثل تداخل جزيئات RNA مع بعضها أي تداخل RNA مع RNA والتي تسمح بتكوين التراكيب الخاصة مثل العروات والانتقافات او السيقان . وهذه التراكيب قد تحوي قطيقات تركيبية مهمة تميزها بعض البروتينات لتعمل تداخلا مع RNA . ومقارنة تواليات عددا من

جزيئات RNA لا تخضع لازدواج القواعد المعهود الذي ينتجه BLAST فضلا عن حقيقة ان تركيب جزيئات RNA يكون أكثر ثبوتا من توالياته (كما سيأتي ذكره لاحقا). فتداخل جزيئات RNA مختلفة وتحدد مسار فعاليتها فمثلا" التطابق المضبوط او غير المضبوط لجزيئات miRNA مع mRNA هي التي ستحدد المسار الذي تعمل فيه الجزيئات الأخيرة فهي اما تفكك بواسطة miRNA او تساعد في إحباط عملية الترجمة ولذلك فإن حدس التركيب سيوضح الاختلاف في إشارات RNA التي تؤدي الى اختلاف التراكيب المعقدة الناتجة والتي بدورها ستحدد أنواع البروتينات التي يمكن ارتباطها . وهذا يكون بتطوير خوارزميات قادرة على التعرف على المجاميع الثانوية التي تخضع لقواعد خاصة بالانضمام والاتحاد لنوع محدد من التواليات ونوع التراكيب الثانوية التي يمكن ان تنتج من التداخلات . اما تداخل جزيئات RNA مع البروتينات والتي تصل أعدادها الى المئات فهي تشكل إحدى ركائز تحديد تراكيب RNA . وهذه التداخلات يكون لها الأثر أفاعل في أداء الوظائف الحيوية . والتجارب تشير الى ان نيوكلوتيدات RNA التي لها القابلية على الارتباط وكذلك عملية الارتباط بحد ذاتها تتأثر بالحوامض الامينية المجاورة لمواقع الارتباط وهي التي تحدد الشريك من RNA ، وهذه التداخلات تعتمد على تمييز RNA للبروتينات التي يراد الارتباط بها وتعتمد على الدومينات والقطيفات الموجودة في جزيئة RNA كذلك ، وهذا الاعتماد يعتمد بدوره على التوالي والتركيب الذي يحدد بالتجارب والحاسوب وتكون هذه المواقع هدفاً لتطوير الأدوية .

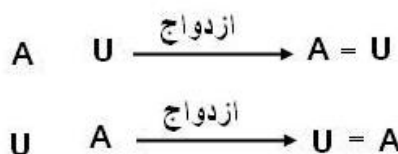
مبدأ اعتماد الطاقة الحرة

تمثل الطاقة الحرة احد الأساسيات في تحديد تراكيب جزيئات RNA وتعتمد في البرامج لتحديد التركيب الثانوي وكذلك تحديد التراكيب الثابتة . وهذا المبدأ مقيد بأساسيات كيميائية وفيزيائية ويمثل الثرموداينمك احدها ، فالجزيئات التي طولها اقل من 700 نيوكلوتيد يمكن حدس تركيبها بدقة تصل الى 70 % من الثرموداينمك لوحده . مما يشير الى ان الثرموداينمك هو المحدد الأكبر للتراكيب ومنها الثانوية وبالتالي يكون القوة الدافعة وراء تطور جزيئات RNA ذات التراكيب الخاصة وخاصة عند بدايات التطور عندما كانت الجزيئات ذات مرونة عالية من حيث التركيب ، وعند الوصول الى الحالة الوظيفية سيكون الثرموداينمك هو القوة المؤدية الى ثبوت التركيب الملائم على حساب الطويات والتراكيب البديلة ، ولذلك فإن التراكيب التي كانت قد تطورت مبكرا" يمكن تحديد تراكيبها بالاعتماد على مبدأ تقليل الطاقة الحرة مقارنة بالتراكيب المتطورة في أوقات متأخرة التي اعتمدت على حركات الطوي والتداخل مع البروتينات ، لذا كان فهم التطور والتركيب الى المستوى الجزيئي ذا أهمية كبيرة .

وأساسيات الطاقة الحرة تعتمد الى إيجاد الاحتمالية لأعلى الازدواجات كما موضح في الآتي :

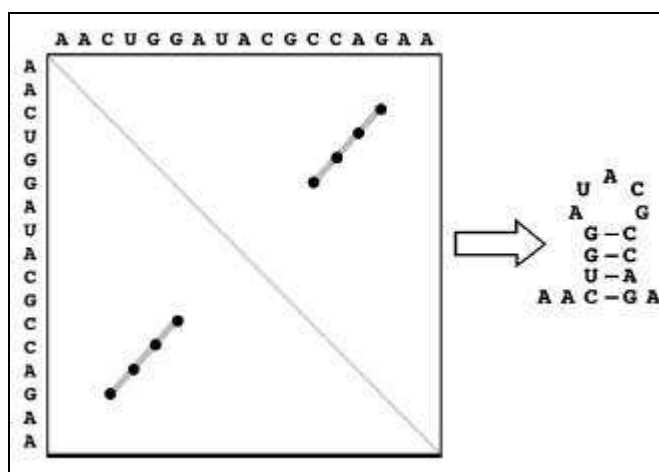


تكون الطاقة الحرة $\Delta G = 1.2$ - كيلوسعة \ مول



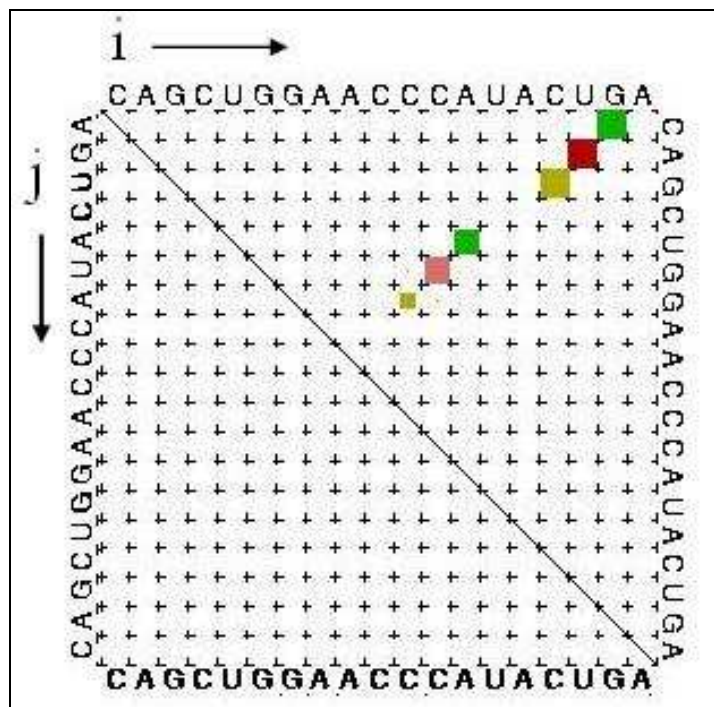
تكون الطاقة الحرة $\Delta G = 1.6$ - كيلوسعة \ مول

وفي هذه الحالة يجب البحث على قيم أوطأ للطاقة لكل أنماط الازدواج الممكنة للقواعد ويكون ذلك بإيجاد المصفوفات باستعمال البرمجة الديناميكية كما موضح في الشكل 142.5 :



شكل 142.5 : البرمجة الديناميكية لصف أزواج قواعد RNA

وعند ظهور أكثر من تركيب يتم اختيار الأفضل . وتصلح البرمجة الديناميكية للتواليات القصيرة ولا تصلح في حالة التواليات الكبيرة والتي تحوي أكثر من تركيب ثانوي وتقيد في تحديد تركيب واحد (وهو الذي يتم اختياره) وهذا قد يكون من معوقات استعمال البرمجة الديناميكية لان جزيئات RNA يمكن ان توجد بأشكال بديلة ولها مستويات متقاربة من الطاقة . ويمكن المساعدة في حالة البرمجة الديناميكية هو استعمال دالة التجزئة Partition function (Q) لإكمال وظيفة البرمجة الديناميكية . وهي تمثل حاصل جمع كل الطاقات الحرة للتركيب (خاصة الثانوي) الذي يمكن ان تنتج من توالي الاستعلام للـ RNA . فهذه الدالة (Q) توفر إمكانية حساب احتمالية ارتباط القواعد لكل زوج ممكن أي تحسب التوزيع الرياضي لاحتمال ازدواج القواعد في حالة التوازن الترموديناميكي لكل التراكيب الثانوية الممكنة في توالي الاستعلام . وتظهر النتائج بشكل مصفوفة الاحتمالية التي تسمى مصفوفة التفتيط وهي مستعملة في العديد من البرامج ويمكن إيضاحها بالشكل 143.5 :



شكل 143.5 : مصفوفة التنقيط Dot matrix لصف تواليات RNA

اذ يمثل الاحداثي السيني (i) والصادي (j) لتواليات RNA فكل موقع ازدواج ممثل بـ (i, j) ، والمربعات الظاهرة في المثلث العلوي (الأيمن) يظهر مربعات مختلفة الأحجام ، الصغيرة تشير الى احتمالية قليلة والكبيرة الى احتمالية كبيرة لتكوين زوج من القواعد (i, j) . وتستعمل النواتج في حسابات أخرى لاختيار عدداً من التراكيب الثانوية دون المثلى بمدى معين من الطاقة ، واعتماد ثابت التوازن الديناميكي .

وفي العموم اذا تمت معرفة الطوي وازدواج القواعد من معلومات الثرموداينمك فإن التراكيب الثانوية ممكن حدسها من التواليات لوحدها بالاعتماد على الطاقة الحرة اذ يكون التركيب الأفضل هو الذي تكون له اقل ما يمكن من الطاقة الحرة . فمثلاً " جزيئة RNA في حالة طوي (F) وفي حالة عدم طوي (U) يكون ثابت التوازن مساوياً " :

$$K = [F] / [U] = e^{-\Delta G^0/RT}$$

K يمثل ثابت التوازن الذي يعطي نسب الجزيئات المطوية وغير المطوية عند حالة التوازن . و ΔG^0 هي فرق الطاقة الحرة القياسية بين الشكل المطوي وغير المطوي ، R يمثل ثابت الغاز ، T تمثل درجات الحرارة المطلقة مقاسة بالكالفن Kelvin .

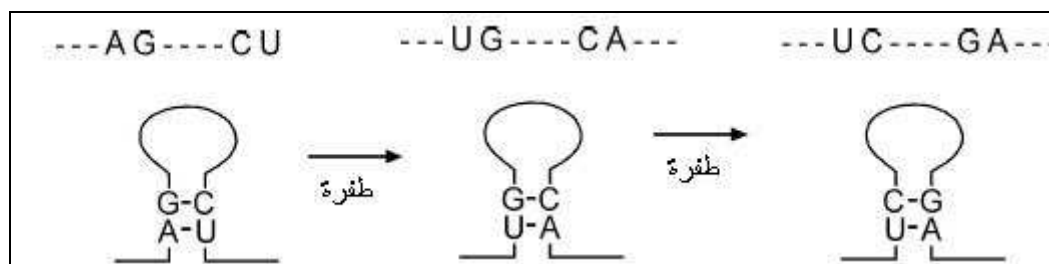
اما النهايات المتدلية Dangling ends من الجزيئات وهي نيوكلوتيديات غير مزدوجة فلها طاقة تثبيت وتوجد بجانب القواعد المزدوجة ، فمثلاً " هذه النهاية يمكن ان تؤخذ في حساب الطاقة الحرة للعروة او أي تركيب يجاورها يكون هناك أكثر من احتمال :

- اما ان تهمل ولا تلعب دورا" في الخوارزمية المستعملة
- او تؤخذ بنظر الاعتبار لكل تشكيلة Combination للقواعد المتجاورة وأزواج القواعد .

مبدأ الاعتماد على التحليل المقارن Comparative approach

المبدأ يمكن ان يعطى دقة أعلى من استعمال الطاقة الحرة كما ذكر أعلاه . فوائد هذا التوجه انها تعطي التركيب للجزيئات المفروض موجود داخل الخلية وتمكن من التعرف على التراكيب التي لا يتحكم بها الترموداينمك ، ولذا تستطيع تحديد تراكيب معينة مثل العقد الكاذبة التي تكون عقبة أمام بعض البرامج وكذلك توضح ازدواجات القواعد غير الخاضعة لنموذج Watson – Crick model . وبذا تساعد في زيادة استغلال المعلومات الموجودة في قواعد البيانات . والطريقة تحتاج الى عمليات اصطفاف عالية الجودة للتواليات المتعددة التي تخضع للدراسة ومثل هذه تكون غير ممكنة للجزيئات النادرة والأخرى عالية التغير، ولذلك يلاحظ ان برامج BLAST المستعمل للمقارنة يكون أدائه ضعيفا" لان البرنامج ينظر ويبحث عن تشابه الحروف (التوالي) وليس التركيب ، فضلا" عن ان البرنامج مصمم للبحث عن الجينات المشفرة للبروتينات ولذلك يلاحظ ان العديد من جزيئات rRNA لا تهتمش مع الجينومات في قواعد البيانات العامة.

وعند صف عددا" من تواليات RNA التي عادة تكون ذات علاقة تطورية لاستنتاج التراكيب المتوافقة باستعمال فكرة التغيرات المترافق واعتمادا" على ان القطيفات الوظيفية للـ RNA تكون ثابتة من ناحية التركيب ، فحدوث طفرات في موقع واحد مسئول عن ازدواج القواعد يجب ان يعوض بطفرة في الموقع المقابل للحفاظ على حالة الازدواج وثبوت التركيب مثل التركيب الثانوي كمل موضح في الشكل 144.5 :



شكل 144.5 : التغيرات المترافق Covariation للحفاظ على التراكيب الثابتة من جزيئات RNA

واعتمادا" على هذه الفكرة يمكن كتابة البرامج للبحث عن نمط التغيرات المتوافقة بعد إجراء عدد من عمليات الاصطفاف لعدد من تواليات RNA . وتساعد في ذلك أي في اختيار التركيب العام الملائم الرسوم المتوافقة Consensus drawing وذلك لان حدس التركيب بالاعتماد على التواليات كل على حدة يؤدي الى أخطاء . وباستعمال الطريقة والرسوم يمكن اختيار التركيب الأكثر احتمالا" وتحذف بقية التراكيب . والخوارزميات المعتمدة على عملية المقارنة

يمكن ان تقسم الى قسمين اعتمادا" على البيانات المدخلة ، الأولى تعتمد على اصطفافات جاهزة والأخرى لا .

الخوارزميات التي تستعمل الاصطفافات الجاهزة Prealignments تحتاج من المستخدم إدخال اصطفاف زوج من التواليات او مجموعة من التواليات المصطفة والتي يمكن الحصول عليها باستعمال برامج الاصطفاف القياسية مثل مجموعة Clustal او T- COFFEE او غيرها . واعتمادا" على الاصطفافات المدخلة تقوم البرامج بحدس التراكيب المتلائمة مع نمط الطفرات مثل التغيرات المترافقة ثم تشتق التركيب المتوافق العام لكل التواليات المدخلة .

وعادة ما يكون حدس التراكيب في هذه الحالة مرتبطا" بحسابات الترموداينمك لتحسين الدقة . ويمكن للطريقة إعطاء نتائج عندما تكون التواليات متشابهة الى حد معقول التي تسمح باصطفاف دقيق وتتفرج بشكل كافى لتسمح بحدس التغيرات المترافق وعند عدم توفر هذه الشروط فيكون من غير الممكن حدس التركيب . ولذلك يلاحظ ان الطريقة تعتمد على جودة المدخلات من التواليات المصطفة ، فاذا كان هناك أخطاء في الاصطفاف فأن التغيرات لا يمكن التعرف عليها . واختيار تركيب متوافق واحد يشكل إحدى معوقات استعمال برامج هذا التوجه وذلك لان البدائل والتراكيب غير الثابتة من الناحية التطورية لا يمكن حدسها .

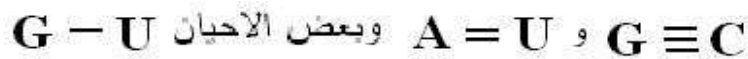
اما البرامج التي لا تستعمل الاصطفاف المسبق ، فهي تقوم بعملية الاصطفاف للتواليات المدخلة واستنتاج التركيب المتوافق ، وعملية الاصطفاف التي تجرى تكون عادة باستعمال البرمجة الديناميكية لدقتها واستنتاج مصفوفة الدرجات التي تشمل تشابه التواليات فضلا" عن إدخال مفاهيم الطاقة . واغلب البرامج المتوفرة تقتصر على إدخال اثنين من التواليات.

الطرق المتحدة Combined methods

وهي الطرق التي تعمل باستعمال التوجهات أعلاه مثل تقليل الطاقة وإدخال عمليات التحوير الكيماوي والاستفادة من البيانات حول ارتباط النيوكليوتيدات القليلة Oligonucleotides binding data وغيرها من المؤشرات المتوفرة . ومثل هذه الطرق يمكن ان تحول لتعمل أوتوماتيكيا" .

حدس التراكيب الثانوية

يمثل التركيب الثانوي أهم التراكيب الموجودة في جزيئات RNA ، ويضم أكثر من حالة منها الحلزونات المعروفة وتراكيب ماشات الشعر واللفات الداخلية وغيرها المارة الذكر ، وهي التي تكون الوحدات البنائية لجزيئات RNA المختلفة وتكون أساسية لوظائفها داخل الخلية . والبعض منها تتداخل مع بعضها مكونة التراكيب ثلاثية الأبعاد . وجزيئة RNA في الأساس جزيئات مفردة الأشرطة ولكن توجد فيها مناطق مزدوجة من



. وثبوت التركيب الثانوي يعتمد على عدد من العوامل منها :

- عدد GC مقابل AU و GU والأواصر العالية الطاقة تكون تراكيب أكثر ثبوتا"
- عدد أزواج القواعد في منطقة الساق (المزدوجة) فالسيقان الطويلة يكون فيها عدد أواصر اكبر.

• عدد أزواج القواعد في تركيب ماشة الشعر فالعروات التي تكون أكثر من 10 أو أقل من 5 قواعد تحتاج طاقة أكبر .

• عدد القواعد غير المزدوجة في العروات والفقاعات ، وزيادة عدد القواعد غير المزدوجة يقلل من ثبوت التركيب .

ان ثبوت التركيب الثانوي يقاس بكمية الطاقة الحرة المنطلقة أو المستعملة عند تكوين أزواج القواعد . والطاقات الحرة تكون ذات طبيعة تراكمية أي يمكن معرفة الطاقة الكلية من تجميع الطاقات الحرة لطاقات التراكيب المتكونة . وتستعمل هذه في حدس التراكيب الثانوية حيث يتم البحث عن التراكيب ذات الحد الأدنى من الطاقة الحرة .

اما الطرق الأخرى لحدس التراكيب الثانوية فهي توجه طرق البداية التي تعمل على حدس التركيب اعتمادا على التوالي للـ RNA والجوهر فيها هو الاعتماد على الطاقة الحرة، وفيها يتم تصميم البرامج للبحث عن التركيب الثابت الذي له أوطأ مستوى من الطاقة الحرة ، وفي العادة فإن تكوين الازدواج تؤدي الى ان الطاقة للجزيئة ستخفض نتيجة لتداخلات التجاذب بين الأشرطة وبالتالي يتم البحث عن التركيب الأكثر ثباتا" وعليه فإن هذه البرامج تصمم للبحث عن التركيب الذي له أقصى ما يمكن من ازدواج القواعد باستعمال طريقة البرمجة الديناميكية . وتحسب الطاقة الحرة اعتمادا على مؤشرات يتم تحديدها بالتجربة ، فمثلا GC يكون هذا الزوج أكثر استقرارا من AU والذي بدوره أكثر استقرارا من GU . والمعروف ان ازدواج القواعد هي ليست أحداث مستقلة وانما تتأثر بالقواعد المجاورة من خلال ما يعرف بالتعاونية Cooperativity في تكوين الحلزونات ، فاذا كان مزدوج من القواعد مجاور لآخر فإن ازدواج القواعد تميل لتثبيت احدها الأخرى بواسطة تداخلات التنضيد او الرص بين التراكيب الحلقية لكل زوج من القواعد ، وهذا التراص والتجاذب يؤدي الى خفض الطاقة ويقع هذا ضمن مفهوم نموذج الجوار الأقرب الذي تكون عنده ΔG^0 هي حاصل جمع الزيادة في الطاقة الحرة لمختلف القطيقات في الجوار الأقرب مثل ازدواج القواعد النتروجينية المرصوصة في حلزونات RNA التي تحصل في التركيب ، وقد تم حساب مؤشرات التعاونية لتكوين ازدواج القواعد من التجارب بدراسة عمليات الصهر وتستعمل في حدس التركيب . والملاحظ ان ازدواج القواعد تكون قليلة الاستقرار عندما تكون مجاورة لعروة او فقاعة، والزعزعة هذه تكون معتمدة على نوع العروة القريبة ، وقد تم حساب طاقة القوى المزعزعة لتدخل ضمن الحسابات اللازمة لتحديد التركيب الثانوي . ولهذا الغرض يتم وضع مخطط للدرجات المعتمدة على الجمع بين التداخلات المثبتة والمزعزعة التي تكون الأساس في حدس التركيب بطرق البداية التي تعمد الى أيجاد التركيب المفضل من ناحية الطاقة بعد اختياره من تراكيب بديلة يمكن ان تظهر .

ويمكن استعمال فكرة Stochastic sampling (وهي الأساس في طريقة مونتني كارلو المارة الذكر) والتي تعنى استعمال متغيرات عشوائية ، وإجراء التحليلات لإيجاد ازدواج القواعد وعنقدها والتركيب الناتج يسمى المركز Centroid . وعملية العنقدة يمكن ان تكون أكثر دقة في حدس التراكيب الطبيعية مقارنة بحدس التركيب باستعمال اقل طاقة حرة لوحدها وهذه تستعمل في حدس جزيئات RNA غير المشفرة مثل تواليات Riboswitch (المذكور في موقع آخر) .

ومن الطرق الأخرى لتحديد التركيب الثانوي للـ RNA هي الطرق الكيماوية أو الفيزيائية . وهذه تؤدي إلى إنتاج عددا " كبيرا" من التراكييب ويكون إيجاد التركيب الصحيح فيها ليس بالعملية السهلة . توجد بعض الطرق الكيماوية التي تعتمد على تحويل القواعد باستعمال مواد خاصة أو تحويل سكر الرايبوز لتحويل مواقع الطوي ، وتغير السكر يعطى نتائج أكثر انتظاما " وأكثر ثبوتا " . والطرق الكيماوية في هذه الحالة تحدد الفعالية النسبية باستعمال مواد كيماوية مختلفة ومنها يمكن تحديد أي القواعد قادرة على الازدواج من الأخرى غير القادرة . ويمكن تحديد التراكييب الثانوية بالاعتماد على الطرق الإنزيمية واستعمال بواي مختلفة والنسخ العكسي الذي يتوقف عند القاعدة (النهاية 3') التي فيها تغيير وتعد هذه طريقة تجريبية ملائمة لبعض الحالات .

اما طريقة الرنين النووي المغناطيسي فتستعمل في بعض البرامج الخاصة وهي تحدد القواعد غير المزدوجة ، ولكن تعد الطريقة غير ملائمة نظرا " للحاجة إلى استعمال النتروجين المعلم (^{15}N) فضلا " عن محدوديتها للعمل مع تواليات تتراوح فقط بين 100 - 300 نيوكلو تيد وتحتاج إلى كميات كبيرة من RNA مقارنة بالطرق الكيماوية .

طرق تحديد التراكييب الثانوية الثابتة

تحديد التراكييب الثابتة يهدف إلى تحديد ومعرفة المواقع ذات التغيرات المترافق أي المواقع التي تظهر طفرات ذات علاقة ، وتزامن هذه الطفرات بين المواقع يوثق التطور الجزيئي للـ RNA . وتمثل الطفرات النقطية أو الموضعية في المواقع المصطفة التي تحافظ على ازدواج القواعد أبسط التغيرات المترافقة للتحديد والتفسير فهي تمثل الطفرات التركيبية الصامتة . وتهدف بعض الطرق إلى تحديد التراكييب الثابتة ولذلك ربما تختلف بعض الشئ عن حدس التراكييب الثانوية العامة . وهناك برامج تستعمل خطوات متعاقبة واستبدالها لإيجاد الأفضل . وفي هذه الحالة يكون من الأفضل تحديد التركيب الثانوي بالاعتماد على التوالي حصريا " والبرامج في هذا المجال توفر التواليات المتعددة لاستخدامها لتقارن بها تواليات الاستعلام ولكن المشكلة تكون بتحديد أقل مستوى من الطاقة ، ولأمثلة هذه التوجهات فهي تحتاج إلى :

- أمثلة عملية الاصطفاف لإيجاد أوطأ مستوى من الطاقة للتركيب .
 - البدء بتواليات مصطفة اعتمادا " على تماثل النيوكلو تيدات وليس تشابهها ثم إيجاد الأزواج الثابتة في الاصطفاف الناتج .
 - حدس التراكييب ذات الطاقة الحرة الأوطأ لكل توالي بشكل منفصل .
- ثم تجري عملية فصل وترتيب للتراكيب المحدوسة لإيجاد التراكييب العامة في التواليات المستعملة . ويمكن أن تجري العمليات بخطوات مختلفة :

❖ **تحديد الطوي والاصطفاف Fold and align** وفيها تستعمل البرمجة الدائيمكية بشكل متزامن للطوي والاصطفاف لمجموعة من التواليات وهي المستعملة في عدد من البرامج لإيجاد التراكييب العامة ذات الطاقة الحرة الأوطأ ، وهذه تصلح لتواليات أقل من 2000 قاعدة . ولتحديد مدى الثقة تستعمل دالة التجزئة (Q المذكورة سابقا) لحساب ثوابت التوازن للتراكيب العامة مع تحديد الدرجات للطاقة الحرة الكاذبة بالاعتماد على الاحتمالية لازدواج القواعد واصطفاف التواليات وهذه تكون ممكنة لاثنتين من التواليات ، اما عند زيادة عدد

التواليات فيستعمل الطريق الموجه المختصر Greedy heuristics لإيجاد التركيب العام باستعمال حسابات للزردواج المتعدد Multiple pairwise calculations

❖ **تحديد الاصطفاف ثم الطوي Align then fold** هذه الحالة يتم إجراء الاصطفاف لعدد من التواليات اعتماداً على المعلومات الموجودة في التوالي ، ثم تحديد التركيب ذات الطاقة الحرة الأوطأ الذي يكون عاماً لمعظم التواليات أو كلها . ثم تجرى حسابات بتغيير قيم الطاقة الحرة الإضافية لتكوين أزواج قواعد في المواقع التي فيها التغيرات المترافقة ، ويحدد التركيب الثابت ولكن ليس التوالي الثابت . وهذه الطريقة سريعة ويمكن ان تستعمل لأي عدد من التواليات . ولكن المشكلة ان جودة الحدس تعتمد على جودة الاصطفاف ، وذلك لان تحديد أزواج القواعد التعويضية (التي فيها Covariation) تعد المفتاح الأهم في جودة تحديد التراكيب . وهذه يمكن تحسينها بإصدار مخرجات مصورة لكل خطوة تنبه المستخدم ليقوم بتقييمها وتعديلها وإعادة عمليات الاصطفاف .

❖ **تحديد الطوي ثم الاصطفاف Fold then align** وفيه يعتمد مؤشر الطاقة الأقل لتحديد التركيب الثانوي لعدد من التواليات ، التي تستعمل لإيجاد التركيب المشترك الذي له أوطأ مستوى من الطاقة . وتطبيق هذا المبدأ يكون غير ملائم حاسوبياً لان التراكيب الناتجة من مؤشر الطاقة الأوطأ للتوالي تكون كثيرة وكذلك فإن عدد التراكيب للتوالي تتغير بتغيير الطاقة الحرة للطوي اذ تكثر الأعداد بشكل تزايدى وبالتالي لا يمكن فرز وتصنيف التركيب الصحيح . ولكن توجد برامج تعتمد مبدأ الطرح أو تلخيص الأشكال لتساعد في ذلك . ومبدأ إجراء الطوي ثم الاصطفاف يكون أسرع من طريقة الطوي والاصطفاف كما انه لا يتأثر بدقة عمليات اصطفاف التوالي .

حدس التركيب الثلاثي

يلعب التركيب الثلاثي دوراً مهماً في أداء بعض الجزيئات لوظيفتها وهناك عدة تراكيب ثلاثية يظهر أكثرها في 16S rRNA و 23S rRNA وينتج من تكوين أوامر هيدروجينية إضافية أو تداخلات كارهة للماء والقوى المشتركة هنا هي أقل قوة من تلك المكونة للتركيب الثانوي وأكثر الخوارزميات والبرامج تهتم بتحديد التراكيب الثانوية . ويتم تحديد التراكيب الثلاثية بالدراسات البلورية للجزيئات . ولكن يوجد عدداً من البرامج التي تبدي محاولات لتحديد التركيب الثلاثي خاصة لجزيئات tRNA .

وبداية فإن تركيب tRNA (ورق البرسيم) Clover leaf الثلاثي يتم تحديده باستعمال الأشعة السينية للتراكيب البلورية . اما في حالة استعمال برامج الحاسوب للتحري عن التواليات في الجينوم فليس الأمر كذلك ، فعدد كبير من التواليات في الجينوم تعطي تركيب ورقة البرسيم هي ليست جزيئات tRNA ولذلك فإن هذه الطرق تعطي نسبة عالية من النتائج الموجبة الكاذبة . اما اذا استعملت مؤشرات الطاقة فإن التغيير في الطاقة الحرة للطوي تعطي مؤشرات جيدة للتفريق بين النتائج الموجبة الحقيقية والكاذبة . وذلك لان الجزيئات غير التابعة للـ tRNA يكون لها تغيير في الطاقة الحرة أقل مقارنة بالجزيئات التي تطوى الى شكل ورقة البرسيم وهي من tRNA الحقيقية .

حس تراكيب جزيئات RNA غير المشفرة

تلعب الجزيئات غير المشفرة ncRNA الدور الأساسي في عدد كبير من الفعاليات الحيوية ، وتحديد الجينات المسؤولة عنها في الجينومات هي حاجة ملحة ولم تحل في المعلوماتية الحيوية . ولكن هناك توجهات عدة منها :

- تحليل المكونات مثل البحث عن GC لتحديد جينات ncRNA
 - استعمال التماثل الجذري في حال توفر البيانات الخاصة بالجينومات
 - استعمال الممهدات ومناطق نهاية الجينات كإشارات للحدس
- ان أيجاد الجينات الخاصة بهذه الجزيئات يساعد في اكتشاف توالياتها وكذلك إمكانية اكتشاف جزيئات جديدة منها .

اما تحليل وحس تراكيب هذه الجزيئات فيتم باستعمال معظم الطرق السالفة الذكر والأفضل اقران كل الطرق الممكنة مثل أيجاد الطويات باستعمال مؤشرات الثرموداينمك مع الخرائط الكيماوية ومعلومات الرنين النووي المغناطيسي ومقارنة التواليات ، فضلا عن استعمال الطاقة الحرة فهذه الجزيئات تكون الطاقات الحرة لها أعلى من باقي الجزيئات ولذلك يستعمل مؤشرات الطاقة في حدس التراكيب الثانوية لها والتي بدورها تحدد التراكيب الثلاثية ان وجدت .

ان أيجاد طرق طوي هذه الجزيئات يساعد في الكشف عن او تحديد المؤشرات الثرموداينمكية التي تكون الأساس في بعض البرامج الخاصة مثل تلك المستعملة للكشف عن miRNA والتي كشفت البرامج ان لها أهدافا في mRNA وأنها قد تكون نشأت في العناصر القافزة خاصة اذا عرف ان 12 % من جزيئات miRNA في الإنسان حاوية على بعض مكونات العناصر القافزة .

ان اعتماد مبدأ الثرموداينمك يكون أساسه ان تركيب جزيئات RNA يكون أكثر ثبوتا من تواليه ولذلك يتم البحث عن التراكيب الثانوية للـ ncRNA باعتماد التغيير في طاقة الطوي الحرة والتي تكون أوطأ من تلك الخاصة بالجزيئات العشوائية . وتستند البرامج على قيم إحصائية خاصة بالموضوع اعتمادا على تراص النيوكلويتيد الثنائية . وتوجد برامج تستعمل عمليات الاصطفاف ثم أيجاد الطوي او بالعكس مثلما ذكر أعلاه لتحديد التراكيب الثابتة فيها . ونظرا لاختلاف الجزيئات غير المشفرة من حيث الطول ومكوناتها من القواعد فيمكن ان توجد برامج متخصصة بكل صنف منها . والتوجهات الحديثة تميل الى استعمال اكبر قدر ممكن من المؤشرات ومنها حزمة او برامج moses (modular sequence suite) ، والبرنامج يستعمل كل الطرق المتوفرة تقريبا واستعماله يعتمد على خلفية المستخدم البايولوجية . والبرنامج يستعمل عدة برامج أخرى وتكون مخرجات برنامج هي مدخلات لما يليها وتدخل القيم الإحصائية فيه لتصفية وترشيح المخرجات للدخول الى برامج أخرى وما يميز البرنامج هو وسائل إظهار الرسوم لكل نتائج الخطوات الوسيطة والتي تساعد في أيجاد الأخطاء وإيجاد الفرضيات فضلا عن المساعدة في تفسير النتائج .

تداخل RNA مع البروتينات

تمثل احد الجوانب التي تهتم بها المعلوماتية الحيوية في عالم RNA ونظرا لأهمية التعرف على هذه التداخلات المشار إليها سابقا طورت البرامج لحدسها . فالتعرف على التداخلات ليس

بالأمر السهل سواء من الناحية التجريبية أو باستعمال الحاسوب . ولذا كان لابد من تطوير برامج حاسوب تحدد هذه المواقع اعتماداً على البيانات المتوفرة حول معقدات البروتين مع RNA . ونظراً لوجود دراسات كثيرة تشير إلى وجود علاقة قوية بين ثملالات التداخل وتركيبها في المعقدات لذا طورت طرق تعتمد على قياس ميل أو نزعة التداخل RNA Interaction propensity وهي تشير إلى حالة التخصص لارتباط ثملالات البروتين مع نيوكلوتيدات RNA مع الأخذ بنظر الاعتبار الثملالات المجاورة من الجهتين لموقع الارتباط وتصل دقة بعض هذه الطرق إلى 84.5 % . وأكثر الطرق تستعمل تقنيات أو طرق جيدة لعمليات تحديد المواقع مثل الشبكة العصبية أو SVM .

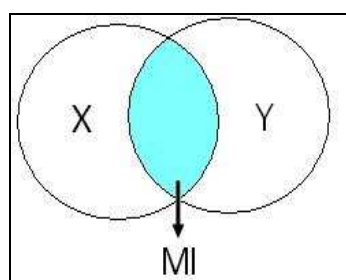
[Support vector machine] وهي خوارزمية تستعمل للتعرف على الأنماط الدقيقة من البيانات المعقدة لذا تساعد في أداء جيد لعمليات تصنيف تمييزية . ويتم تدريب البرنامج على معلومات وتصنيفات سابقة ، ويستعمل في العديد من برامج المعلوماتية الحيوية التي تتناول المواضيع الدقيقة مثل إيجاد مواقع بدء الانتساخ وتحديد التماثل الجذري البعيد للبروتينات وفي التعرف على طوي البروتينات ، وكذلك في تحليل التعبير الجيني في المصفوفات الدقيقة وتحديد التصنيف الوظيفي لمناطق الممهد وكذلك تداخلات البروتينات مع بعضها وتشخيص الببتيدات من المعلومات الناتجة من بيانات قياس الطيف . والخوارزمية مرنة لاستعمالها في العديد من الدوال والعلاقات لأنها تعتمد على ما يسمى بدالة Kernel function والأخيرة توفر المرونة للخوارزمية بسماعها لتحويل البيانات] .

ونظراً لدقة وتعقيد حدس التداخلات أدخلت خوارزمية SVM لدقتها العالية والنجاحات التي حققها عند الاستعمال في حدس تداخلات RNA والبروتينات وتساعد SVM في إعطاء نتائج اعتماداً على تواليات الحوامض الأمينية والتراكيب الثانوية في البروتينات ، وذلك لأن الحوامض الأمينية المجاورة لموقع الارتباط والطبيعة التركيبية للثملالات تؤثر كمشارك قوي في عملية الارتباط . وبالأخذ بنظر الاعتبار طبيعة التداخل بين البروتينات و RNA التي تشمل تكوين الأواصر الهيدروجينية وقوى كهربائية مستقرة وكراهية الماء وقوى فاندرفال ، فهذه تحسب من الاعتماد المشترك بين ثملالات البروتين و نيوكلوتيدات RNA مع الأخذ بنظر الاعتبار الثملالات المجاورة لتحديد التخصص ، ووجد أن 9.87 % من ثملالات البروتين قادرة على الارتباط أما الباقي فلا ترتبط .

ولأهمية المناطق المجاورة وصفات ثملالات البروتين التي تعد الأهم في تحديد موقع الارتباط فالعديد من البرامج تأخذ بنظر الاعتبار العديد من صفاتها مثل :

- الصفات الفيزيوكيميائية وتشمل عدد الذرات وعدد شحنات الكهربائية المستقرة وعدد أواصر الهيدروجين الفعالة وهذه تحسب بغض النظر عن التركيب الذي تشارك فيه الحوامض الأمينية .
- حساب كراهية الماء بالرجوع إلى الفهارس الخاصة Hydrophobicity index .
- التراكيب الثانوية التي تشارك فيها الثملالات المجاورة مثل الحلزونات والصفائح واللفات التي تزود ببرامج خاصة .
- قياس المساحة السطحية النسبية المعرضة للماء .

- استعمال درجات التوافق Consensus scores باستعمال PSI – BLAST لاستنتاج درجات ثبوت التواليات . ومصفوفات البروتينات الخاصة لتسجيلها.
- طاقة السلاسل الجانبية وتتمثل بقياس pKa للسلاسل الجانبية للحوامض الامينية والتي توجد لها جداول خاصة ، فضلا عن استعمال دالة المشاركة Mutual information (MI) function وهي طريقة لقياس المشاركة لمتغيرين عشوائية وتقاس بـ bit عند استعمال اللوغاريتم للأساس 2 . وهي موضحة في الشكل 145.5 :



شكل 145.5 : دالة المشاركة Mutual information function

وتستعمل لقياس الانتروبي للثملات عند حدس التراكيب الثانوية للـ RNA . فالملاحظ عندما يكون X و Y مفردات مستقلة فإن قيمة MI تكون مساوية للصفر وعند زيادة قيمة MI فهذا يعني ان هناك مشاركة واعتماد كل من المتغير على بعضهما . كل هذه تدخل الى برامج خاصة مدربة على معلومات مسبقة لتعطي قيم إحصائية التي على ضوءها يتم تصنيف النتائج . واهم القيم الإحصائية المستعملة في هذا المجال هي نزعة التداخل المشتركة Mutual interaction propensity باستعمال الدالة المذكورة آنفا .

حدس البروتينات غير معروفة الوظيفة المرتبطة بـ RNA

البروتينات غير معروفة الوظيفة Proteins of unknown functions (PUFs) والتي تصنف غالبا على انها بروتينات افتراضية تصل نسبتها من 18 - 40 % وتزداد النسبة كلما زاد تحديد المزيد من تواليات الجينومات . ومن المعروف ان حدس الوظيفة يكون ضروريا لوضع الخطط للعمل التجريبي مثل إجراء عمليات التطهير وإحدى الطرق المهمة لمعرفة الوظيفة هو معرفة وتحليل الربائط التي ترتبط بالبروتينات واهم هذه الربائط هي RNA والذي تصل نسبة البروتينات التي ترتبط اليه والمسجلة في بعض قواعد البيانات المعتمدة الى 15 % . وفي هذا المجال من الدراسة لابد من :

- تحديد التوالي البروتيني الذي يرتبط الى RNA
 - تحديد وظيفة عملية الارتباط الافتراضية للبروتينات
- وهناك العديد من التقنيات التي تعنى بالحالة الأولى . ونظرا لصعوبة تحديد عملية ومواقع الارتباط كان من الواجب إدخال مواصفات أخرى في الخوارزميات المستعملة وذلك لان تحديد هذه المواقع يكون مهما واضطرابه يؤدي الى العديد من الأمراض في مقدمتها حدوث

السرطانات ، وبالتالي توفر الفرصة لإيجاد أهداف دوائية جديدة . ومنها ما يخص البروتينات مثل :

- حجم قوى فاندرفال باستعمال طرق خاصة واختيار المسافات التي تكون اقل من 3.9 انكستروم .
- تحديد الأواصر الهيدروجينية
- تحديد القطبية ، وتحديد الشحنة
- الشد السطحي
- الإتاحة للمذيبات التي تؤثر في إمكانية تعرض الشمالات للمذيب
- التركيب الثانوي
- مكونات البروتين من الحوامض الامينية .
- طول توالي البروتين

وكل هذه المؤشرات تدخل الى البرنامج بخطوات متتالية . ومن أهم الطرق العاملة في هذا المجال طريقة PiRaNhA . والطريقة تستعمل نافذة كبيرة تشمل 23 ثمانية لتأخذ بنظر الاعتبار الثمالات المحيطة بمكان الارتباط أي تدخل البيئة المحيطة بالتوالي المعني بالارتباط في الحسبان ووجد ان دقة الطريقة تصل الى 87.2 % ، والأخذ بنظر الاعتبار البيئة المحيطة يعود الى ان ارتباط RNA الى البروتين قد يكون غير مباشرا" وانما يمكن ان يرتبط الى ثمالات مجاورة تشارك في عملية الارتباط نظرا" لاعتبارات الطاقة .

حس ترانكيب RNA داخل المناطق المشفرة

المعروف ان جزيئات RNA غير المشفرة ncRNA تعمل بمساحتين مهمة الأولى هي كونها Trans-acting أي تعمل عن بعد وتذهب الى الأهداف البعيدة لتمارس فعاليتها في التنظيم . المساحة الثانية هي عملها كعناصر تنظيم داخلية Cis-acting elements ومثل هذه تكون في جزيئات mRNA في المناطق غير القابلة للترجمة (UTR) وكذلك توجد في جينومات الفيروسات التي تكون جينوماتها من نوع RNA التي تشترك في تنظيم فعالية الجينات . ومثال آخر في خلايا حقيقية النواة فإن مشجعات الاكسونات Exonic enhancers تكون تراكيب من RNA . ولتحديد تواليات RNA في مثل هذه الحالات يطبق مبدأ الترموداينمك على المناطق المشفرة للبروتين وإيجاد RNA بالمقارنة بين اقل مستويات من الطاقة الحرة .

البرامج والوسائل في هذا المجال لا تزال قليلة ومن أهمها RNA – Decoder الذي يستعمل لإيجاد تراكيب RNA في المناطق المشفرة للبروتين وتحديد تداخلها ويأخذ بنظر الاعتبار العلاقات التطورية . ويدرب البرنامج على جينومات الفيروسات مثل HCV التي تعد من مظاهر التطور الثابتة فضلا" عن اعتماده على قياسات الطاقة والصفات الفيزيائية . فالتراكيب الفعالة في هذه الحالة (بشكل عام) لا تميل الى المستوى الأقل من الطاقة (كما هو الحال مع تراكيب RNA الأخرى) مقارنة بالمناطق الأخرى . وتستعمل في الحس تراكيب السيقان Stem pairs لحس المناطق الأكثر ثبوتا" . ولكن التوجه في معرفة تراكيب RNA ضمن المناطق المشفرة يلاقي العديد من الإخفاقات وذلك يعود الى استعمال جينومات الفيروسات مثل

HCV في التدريب وهي غير شائعة وكذلك صغيرة ، لذا كانت الحاجة ملحة الى مؤشرات أخرى في التدريب منها استعمال معلومات التطور لحداث تراكيب RNA الفعالة في او قرب المناطق الجينومية المشفرة للبروتينات.

قواعد البيانات

هناك عدداً من قواعد البيانات التي توفر المعلومات حول تواليات ووظائف جزيئات RNA لجميع ممالك الحياة (البكتريا ، الاركيا ، والخلايا حقيقية النواة) . ومعظمها تعتمد على قاعدة بنك الجينات بالإضافة الى استعمال قواعد بيانات أخرى للحصول على معلومات إضافية ولهذه القواعد إصدارات مختلفة تجدد مع مرور الزمن . واهم هذه القواعد والمستعملة بكثرة في البرامج هي Rfam فهي تضم تواليات لـ 248 نوع من البكتريا الممثلة بعدد 19349 من التواليات والاركيا 24 نوع والممثلة بعدد 1098 من التواليات ، والخلايا حقيقية النواة 152 نوع ولها 5586 من تواليات RNA وفق الإصدار الأخير (عام 2010) وفيها قواعد ثانوية متخصصة ببعض الأحياء التي تعد نماذجاً دراسية مثل القاعدة الخاصة برشاد الصخر تحت رمز ASRP وأخرى خاصة بخميرة الخبز .

اما أعداد عوائل جزيئات RNA الممثلة في القاعدة وأنواعها موضحة في الجدول 11.5 :

جدول 11.5 : أنواع جزيئات RNA وأعدادها الموجودة في قاعدة البيانات Rfam

التصنيف	عدد العوائل	عدد التواليات
Gene	31	3475
Gene ; Antisense	12	401
Gene ; miRNA	46	1391
Gene ; rRNA	3	96818
Gene ; ribozyme	11	1305
Gene ; sRNA	44	800
Gene ; snRNA	1	353
Gene ; snRNA ; snoRNA ; CD-box	199	4306
Gene ; snRNA ; snoRNA ; HACA-box	83	1565
Gene ; snRNA ; snoRNA ; scaRNA	16	141
Gene ; snRNA; splicing	10	4726
Gene ; tRNA	1	85463
Intron	2	28274
CIS- reg	99	34972
CIS- reg ; IRES	27	6601
CIS- reg; Frameshift	5	4706
CIS- reg ; riboswitch	10	1873
CIS- reg ; thermoregulator	3	157
CIS- reg ; leader	4	142

ومثل هذه القواعد تخضع للتدقيق والتشذيب لبياناتها للابتعاد عن التكرار . وفي قاعدة بيانات Rfam يمكن البحث عن التوالي تحت اسم الكائن أو رمز جزيئة RNA أو استعمال رقم التسجيل التابع لبنك الجينات ، كما يوفر موقع القاعدة الالكترونية إجراء المقارنات باستعمال BLAST

البرامج المستعملة لدراسة RNA

تختلف البرامج في الأسس التي تبني عليها وبالتالي تختلف المخرجات الناتجة منها . وأكثر المواقع تعاملًا مع RNA هو مزود الخدمة Vienna RNA webserver الذي يوفر حزمة من البرامج التي تتعامل مع مختلف المؤشرات المستعملة في حدس RNA بصفاته المختلفة وبعض البرامج تعطي التركيب الحلقي Circular folding لجزيئات RNA هي مفتوحة ولكن يحدث وجود حلقة مفردة ولذلك يجب ان تؤخذ بنظر الاعتبار ، وأفضل البرامج هي التي تعطي نتائج مصورة مع تهميشها .

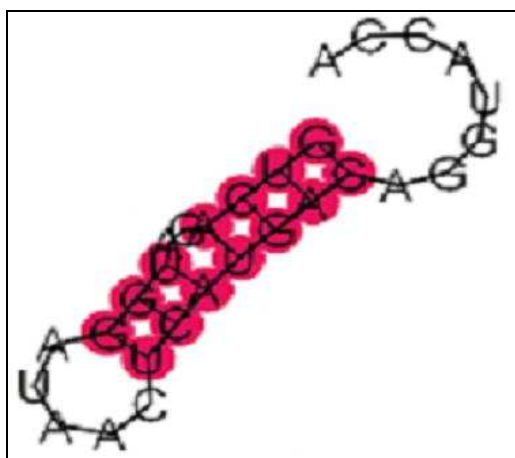
اما البرامج الأخرى فتظهر أكثر من صيغة للمخرجات فمثلاً حزمة البرامج التابعة للـ Vienna RNA فتظهر مخرجات مصورة فضلاً عن إظهار التوالي فمثلاً التوالي الآتي

GUCAGUGGAUAACUCAUGACAGGUACCA

يظهر الشكل كالاتي

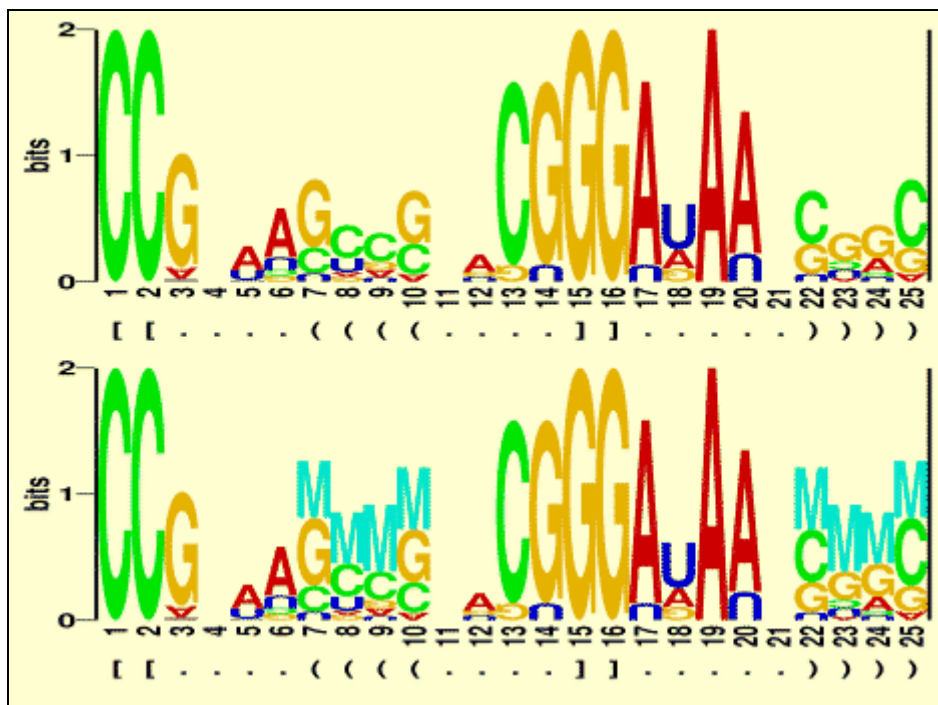
((((((.....)) *)))))

اذ تمثل النقاط مواقع القواعد غير المزدوجة ، في حين يمثل () موقع القاعدة من الموقع (i) اما () فيمثل موقع القاعدة (j) في المصفوفة المستعملة . والشكل الآخر يظهر التوالي أعلاه وبشكل رسم كما موضح في الشكل 146.5 :



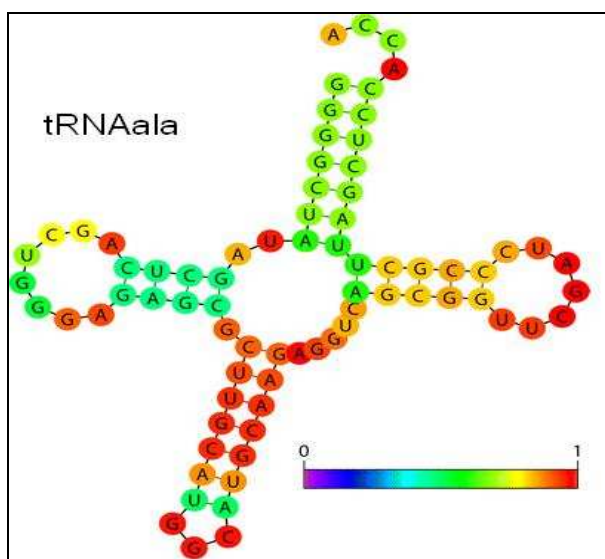
شكل 146.5 : التمثيل الصوري للتوالي

ويمكن ان تربط هذه بأحد برامج الشعار لتوضيح النتائج كما موضح في 147.5 :



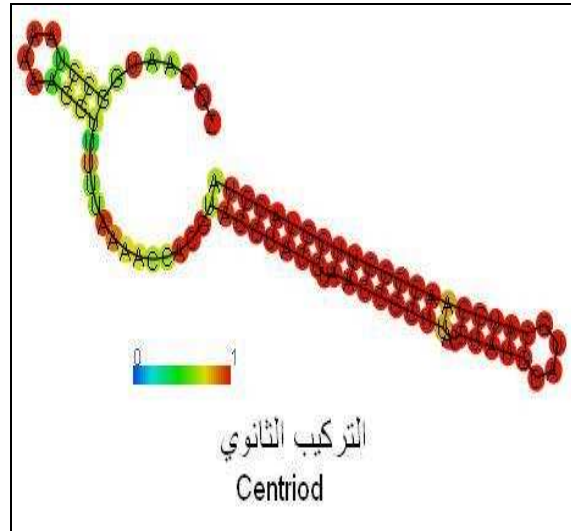
شكل 147.5 : توضيح المقارنة باستعمال برنامج الشعار Logo

ويعطى البرنامج تراكيب معتمدة على الطاقة ومناطق الازدواج باستعمال الألوان التوضيحية لذلك كما موضح في الشكل التالي (148.5) الذي يوضح تركيب جزيئة tRNA ala



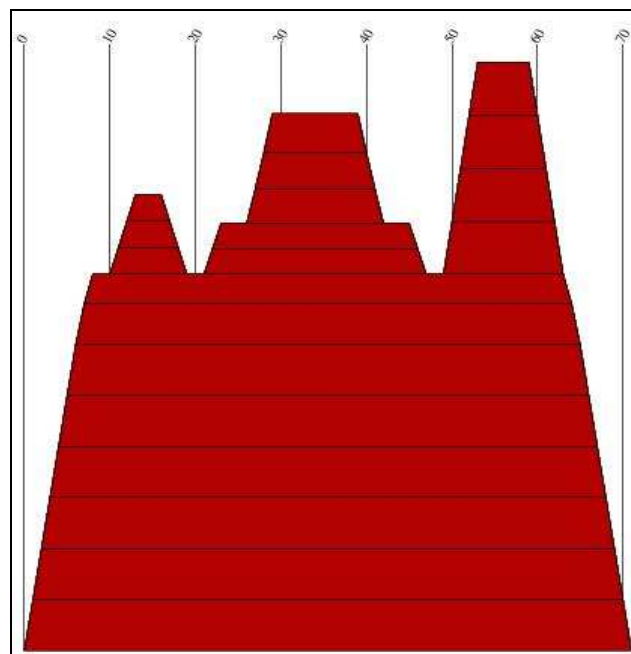
شكل 148.5 : التركيب الثانوي لجزيئة tRNA ala

ويظهر البرنامج Centroid structure الذي يمثل التركيب الثانوي وباستخدام مؤشرات اللون كما موضح في الشكل 149.5 :



شكل 149.5 : التركيب الثانوي Centroid

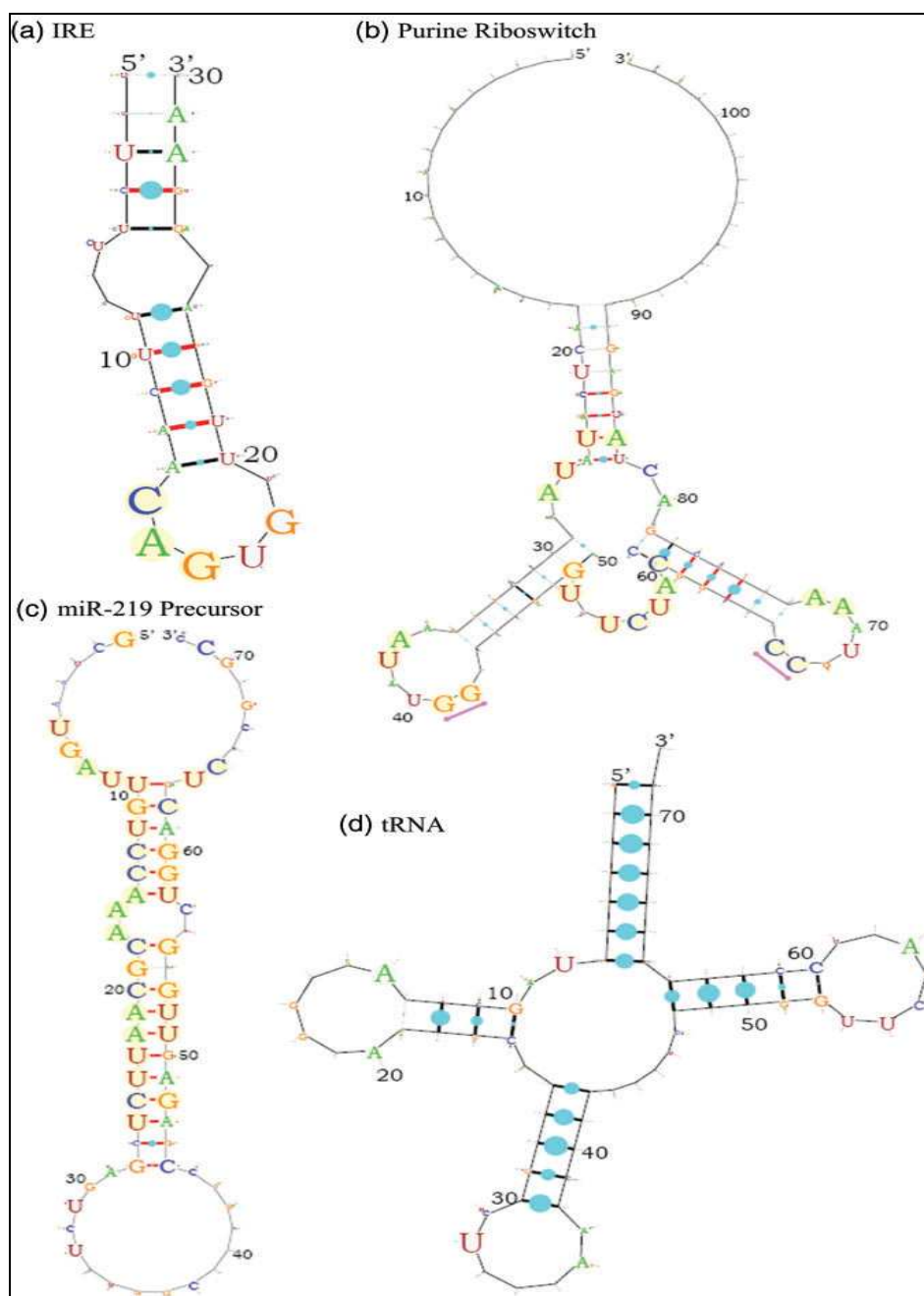
وتظهر البرامج رسم الجبل Mountain plot الذي يأخذ بنظر الاعتبار العديد من المؤشرات الإحصائية ويوضح بعض المؤشرات كما في الشكل التالي (شكل 150.5) وكذلك الانتروبي عند كل موقع .



شكل 150.5 : رسم الجبل التمثيلي لجزيئات RNA

ويمكن ان تمثل كل فقرة من الرسم احد التراكيب الثانوية وعلى العموم فكل برنامج يختص بالمرجات التي يظهرها مع توضيحها.

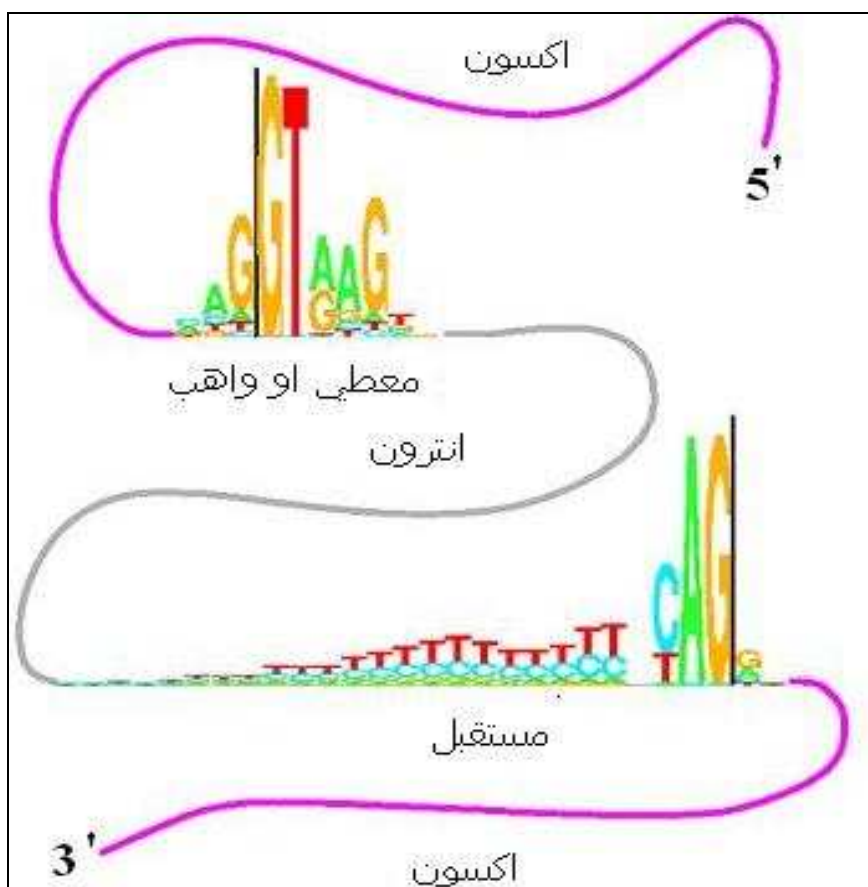
ومن البرامج الحديثة RNALogo الذي يأخذ العديد من المؤشرات حول تواليات RNA بالحسبان وباستعمال برامج أخرى وفي خطوات متعددة تظهر مخرجات مختلفة كما موضح في الشكل 151.5 :



شكل 151.5 : بعض تراكيب جزيئات RNA كما تظهر باستعمال برنامج RNALogo

ويكون لطبيعة الرسم علاقة بمواصفات الجزيئات ، اذ تشير الحروف الكبيرة الى ثبوت الثمالة أثناء التطور والألوان تعني مواصفات خاصة وكل هذه يمكن الرجوع اليها الى الملفات المساعدة المرافقة للبرنامج .

ويمكن ان تستعمل برامج الشعار في تحديد بعض المواقع كما في تحديد المناطق الفاصلة بين الانترونات والاكسونات في جزيئات Pre-mRNA التي يمكن ان تنتج من الانتساخ كما موضح فش الشكل 152.5 :



شكل 152.5 : إيجاد مواقع الانفلاق باستعمال برنامج الشعار Logo

تقييم أداء البرامج

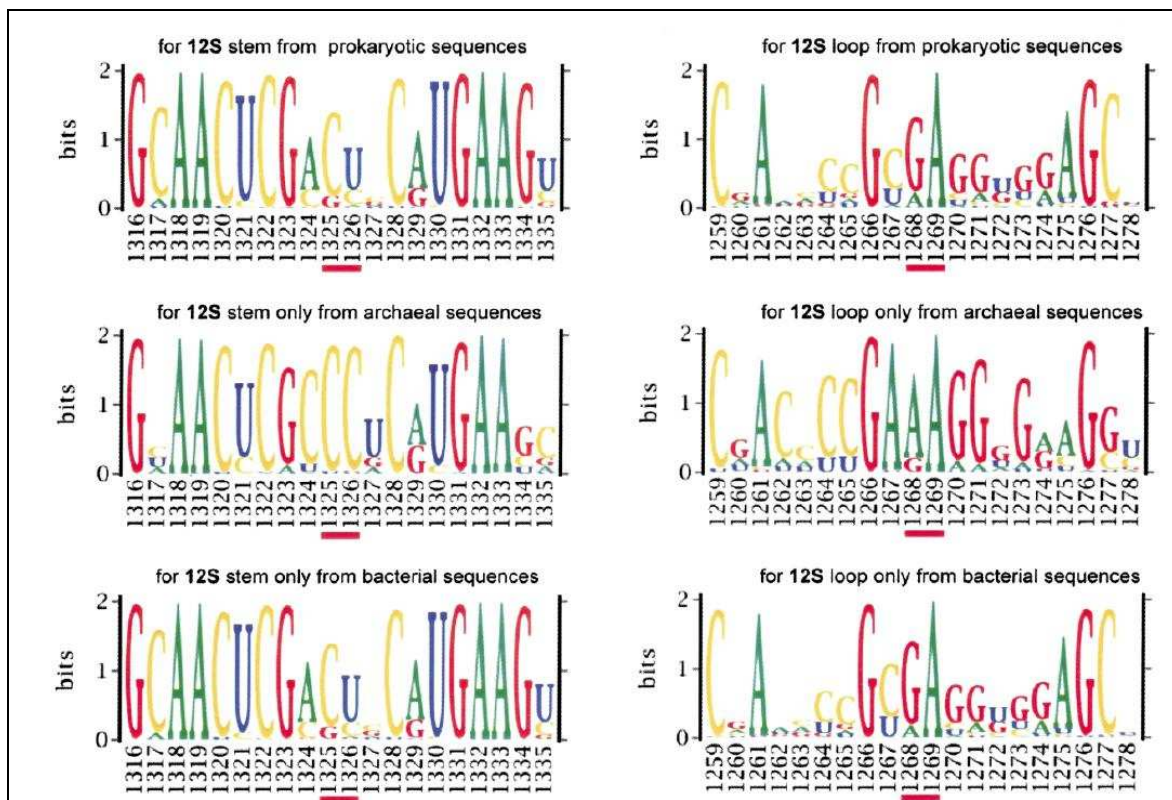
تختلف دقة أداء البرامج اعتماداً على الأسس المستعملة في تصميمها والمعلومات المتوفرة حول التواليات ، لذلك كان التقييم الصارم لأداء الحدى يعاق بشكل تقليدي بقلة المعلومات حول التراكيب خاصة الثلاثية لجزيئات RNA . ومن جهة ثانية فهناك الطرق الفيزيائية التي حلت بعض الجوانب الغامضة باستعمال التراكيب البلورية وخاصة للرايبوزومات ومن ورائها rRNA . وتم ايضاح بعض التفاصيل حول التراكيب التي لها علاقة بأنواع مختلفة من جزيئات RNA . وعند إجراء تقييم لدقة الحدى فبعض الطرق تعتمد مؤشراً واحداً مثل معامل

الارتباط (r) الذي يحسب بين نتائج البرامج والوقائع الحقيقية ومعامل الارتباط يؤخذ بالحسبان كل المعلومات حول الحساسية والانتخابية ، ونتائج التقييم أشارت الى ان طرق البداية تسجل دقة تعتمد على طول التوالي فهي بالنسبة للتواليات الطويلة تقع بين 20 - 60 % ولكن بالنسبة للتواليات القصيرة فتصل دقتها الى 70 % .

ومن جهة ثانية فأن تحديد دقة البرامج المعتمدة على مؤشر واحد مثل مؤشرات الطاقة لوحدها يكون صعباً وذلك لان المؤشر المستخدم (الطاقة) يكون غير كافياً للتمييز بين التراكيب المختلفة المحتملة للجزئية الواحدة . ومن جهة أخرى تتأثر دقة أداء بعض البرامج لأنها تعتمد على نتائج برامج أخرى مثل تلك التي تستعمل تواليات مصطفة جاهزة والتي يكون فيها احتمالية الخطأ بدرجة لا يمكن غض الطرف عنها ، وتزداد نسبة الخطأ في الاصطفاف وكما معروف مع التواليات الطويلة ، وكذلك الحال مع برامج المقارنة . ولكن دقة الحدس باستعمال التوجيهين يمكن ان يصل الى أقصى درجة 100 % مع التواليات القصيرة مثل tRNA غير المتداخلة . وعليه يمكن تفضيل توجه المقارنة على توجه طرق البداية . ومن الجدير بالذكر ان معظم البرامج المطروحة والمستعملة تتجاهل العقد الكاذبة في تراكيب RNA لأنها تعقد عمل برنامج الحاسوب ، ولذلك كان لابد من إجراء التحسينات على البرامج لتجاوز هذه العقبات كما حصل في برنامج RNALogo المذكور أعلاه .

البرامج الملونة في مجال دراسة RNA

للبرامج الملونة دور كبير في إظهار دقائق التفاصيل فمثلاً ثبوت تراكيب RNA مقارنة بالتوالي الحاوي على الطفرات التعويضية اذ تستدعي هذه الحالة إظهار مواقع هذه الطفرات بألوان تختلف عن باقي التوالي كما ان النيوكليوتيدات غير القادرة على الازدواج يجب ان تظهر بلون مختلف مثل لون باهت مقارنة بباقي التوالي ولذلك تستعمل برامج خاصة للتلوين مكتوبة بلغات برمجة مختلفة ويمكن ان تعطي ملفات للمخرجات مختلفة مثل pdf او ESP (التابع لبرنامج GSview) ، وأشهر البرامج الملونة هي مجموعة الشعار مثل WebLogo و SequenceLogo الذي يمثل احد برامج الشعار الخاصة بالتواليات و CorreLogo و enologo . ومعظمها يظهر النتائج بشكل منضد فالشكل التالي (153.5) يوضح مخرجات عملية اصطفاف عدد من التواليات .

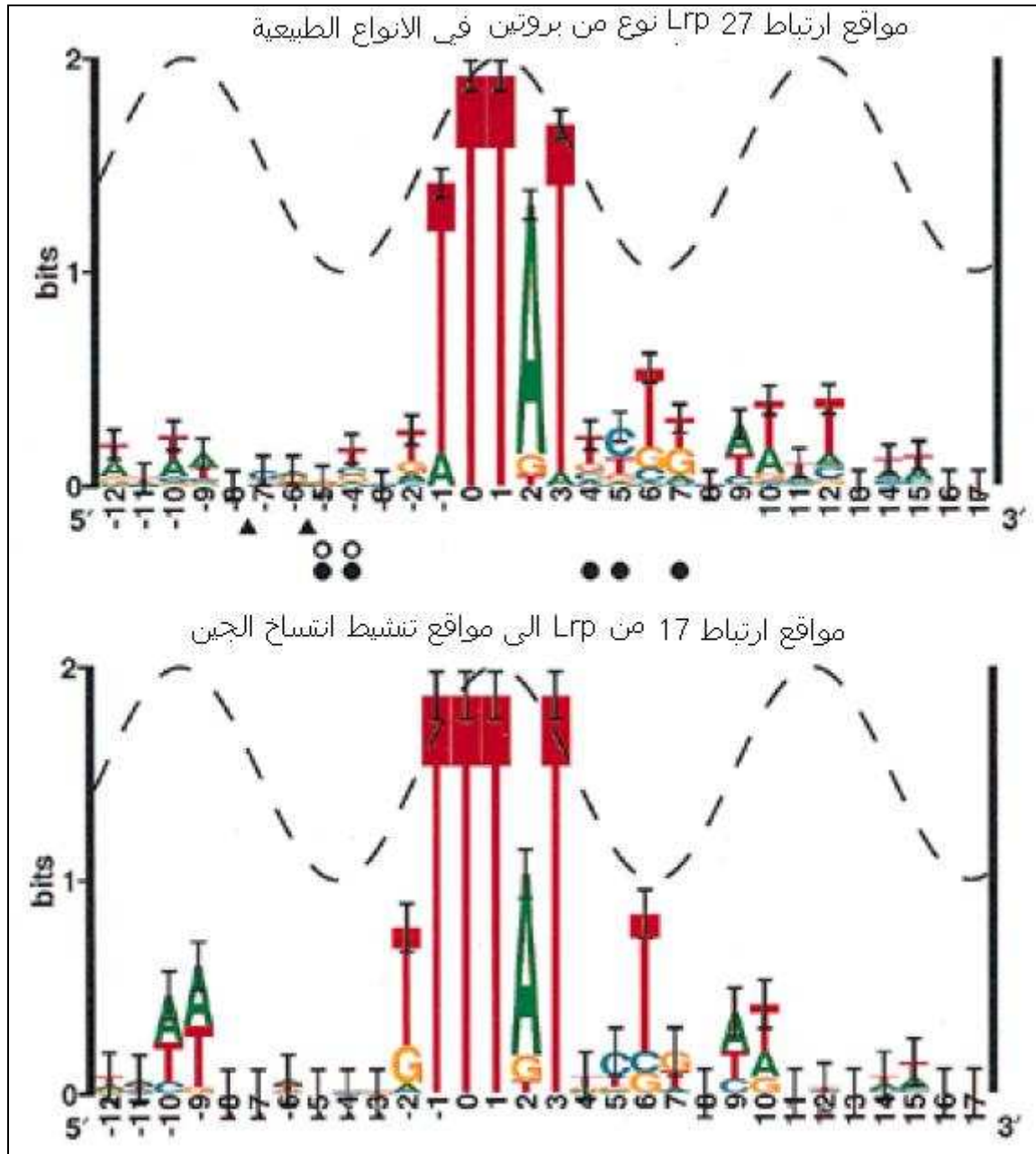


شكل 153.5 : مخرج احد برامج Logo لاصطفاف عدد من تواليات RNA

يمثل الشكل اصطفاف تواليات RNA . يكون علو كل نضد في الموقع الواحد ممثلاً لمحتويات التوالي من القواعد او الثمالات التي تحدد بطرق ومعادلات معينة متضمنة في البرنامج مثل قياس قيمة Rseq التي تعطي الفرق بين الانتروبي الأعظم والآخر الذي تم ملاحظته ، وتعتمد على تردد الثمالة او القاعدة في تواليات الاستعلام التي تمثل بعلو كل حرف او ثمالة . وفي العموم تكون القيم الصادية مقاسة بـ bits التي تكون في العادة 2 وبعض الأحيان تصل الى 4 في حالة البروتينات .

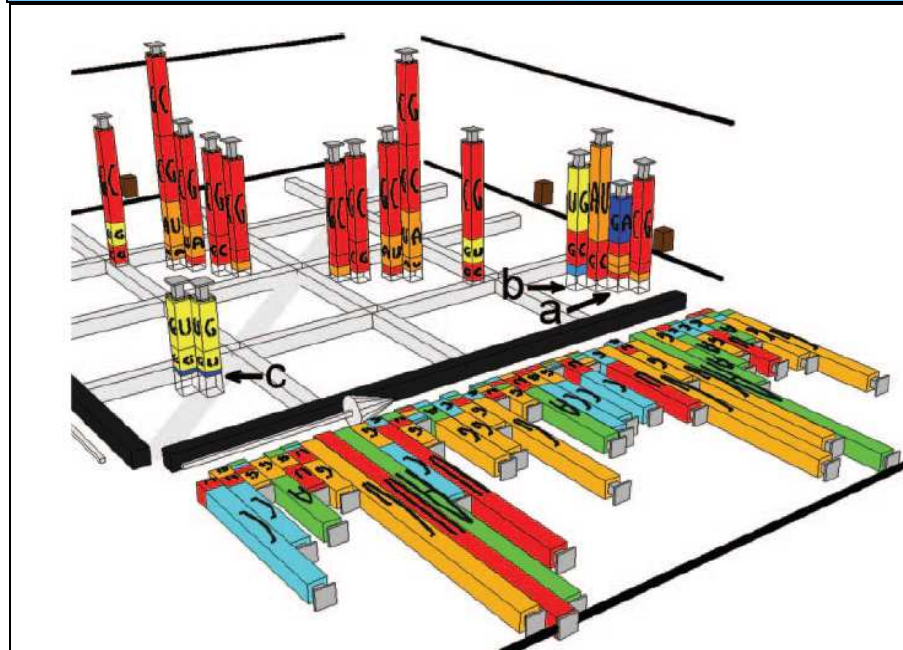
وبرامج الشعار يستعمل لإيضاح مواقع الارتباط او غيرها من المؤشرات لعدد من التواليات كما موضح في الشكل (154.5) بالنسبة للبروتين المنظم Lrp الشائع في بعض اجناس العائلة المعوية من البكتريا الذي يمكن ان يرتبط بعدد من الممهدات .

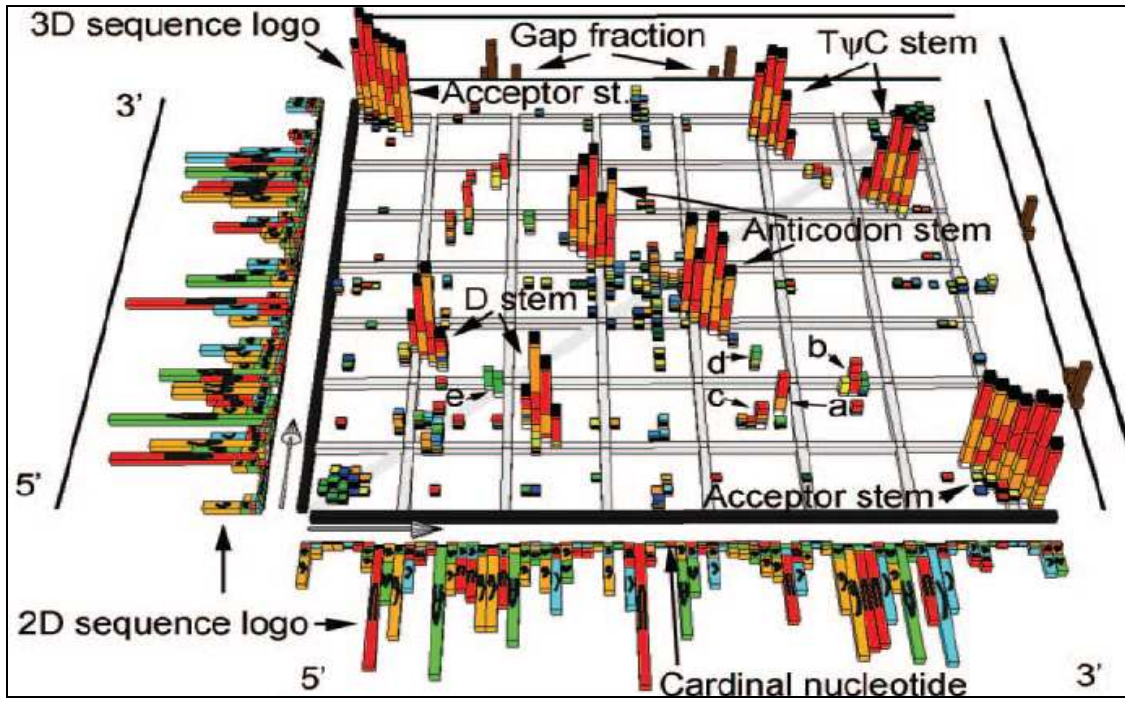
اما enologo فيستعمل لرسم تواليات RNA والمخرجات تشمل قياس الطاقة ومصنفات الاحتمالية ومصنفات الاصطفاف لتمثيل المعلومات المشتركة للمواقع المختلفة في التواليات ، والصورة التي ينتجها مثل الأشكال الخاصة بـ SequenceLogo و WebLogo المذكورة أعلاه .



شكل 154.5: مخرجات برنامج Logo لتوضيح مواقع الارتباط في بعض الجينات

CorreLogo برنامج يربط بين محتوى البيانات حول التواليات والمعلومات المشتركة لمواقع مختلفة وينتج رسم ثلاثي الأبعاد لإيضاح العلاقة بين القواعد وإمكاناتها للارتباط ، والشكل 155.5 يوضح بعض هذه المخرجات :





شكل 155.5: مخرجات برنامج CorreLogo المجسمة لاصطفاف جزيئات RNA

الفصل السادس
المعلوماتية الحيوية وبعض جوانب الحياة

المعلوماتية الحيوية وعمليات الايض
المعلوماتية الحيوية والبيئة
المعلوماتية الحيوية والجوانب الصحية للإنسان
مكون الإنسان الميكروبي Human microbiome
المعلوماتية الحيوية والجهاز المناعي
الحساسية Allergy
الأغذية المهندسة وراثيا

المعلوماتية الحيوية وبعض جوانب الحياة

بعد استعمال المعلوماتية الحيوية في دراسة جزيئات الحياة من مختلف الجوانب ، جاء دور استعمال المعلوماتية الحيوية في جوانب أكثر تعقيدا من جوانب الحياة والفعاليات الحيوية ومنها الجوانب الايضية والجوانب البيئية والصحة وغيرها أي ان التوجه انحرف ليتعامل مع العلاقات والتداخل بين جزيئات الحياة وذلك يعود الى التطور الكبير في الطرق السريعة لأداء التجارب ومن هنا بدا إنشاء قواعد بيانات خاصة بالمسارات الايضية وليس بالجزيئات المشاركة فيها .

المعلوماتية الحيوية وعمليات الايض

مسارات الايض تقع في صميم اهتمام المعلوماتية الحيوية ويمكن تعريفها على إنها مجموعة من التفاعلات الموجهة تتداخل مع بعضها تحت الظروف الفسلجية عن طريق جزيئات ومركبات وسطية بسيطة بتركيز مختلفة في المسار الايضية ، فارتفاع تركيز مادة معينة يعني انها تكون في نقاط تقاطع أو تفرع ، فمثلا في نظام تحليل السكر فان دورة TCA تعد إحدى نقاط التفرع ويحصل منها ضخ للمركبات لتذهب الى تفاعلات جانبية .

والجزيئات الحيوية التي تناولتها الفصول السابقة تعمل ضمن شبكات الايض، لذلك فان دراسة المسارات الايضية توفر الفرص للوصف الجيد للشبكات والدوائر الخلوية . ومقارنة مسارات الايض يعطي الفروق الايضية بين الأنواع ، ومقارنة المسارات الايضية تحتاج الى الجمع والربط بين تأثير عدة مسارات ايضية تحت شبكة موحدة وكذلك الكيفية التي يحل بها التطور المشترك للمسارات الايضية ، ومن هذه يمكن إيجاد العلاقات . فالظروف البيئية يمكن ان تملئ وجود مسارات ايضية معينة ومن هذه المقارنات يمكن اكتشاف مسارات ايضية جديدة ، ولإجراء المقارنات كان من الضروري إيجاد قواعد بيانات للمسارات الايضية وهي بطبيعة الحال تكون مرتبطة بقوة مع قواعد بيانات الإنزيمات باعتبار الأخيرة هي جوهر حدوث المسارات الايضية . فضلا عن الحاجة الى التعريف بالجزيئات والمركبات الداخلة ضمن المسارات وتحديد كميتها في المسارات الايضية .

قواعد بيانات الايض وطرق الاصطفا

من المعروف (على الأقل في الوقت الحاضر) ان البيانات في قواعد بيانات الايض تكون متداخلة مع بعضها فالبروتينات يمكن ان توجد في القواعد الخاصة بالجينومات وكذلك فان الإنزيمات يمكن ان يدرج بعضها في قواعد البروتينات او الإنزيمات وكذلك يتفرق البعض الآخر في قواعد بيانات المسارات الايضية . ومن جهة ثانية فان التناسق بين التعبير الجيني بين مسارات الايض الخاصة هي التي تحدد نوعية المتأيضات التي يمكن ان تخلق وبالتالي ومثل هذه يمكن ان تستعمل في الكشف عن وضعية الشبكات الخلوية ويمكن الوصول إليها من دراسة مكونات قواعد البيانات المختلفة .

وقد وجد من الدراسات انه حتى في الأحياء البسيطة تكون الشبكات الخلوية معقدة جدا والذي يمكن ان يخفي المسارات الايضية الرئيسية التي تحدد بالاستجابة للمؤثرات الخارجية .

وابسط الطرق في هذه الحالة استعمال قياس اتحاد العناصر المتكافئ الكميات Stoichiometry لكل تفاعل ولكن هذا التوجه يحدث في مسارات قليلة من مسارات أخرى التي تظهر في النهاية الشبكات الكلية . وهناك بعض التوجهات التي تستعمل في مثل هذا المجال .

• تحليل التوازن الدفقي Metabolic flux balance analysis (FBA) توجه يستعمل ويعتمد التحليل هذا على الموديلات أو النماذج الفيزيائية للكيمياء التي تسيطر على التدفق في التفاعلات لتحديد كميات المواد الناتجة وهي طرق تعتمد على المواصفات الكيماوية للشبكات ولا تدخل فيها التأثيرات الوراثية . ولذلك تستعمل طرق وتوجهات لتحليل الشبكات الثانوية Subnetworks لترتيب المسارات بشكل أفضل . وفي هذه الحالة يعد الجين هو العقدة Node التي تعتمد عليها عدداً من البرامج مثل Markov mixture model (3M) وتحويراتها المختلفة مثل برنامج PathRanker الذي يكون ضمن حزمة أو مجموعة من البرامج الأخرى التي تستعمل بشكل خاص في تحليل ايض الاجهاد وغيرها من الجوانب .

مما يساعد في التوجه FBA ان كل من تحديد تواليات الجينومات واستعمال المعلوماتية الحيوية في تحليلها وفرت تفاصيل كثيرة حول المكونات الجزيئية الموجودة في العديد من الأحياء بدائية النواة — كمرحلة أولية — وساهمت في وضع بيان مصور Parts catalog لأجزاء الخلية الميكروبية .

واستعملت إمكانية المعلوماتية الحيوية لوضع تمثيل افتراضي للوظائف الايضية الأساسية بحيث يمكن ان يعاد تركيبها وتحليلها باستعمال توازن الدفق الأيضي ويكون تحليل توازن الدفق الأيضي ملائماً لدراسة الشبكات الايضية وقد استعمل لتفسير الإمكانات الايضية للـ

E. coli حيث أمكن وضع خرائط حاسوب للبيكتريا وتحديد أفضل استهلاك للمواد في المسارات الايضية كدليل للمتغيرات البيئية . وباستعمال التحليل الحاسوبي تم التعرف على ان هناك عدد محدد من الجينات خاصة بالايض المركزي مثل تحلل السكر Glycolysis ودورة الأحماض ثلاثية الكربوكسيل TCA ومسار فوسفات السكريات الخماسية Pentose phosphate ونظام نقل الالكترونات التي تكون ضرورية للنمو الهوائي للبيكتريا في وسط غذائي أدنى Minimal medium ، وأدى الى التعرف على ان هناك خمسة عشر جينا تكون ضرورية للايض اللاهوائي . كما أمكن التنبؤ بعدد من الطفرات الافتراضية ورسم الخرائط لها وقابليات هذه السلالات المتناظرة Isogenic strains ، لذلك فان مثل هذه النتائج تمكن من فهم اكبر للعلاقة المعقدة بين النمط الجيني والنمط المظهري

ومن هذا أمكن تصميم كائن مثل *E. coli* على الحاسوب سميت *In silico E. coli* بالاعتماد عن تواليات الجينومات المعلنة والمراجع وقواعد البيانات . وقد وجد ان السلالة الافتراضية *In silico E. coli* وبعد مقارنتها بصفات *E. coli* الحقيقية أو الطبيعية Wild type انه يمكن تفسير تصرف عدد من السلالات المطفرة من البيكتريا ، اذ ان استعمال الجينات يكون معتمداً على مصدر الكربون وتوفر مواد الأساس الأخرى أي ان النمط المظهري للطفرات يعتمد على العوامل البيئية أيضاً .

وقد استعملت التوجهات أعلاه باستعمال *In silico media* ، وبعدها يمكن حذف احد الجينات *In silico gene deletion* من مسارات الايض المركزي لإيجاد التفاعلات التي تجري تحت إمرة هذا الجين ، ومن هذه النتائج يمكن معرفة وحساب الاحتياجات Auxotrophic requirements التي تؤثر في الشبكة الايضية وتسجل كجزء او نسبة من القيم المثلى للسلالات الطبيعية . وقد ساعدت طرق الحاسوب إيضاح الارتباط بين الظروف

البيئية والاستهلاك الأمثل للمواد في المسارات الابيضية مثل معرفة تأثير حذف الجين في كل الأطوار التي تشارك فيها مادة أساس مثل الكلوكوز تحت الظروف الهوائية ، وكذلك إجراء المقارنات بين سلالات الحاسوب المتناظرة *Isogenic in silico strains* . ومن هذه الدراسات أمكن تقسيم الجينات الى مجاميع :

- الجينات الضرورية *Essential genes* والتي تكون نواتجها ضرورية للنمو والفعالية تحت ظروف محددة ، بالاعتماد على نواتج الجينات المدروسة .
- الجينات الحرجة *Critical genes* وهي الجينات التي عند غيابها يقل حاصل النمو الى اقل من 95 % مقارنة بالسلالات الطبيعية .
- الجينات غير الضرورية *Non-essential genes* والتي عند حذفها يكون النمو مقارب للأنواع الطبيعية ويكون حاصل النمو اكبر من 95% تحت ظروف معينة فمثلا الجينات اللازمة لتكوين ATP تكون ضرورية تحت الظروف الهوائية باستعمال طريقة الفسفرة التأكسدية في حين يتم الاستغناء هذه الجينات تحت الظروف اللاهوائية التي يتم فيها تخليق ATP بطريقة الفسفرة بمواد الأساس .
- الجينات الضرورية الظرفية *Conditionally essential genes* وهي الجينات اللازمة والضرورية تحت ظروف معينة .

وفي جميع الأحوال تكون صفة الضرورة أو عدمها معتمدة على طور النمو . ومما لا شك فيه فان تحديد التواليات الجينومية حمل في طياته القابلية الكبيرة على تحديد البصمة الوراثية *Blue prints* لحياة الخلية ، فهو يحوي على المعلومات الخاصة بالتكاثر وغيرها من الفعاليات التي تكون موجودة في أنظمة اطر القراءة المفتوحة وما يتبعها من معلومات تنظيمية للوظائف ، والتوجه الحاسوبي تطور ويتطور باستمرار لاستثمار أو استنتاج المعلومات المشفرة في DNA ولكن مع هذا فان الشواهد تشير الى تعقيد الأنظمة والوظائف الحيوية الأساسية التي تشارك فيها التدخلات بين المكونات الحيوية داخل الخلايا. ولفهم التعقيد المستورث في الشبكات الخلوية يتم التأكيد على الحاجة الى صفات عامة للشبكات الخلوية لتكون المنطلق . كما ان تعقد الأنظمة الخلوية الحية أدت الى مفهوم واضح هو انه لا يمكن دراسة العمليات الحيوية أو تحليلها أو التنبؤ بها بالاعتماد على وصف ودراسة مكون واحد وهذا يعني الابتعاد عن الطرق التقليدية في علوم الحياة والتوجه الى طرق الدمج والتكامل لفهم العلاقات بين وظائف الجينات ودور كل جين في حالة الوظائف الوراثية المتعلقة بوظيفة الخلايا أو ما يطلق عليه بالدوائر الوراثية *Genetic circuits*.

ولبناء طريقة هندسية لإجراء التحليل أو تصميم الأنظمة المعقدة لا بد من وجود موديل رياضي أو حاسوبي مثل المحاكاة الديناميكية للعمليات الخلوية التي تعتمد أساسيات قوانين الفيزياء الكيماوية. وفي ستينيات القرن الماضي وضعت موديلات لشبكات الايض من الأنظمة الحيوية ولكنها لم تلاق النجاح نظرا لعدم وجود المعلومات الحركية حول ديناميكية تنظيم الايض ولكن مع هذا وفي ظل غياب المعلومات الحركية يمكن الوصول الى قابليات وإمكانات وان كانت نظرية لاستعمال توجه توازن الدفع الأيضي .

وتوجه توازن الدفع الأيضي يمثل الأفضل فيما اذا كان متوافقا مع بيانات التجارب المختبرية خاصة تحت ظروف ملائمة مثل التغذية الجيدة ، اذ انه يأخذ بنظر الاعتبار عمليات التنظيم بحيث يكون التصرف الأيضي في حالته المثلى وهذا الافتراض أدى الى الحصول على نتائج متوافقة مع بيانات التجارب المختبرية ، ومن الواضح ان هذه التوافقية تضطرب عند حصول الطفرات. وبإتباع هذا النهج وبوجود مصفوفات التكافؤ الكيماوي *Stoichiometry matrices* يمكن الكشف عن معلومات جديدة حول الايض وكذلك التعرف على الوظائف المفقودة أو المعرفة بشكل خاطئ والدراسات في هذا الحقل (تمثيل الايض بشبكات على الحاسوب) مثلت احد الحقول السريعة التطور في البيولوجي الحاسوبي *In silico biology* ، اذ مكنت من تحديد السلالات التي تم الإطاحة بأحد جيناتها *Well-defined knockout* للـ *E. coli* من فتح الآفاق أمام إمكانية التقييم المهم والخرج بين التمثيل الحاسوبي لتصرف الطفرات والايض وشبكات الايض الحيوي داخل الخلايا تحت ظروف وراثية وبيئية ، وتشكل إستراتيجية مهمة لاختبار الجينات ويزود بنظرة فاحصة في الايض الحيوي ، ومثل هذا يفيد في عمليات انتخاب السلالات المستعملة في عمليات الإنتاج الحيوي ، وخاصة ان له القابلية على هندسة الايض بالاعتماد على تحول الكتلة ، وإخضاعها للحسابات الرياضية للوصول الى أفضل النتائج .

• **التوجه الآخر في دراسات الايض هو إجراء عمليات الاصطفاف *Metabolic pathway alignment* وهذا يتطلب وجود قواعد بيانات للمسارات فضلا عن ضرورة التعريف بالمسارات ، فهناك عمليات ايض بنائي *Anabolism* وعمليات ايض هدمي *Catabolism* فيوصف المسار على انه هدمي اذا كان الوزن الجزئي والكتلة الجزيئية للنواتج اقل من الوزن الجزئي والكتلة الجزيئية لمواد الأساس الأولية ، وفي حالة العكس فيوصف بأنه بنائي . ولكن بعض الأحيان تكون الحالة مختلفة ، فمثلا تفكيك مواد عضوية معقدة يمكن ان يؤدي الى تكوين *Acetyl – CoA* أو *Succinyl – CoA* وهي أثقل من مواد الأساس الأولية ، فضلا عن انها مواد تفرع وسطية وليست نواتج نهائية لمثل هذه المسارات الايضية ، لذا فهي تصنف على انها مسارات بنائية وهذا يخالف الوظيفة الظاهرية لعملية الهدم ، ولذلك غيرت تسميته لبعض المسارات الايضية لغرض إدخالها في عمليات الحاسوب . وفي هذا التوجه يتم وضع البوم الكتروني للخريطة الايضية والتي تكون على خطوات أو أطوار :**

• **الطور الأول:** يتم البحث عن اطر القراءة المفتوحة التي قد تم تحديد توالياتها ووظائفها

• **الطور الثاني:** يتم تعيين الوظائف التي تلاؤم المسار وتضبط المسارات الايضية والوظائف مثل ضبط الموديلاات الرياضية والتي تسبقها عمليات متكررة لغرض الوصول الى أفضل موائمة *Matching* لبيانات التواليات المتوفرة والاستعانة بالكمياء الحيوية والمعلومات عن النمط المظهري ، فضلا عن الحاجة الى تحديث المسارات للحصول على المسارات الجديدة والإصدارات الجديدة الناتجة من تحديث إصدارات قديمة . وتراكم المعلومات الكيماوية الحيوية ومعلومات عن النمط المظهري ساعد في الكشف عن الجينات المفقودة وتم إيضاح ان الوظائف معتمدة بعضها على البعض الآخر بين بروتينات المسار الواحد وكذلك الاعتماد المتبادل بين المسارات في كل عمليات الايض الخلوي ، وقد تم توزيع المسارات الايضية وفقا لطولها أي عدد الخطوات ، فالمسارات التي طولها أكثر من خطوة تشكل نصف مجموع المسارات ، ولكن طول المسار يكون مقياسا ضعيفا للاعتماد الوظيفي المتبادل بين البروتينات

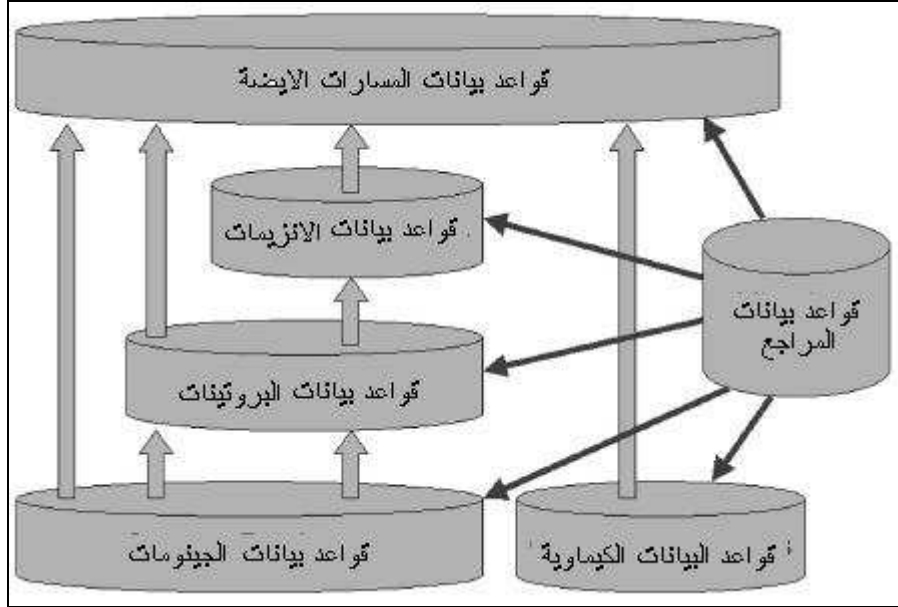
نظرا لان عدد من المسارات ذات الخطوة الواحدة مثل آليات النقل عبر الأغشية يمكن ان تعتمد على تناغم وانسجام تأثير مجموعة من البروتينات ، لذلك فان قوة أو كفاءة الحدس لعنقود الوظيفة يعتمد على عدد البروتينات في ذلك العنقود . ومن السهولة ملاحظة ان العنقدة الوظيفية للبروتينات تمتد الى ما بعد المسارات الايضية وحدودها ، وعليه فان كل مركب في الخلية لابد ان يكون له مصدرا " ومغطسا" وهذه الاحتياجات تشير الى الحاجة الى عدد كبير من العناقيد الوظيفية التي تتكون من عدة خطوات في المسار الأيضي ، وعلى العموم فان هذه الدراسات بعيدة عن الكمال على الأقل في الوقت الحاضر. وفي ضوء البيانات والإيضاحات أعلاه انطلق مفهوم دراسة مكنون الأيضي Metabolomics .

دراسة مكنون الايض

هذه الدراسات تكون قريبة جدا من دراسات مكنون الجينومات وكذلك مكنون البروتينات من مختلف النواحي ، وذلك لان دراسة المكنون الأيضي يرتبط ارتباط وثيق بالنمط الجيني للكائن وكذلك فسلجته وبيئته أي ماذا يأكل وماذا يتنفس وعليه فهو يوفر فرصة فريدة للنظر في العلاقات بين النمط الجيني – النمط المظهري Genotype – phenotype وكذلك النمط الجيني – مع النمط البيئي Genotype – envirotype . وتستعمل دراسة مكنون الايض في العديد من التطبيقات تأتي الجوانب الصحية في مقدمتها ومنها إيجاد الأدوية وإجراء التجارب الأولية عليها وكذلك دراسة السموم ونقل الأعضاء والكيمياء السريرية بشكل عام .

ومن الواضح ان مقارنة واصطفاف المسارات الايضية يعطي نتائجاً حول تشابه الأحياء على الصعيد الوظيفي . فعند صف المسارات تنتج درجات تشابه ودرجات اختلاف بين المسارات الايضية وإيجاد العلاقات والدرجات Scores للاصطفاف تشير الى التغيرات التطورية التي حصلت للمسارات الايضية وبالتالي يوضح العلاقات التطورية بين الأنواع باستعمال الصفات الوظيفية .

وعلى غرار ما ذكر في عمليات الاصطفاف السابقة فأن اصطفاف المسارات الايضية يحتاج الى قواعد بيانات للمقارنة معها. وقواعد بيانات الايض تحوي على المسارات الايضية وكذلك مسارات نقل الإشارات والإنزيمات والبروتينات وبيانات الجينوم وفي العموم يكون الهيكل العام لمثل هذه القواعد كما موضح في الشكل 1. 6:



شكل 1.6 : أساسيات قواعد بيانات الايض

والملاحظ ان قواعد البيانات الخاصة في هذا المجال تشمل قواعد الايض مثل قاعدة كيوتو KEGG و EcoCyc و MetaCyc وغيرها والمسارات المذكورة في موقع ExpASY المارة الذكر وترتبط بقواعد خاصة بالجينومات مثل بنك الجينات وقواعد بيانات الموقع الأوروبي EMBL والياباني DDBJ والقواعد الخاصة بالأحياء مثل الأحياء المجهرية MGDB وقاعدة بيانات ذبابة الفاكهة FlyBase وغيرها . فضلا عن ارتباطها بقواعد مثل قاعدة البروتينات السويسرية و TrEMBL وبنك بيانات البروتينات PDB وعدد آخر من القواعد ولا بد ان ترتبط قواعد البيانات هذه مع قواعد البيانات الخاصة بالإنزيمات مثل BRENDA و ENZYME وغيرها ، وتكون الحاجة ملحة ان ترتبط هذه القواعد بقواعد بيانات كيميائية توفر الصفات الكيميائية للجزيئات مثل ChemFinder وغيرها ومن الطبيعي انه ترتبط بقواعد للمراجع العلمية التي وفرت المفردات المدرجة في القاعدة مثل قاعدة المراجع .

وتختلف قواعد بيانات الايض الموجودة في الوقت الحاضر عن سابقتها في جودتها ومعلوماتها والتسميات المستعملة فيها وتعتمد نظام الألوان في التأشير لمحتوياتها ، فالبعض يستعمل ألوان خاصة لكل مجموعة من الأحياء مثلا استعمال اللون الأخضر لمسارات ايض النباتات واللون الأزرق للحيوانات والأحمر للأحياء المجهرية واللون الأسود للمسارات الايضية العامة ، فضلا عن توفيرها خرائط للتفاعلات الإنزيمية وغير الإنزيمية ، وكل قاعدة تحوي على وسائل لإجراء الفحوص المختلفة ومنها وسائل المحاكاة Simulation tools .

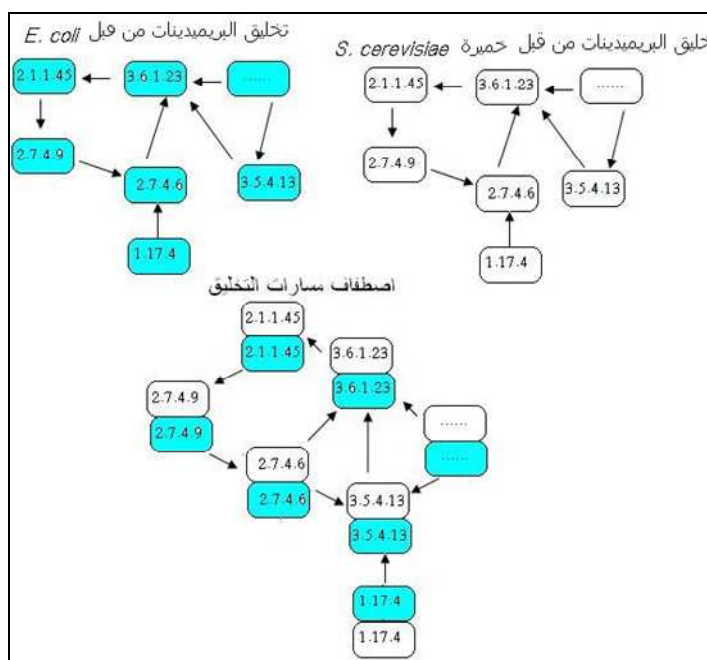
وبصورة عامة فان عمليات الاصطفاف تحتاج الى مقارنة مجالين من البيانات :

الأول : طبوغرافية المسارات

الثاني : التغير في محتويات المسار من الإنزيمات لغرض إيجاد التشابه وليس التماثل ، اذ ان بعض الإنزيمات تتوافق في المهام التي تؤديها ولكنها لا تتوافق من حيث تواليات حوامضها الامينية .

وهذان يؤثران في عمليات الاصطفاف والتصنيف الصحيحة. وهناك العديد من البرامج المستعملة لغرض صف المسارات الايضية البعض منها يختص بمسارات ايض تفكك المواد مثلا من قبل الأحياء المجهرية ، والآخر بتخليق مواد الايض الثانوي خاصة في النباتات كما في برنامج PathPred .

وفي العموم ترتبط البرامج بقواعد الايض التي يجب ان تكون ذات نوعية جيدة ومدققة ومشذبة مثل EcoCyc وغيرها وأكثرها استعمالا قاعدة كيوتو. وفيها تمثل الإنزيمات بالعقد والتفاعلات بخط يربط هذه العقد لتكون مطواعة للاستعمال في الحاسوب كما موضح في الشكل (2.6) .

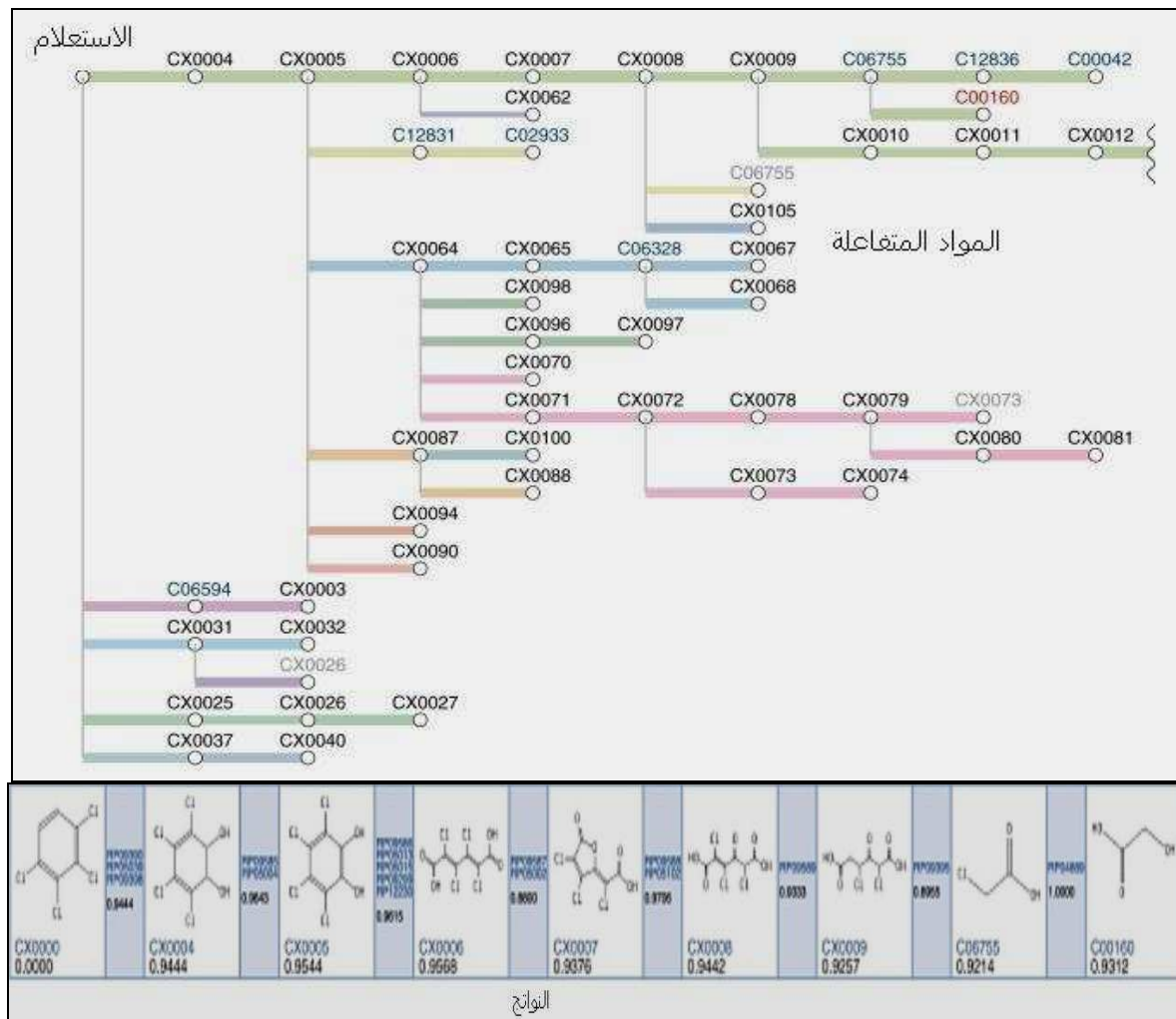


شكل 2.6 : تمثيل الإنزيمات وخطوات التفاعل عند صف المسارات الايضية

وعند عدم معرفة إنزيم إحدى العقد يقوم البرنامج بوضع مقترح للإنزيم . والإنزيمات في العادة تمثل بأرقام EC لها وبذا يمكن البحث عن الإنزيم المفقود من رقم EC له . اما درجات التفاعل فيشار لها بسمك الخطوط الموجودة في الشجرة . وفي حالة عدم وجود تواصل في الشبكة الايضية ينشأ البرنامج شبكة ثانوية Subtree .

وفي برامج أخرى تكون المركبات هي الممثلة بالعقد وتعطى أرقام خاصة في العديد من البرامج لتكون مقبولة في الحاسوب مثل COO160 للـ Glycolate. ويشار الى المركبات

بـ CX يليها رقم عندما تكون هذه المركبات غير معرفة بقاعدة البيانات ذات الاستعمال العام مثل قاعدة كيوتو ، كما موضح في الشكل 3.6 :



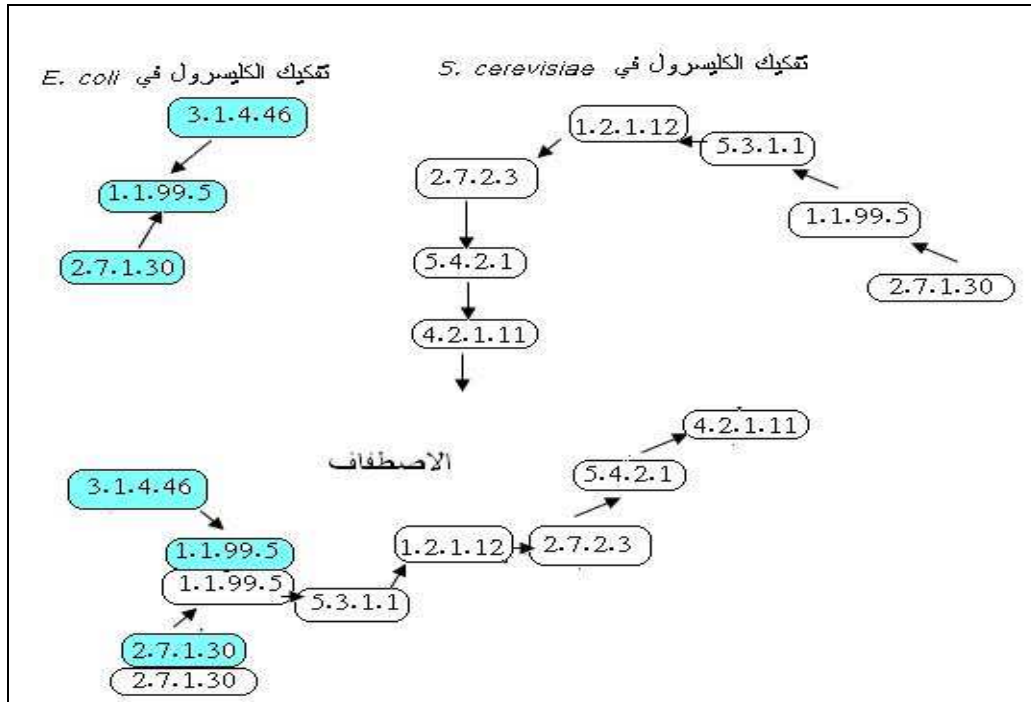
شكل 3.6 : التعامل الحاسوبي مع المسارات الايضية

ووقت تنفيذ للبرامج يعتمد كما هو معروف على حجم قواعد البيانات التي تستخدم للاصطفاف ، وترتب المسارات الايضية وفق التشابه والقيم الاحصائية وتعطي نتائج مصورة. وتتم عمليات صف المسارات الايضية بشكل مشابه في الأطر الخارجية لصف جزيئات الحياة ، فقد تكون عمليات الاصطفاف موضعية أو عمليات صف شاملة ويتم تسجيل الدرجات مع إعطاء قيم للحدود الفاصلة أو حد العتبة وتستخدم في ترتيب النتائج وفق القيم الاحصائية كما ذكر آنفا . ومن الطبيعي يكون التعامل مع الإنزيمات التي تقوم بوظيفة واحدة أسهل من التعامل مع إنزيمات متعددة الوظائف والتي تمثل عادة بأشجار أو تفرعات ثانوية ، ويكون الأساس

المعتمد ان الإنزيمات ذات EC المتشابهة تكون متشابهة في الوظيفة وليس في الضروري ان يكون هناك تشابهها" في التوالي .

والطرق المستعملة هي اما الموجهة أو تستعمل بعض الأحيان البرمجة الدائيميكية البطيئة ، ويتم فيها حساب الفجوات وما يتعلق بها من جزاءات ، وفي جميع الأحوال فان إتمام الاصطفافات والحصول على قياسات كمية لابد من ان تمثل رياضيا لغرض ان يقوم الحاسوب بالتعامل معها . وتجرى المقارنات ضمن النوع الواحد والسلالات المختلفة أو بين الأنواع المختلفة .

والأهداف الرئيسية من صف المسارات الايضية يكون لإيجاد المسارات الثابتة على مدى عمليات التطور ، في دراسة مقارنة بين *E. coli* وخميرة الخبز وجد ان نسبة محددة من المسارات كانت متشابهة وهذا يشير الى ان هذه المسارات قد وجدت قبل افتراق الأحياء الى ممالك الحياة الثلاث أثناء عملية التطور، كما موضح في الشكل 4.6 .



شكل 4.6 : مقارنة واصطفاف بعض المسارات الايضية بين *E. coli* وخميرة الخبز

ان إيجاد القطع الثابتة من المسارات الايضية يساعد في إمكانية إعادة توجيه المسارات الايضية وكذلك هندسة المسارات الايضية ضمن ما يسمى بتعديل المسارات الايضية Metabolic pathway tinkering لتلائم متطلبات جديدة . من أهم قواعد البيانات المهمة لمقارنة المسارات الايضية الهدمية هي قاعدة UM-BBD المارة الذكر التي تتعامل مع تفكيك المواد الملوثة في البيئة ، فهي ترتبط الى قواعد بيانات للبيض مثل قاعدة كيوتو وكذلك قاعدة بنك

الجينات و ChemFinder وترتبط بقواعد الإنزيمات . والمركبات المستعملة فيها تكون معرفة وفق الأرقام رسمية CAS No. كما انها تستعمل صيغ خاصة للمركبات هي SMILES . فضلا عن استخدامها لطرق خاصة بها لتصنيف المسارات الايضية .

والبعض من العمليات التي تتم في هذه القاعدة يعتمد على الحدس لذلك يتوقع ان يكون فيها أخطاء لذا كان لابد من إيجاد وسائل للتأكد منها . وعلى العموم فان القاعدة تحتاج الى توحيد الملفات المستعملة واستعمال التسميات القياسية وذلك لان معظم الأخطاء آتية من عدم تجانس المعلومات وصيغها في قاعدة البيانات والذي يؤدي بدوره الى عدم انسجام البيانات المستخلصة من مصادر مختلفة ترتبط اليها القاعدة . وسيأتي تفصيل أكثر للقاعدة عند مناقشة موضوع المعلوماتية الحيوية والبيئة لان القاعدة أكثر التصاقا بالآخرية .

وربما هناك نقطة يمكن الإشارة اليها في نهاية هذا الموضوع هو بدايات لعمليات اصطفاف الشبكات الحيوية وليس فقط المسارات الايضية وهذه الأخرى يمكن ان تكون موضعية أو شاملة والمظهر الرئيس فيها اعتماد تداخلات البروتينات Protein-Protein interaction (PPI) ومن البرامج الخاصة هي QNet لشبكات الاستعلام وبرامج QPath الذي يرافقه في الاستعلام عن مسارات الايض المستقيمة .

قواعد بيانات الايض الخاصة

بعد المناقشات أعلاه استمر التطور في بحوث الايض ونتجت عنه إنشاء مواقع وقواعد بيانات خاصة منها :

• MetaCrop وهي خاصة ببيانات المحاصيل المهمة وفيها تدرج المسارات الايضية والإنزيمات والمواد الكيماوية (بذكر صيغها الكيماوية) فضلا عن ذكر مواقعها في النبات سواء في عضو أو نسيج مع ذكر مرحلة التطور النباتي الذي توجد فيه المركبات ، وذكر الايض المركزي للنبات والبيانات الحركية للتفاعلات فضلا عن توفر المراجع الخاصة . وتحتوي القاعدة على قواعد خاصة بالمحاصيل مثل الرز RiceCyc و AraCyc والخاصة برشاد الصخر الذي يعد النموذج الدراسي للنباتات وتحتوي على مسارات الايض الثانوي والمعلومات غير المتوفرة حول نوع معين يمكن تحديدها بشكل تقريبي من الأنواع القريبة منه

• قواعد البيانات الخاصة بالإنسان : وفي هذا المجال فان الحقل غير معرف بشكل كامل ومن هذه القواعد قاعدة مكنون ابيض الإنسان Human metabolome database (HMDB) ، قاعدة متوفرة على صفحات الانترنت تحوي العديد من البيانات المفصلة حول مواد الايض الصغيرة الموجودة في جسم الإنسان ، وقد خصصت للاستعمال في مجال دراسة مكنون الايض والكيمياء السريرية والكشف عن الواسمات الحيوية . وصممت القاعدة لتكون ذات روابط الى :

- بيانات كيماوية .
- بيانات سريرية .
- بيانات في مجال البايولوجي الجزيئي والكيمياء الحيوية .

للقاعدة عدة إصدارات الحديث منها (الإصدار 2.5) يحتوي على أكثر من 7900 مدخل لمواد الأيض وتشمل مواد الأيض الذائبة في الماء وكذلك الذائبة في الدهون فضلا عن مداخل للمواد وفق وفرتها مثلا تلك المتوفرة بتركيز عالية أكثر من 1 مايكرومول والمواد النذرة التي تراكيزها أقل من 1 نانومول وترتبط القاعدة بما يقرب من 7200 توالي للبروتينات و DNA التي لها علاقة بمداخل مواد الأيض التي تحتويها . وكل مدخل MetaboCard يحوي على أكثر من 110 حقل من البيانات تلتين منها مخصصة للبيانات الكيماوية والسرييرية والثالث الآخر مخصص للبيانات الكيموحيوية والإنزيمات . وكل حقل من البيانات يرتبط بقواعد بيانات أخرى مثل قاعدة كيوتو و MetaCyc وقاعدة البروتينات السويسرية وبنك بيانات البروتينات PDB وغيرها . توفر القاعدة قاعدة عريضة للمراجع والتواليات والتراكيب الكيماوية . تحوي القاعدة على أربع قواعد إضافية متخصصة منها ما يخص بنك الأدوية وأخرى خاصة بالسُموم والجزيئات الصغيرة والأغذية (كما سيأتي ذكره لاحقا)

Human metabolome project (HMP) ويمثل مشروع الأيض البشري وتكون البيانات فيها متاحة من خلال Human metabolome library والموقع أو القاعدة تزود بالمعلومات حول المركبات الكيماوية للأغراض العلمية . ومثل هذه المواقع تكون مساعدة في دراسة الأيض كما في مشروع الجينوم البشري ، والمساعدة في فهم العديد من اضطرابات الأيض . والقاعدة توفر معلومات حول جزيئات الأيض الصغيرة في جسم الإنسان وتستخدم في الكيمياء السرييرية والكشف عن الواسمات الحيوية ، ترتبط الى بيانات سرييرية وأخرى كيماوية مع تبيان المعلومات عنها . وتحوي القاعدة بيانات مثل :

- الجزيئات الذائبة في الماء
- الجزيئات الذائبة في الدهون
- الجزيئات العالية التركيز Macro التي تركيزها أعلى من 1 مايكرومول
- الجزيئات النادرة Micro التي تركيزها أقل من نانومول
- وكل بطاقة MetaboCard يحوي معارف حول جزيئات الأيض في حقول كثيرة وله ارتباطات مع قواعد بيانات متعددة للأبيض والتراكيب وترتبط الى بنك الأدوية DrugBank وقواعده الثانوية مثل T3DB الخاصة بالسُموم والملوثات وكذلك FoodDB الخاصة بالأغذية ومكوناتها ومضافاتها و SMPDB الخاصة بالأمراض البشرية .
- وتحوي أيضا على معلومات حول السوائل الجسمية Biofluids التي تمثل بجدول فيها قوائم بتركيز مواد الأيض الطبيعية وغير الطبيعية في سوائل الجسم أي انه يشبه القاعدة HumanCyc database وهي إحدى قواعد البيانات التابعة لقاعدة البيانات الشاملة BioCyc وتحوي العديد من البيانات وتحوي عدد كبير من المداخل لغاية 2010 هي:

- المسارات الايضية ويصل عددها الى 255

- جينوم الإنسان

- القليل من مسارات نقل الإشارات

- التفاعلات الحيوية ويصل عددها الى 1862

- البروتينات الناقلة
 - مواد ايض بشرية يصل عددها الى 1359
 - مكونات الخلية الثانوية
 - إنزيمات ويصل عددها الى 3315 من الإنزيمات البشرية
- وترتبط بعدد من قواعد البيانات الخاصة سواء بالبروتينات ، أو الجينومات و كذلك المايكوكوندريا . وهي تمثل المرحلة بعد الجينوم Beyond the genome (BTG) لتكون البوابة للمسارات الايضية (PGDB)Pathway/Genome databases وتتميز القاعدة بكونها ذات نوعية جيدة اذ ان المعلومات فيها تكون مدققة ومشذبة وتستعمل في تحليل بيانات -omics التي تخص المسارات الايضية وتساعد في ذلك من خلال برامج Softwares التي تقوم بحدس الشبكات الايضية للكائنات الحية فضلا عن إمكانية تحديد الإنزيمات المفقودة أو غير المعروفة في المسارات الايضية . ومما يساعد في ذلك هو وجود إمكانيات استعمال الألوان والتكبير لمخططات الشبكات الايضية . واحتواء القاعدة على ارتباطات لمواد الايض مع التفاعلات والإنزيمات والجينات يسهل دمج أنواع من بيانات -omics .
- ترتبط القاعدة بالمصادر الخاصة بالجينوم البشري مثل Ensemble وفيها بطاقات الجين GeneCards خاصة ، وكذلك ترتبط الى Entrez gene و OMIM والمصادر الخاصة بالمواد الكيماوية PubChem ، فضلا عن ارتباطها بقواعد بيانات البروتينات مثل UniProt والقاعدة تتعرض لنشاطات كبيرة لتحسينها وإضافة المزيد من المداخل عليها مثل البروتينات الناقلة وبيانات حول أجزاء الخلية وغيرها من الإمكانيات.

المعلوماتية الحيوية والبيئة

بعد حقبة استعمال عمليات الزرع للأحياء المجهرية التي استمرت طويلا وفق الطرق المتاحة ذلك الوقت تبين ان الصورة التي يتم الحصول عليها غير مرضية تماما اذ ان ما يمكن زراعته في المختبر لا يتجاوز 1 % من الأحياء الموجودة في بيئة معينة . وبعد تطور طرق التحليل والكشف خاصة على المستوى الجزئي ظهر توجه استعمال الجينوم البيئي او جينوم المجتمع Metagenome الذي يمثل كل (تقريبا) جينومات الأحياء الموجودة في بيئة ما والذي يتم تحديده والكشف عنه بالطرق الجزيئية بعيدا عن عمليات الزرع التي أثبتت فشلها في هذه الناحية . وقد أطلق عليها الجينوم ألتجوازي (Meta-) أو الجينوم المجتمعي لانه يتجاوز الدراسات الجينومية العادية ليعبر الى DNA البيئي وبدا استعمال المصطلح عام 1998 للوصف والتحليل الوظيفي لفلورا التربة وسمي عندها Soil metagenomics . والتعامل مع هذا المجال يعرف ايضا بدراسة الجينومات البيئية Ecogenomics أو Environmental genomics .

ويؤثر الكشف عن الجينوم البيئي في تحديد مفهوم التنوع الحيوي Biodiversity الذي يمثل الضروب والمتغيرات بين الأحياء والمعقدات البيئية التي توجد فيها أي درجة من التغيرات في أشكال الحياة في نظام بيئي أو نظام حيوي Biome على كوكب الأرض. ولا بد من وقفة عند هذا الموضوع اذ لم يتم تناوله بشكل كاف فيما سبق .

فمصطلح التنوع الحيوي يمثل أنظمة بيئية مختلفة والأنواع الحية التي توجد فيها فضلا عن الجينات المختلفة في بيئة ما والتي تحدد وفرتها النسبية عند إجراء الدراسات . فالتنوع الطبيعي يكون مرادفا تقريبا للتنوع الحيوي Biological diversity ، اذ يمثل عدد الأنواع المختلفة وأعدادها وأفرادها في بيئة ما أو منطقة جغرافية . ومن المتوقع ان تحدث تداخلات مختلفة بين الأنواع المختلفة في البيئة الواحدة مما يؤدي الى تغير وراثي بين الأفراد أو ضمن النوع الواحد ، وهذا يؤدي الى التنوع الوراثي الذي يعني التنوع الوراثي ضمن النوع الواحد أو التغيرات الوراثي ضمن مجتمع ما . وعلى أساسه يحصل تنوع الأنواع أو تنوع تصنيفي Species diversity أو Taxonomic diversity في منطقة محددة ويوضح علاقة الأنواع مع بعضها مثلا بيئة فيها 3 أنواع من الطيور تكون اقل تنوعا من البيئة التي فيها 2 نوع من الطيور ونوع واحد من السحالي.

وبعد التنوع الحيوي احد المقاييس المهمة لصحة الأنظمة الحيوية وقد خصصت سنة 2010 لتكون سنة التنوع الحيوي من قبل المنظمات المعنية بعد ملاحظة التدهور الكبير الذي حصل في التنوع الحيوي للعديد من البيئات في بقاع العالم المختلفة وقلت المناطق الساخنة للتنوع الحيوي Biodiversity hotspots التي تمثل مناطق ذات مستويات عالية من التنوع كما في المناطق الاستوائية وخاصة الغابات .

ومن النواحي الأخرى التي يشارك في دراستها الجينوم البيئي او المجتمعي هو تكوين الأغشية الحيوية من قبل الأحياء المجهرية سواء كانت في النواحي السلبية عند حث الأمراض أو النواحي الايجابية عند معالجة تلوث البيئة . وعيش البكتريا في الأغشية الحيوية يشير الى الحياة الاجتماعية التي تعيشها في الطبيعة وان العيش والنمو في الأوساط الزراعية هو الاستثناء

، وتستعمل في مثل هذه الحالات اللغات الخاصة بها للتواصل عبر ظاهرة تحسس الزحام Quorum sensing وترتيب تصرف المجتمع البكتيري وفق الزحام الحاصل للأفراد في حيز محدد . ولدراسة مثل هذه الظواهر لابد من إجراء تحليل DNA لكل الميكروبات القابلة وغير القابلة للزرع أي دراسة جينوم المجتمع ، والطريقة الفعالة في ذلك هو بناء مكتبات خاصة Metagenomic libraries باستعمال نواقل مختلفة مثل بلازميدات أو الكوزميدات والفوزميدات Fosmids وكروموسومات البكتيريا الصناعية BAC المعروف تداولها في طرق الهندسة الوراثية . ثم إجراء البحث في هذه المكتبات لإيجاد الجزيئات الجديدة والفعالة والتي تكون لها تطبيقات في التقنية الحيوية والصيدلانية . وهناك طريقتان لإجراء المسح منها

- الطرق المعتمدة على التواليات مثل استعمال PCR أو التهجين وتحديد التواليات.

- الطرق المعتمدة على الوظائف مثل تشخيص التغيرات في الفعاليات الإنزيمية وتكوين الهالات الشفافة حول المستعمرات في بعض طرق التقدير المختبري وغيرها من الطرق.

ففي الأغشية الحيوية يعيش أكثر من نوع من الكائنات وحيدة الخلية تكون للمجتمع من الكائنات الحية وعليه فإن مفهوم الجينوم البيئي يتجاوز الجينات والجينومات الخاصة بالأنواع ويذهب إلى أبعد من دراسات الجينومات العادية ، وفي هذا المجال وفرت المعلوماتية الحيوية البرامج الملائمة للكشف عن متأصلات تحسس الزحام في الأحياء القابلة للزرع وكذلك وفرت برامج للأحياء غير القابلة للزرع بالاعتماد على جينوم المجتمع البيئي من أغشية حيوية مختلفة . ومن البرامج المستعملة Orphelia الذي يستعمل لحدس الجينات من قطع DNA القصيرة التي تعود إلى جينومات الأحياء التي لا يمكن أن تنمو ومن أهم معوقات دراسة الجينومات البيئية تكمن في قصر المقروءات لذلك يتم البحث عن مقروءات قصيرة في قواعد البيانات مباشرة للجينات المعروفة أو البروتينات المعروفة ومقارنتها للكشف عن التواليات ذات التماثل الجذري .

ومن البرامج المساعدة في هذا المجال Metagenomic ORFome assembly (metaORFA) فعند إيجاد أطر القراءة تجمع في المجمع ORFome وتستعمل للبحث في قواعد البيانات عن المتشابهات الجذرية وبالتالي تمكن من دراسة وتحليل التنوع . إن عملية التجميع تزيد من حساسية عملية البحث عن التشابه الجذري وبالتالي تحسن من تحليل التنوع في بيانات جينوم المجتمع وهذا يكون مفيداً وملائماً عندما يكون تجميع الجينوم غير ممكن والذي لا يؤدي الغرض نظراً لقلّة تغطية التواليات بالطرق العادية .

الطرق المستعملة في تحديد الجينوم البيئي

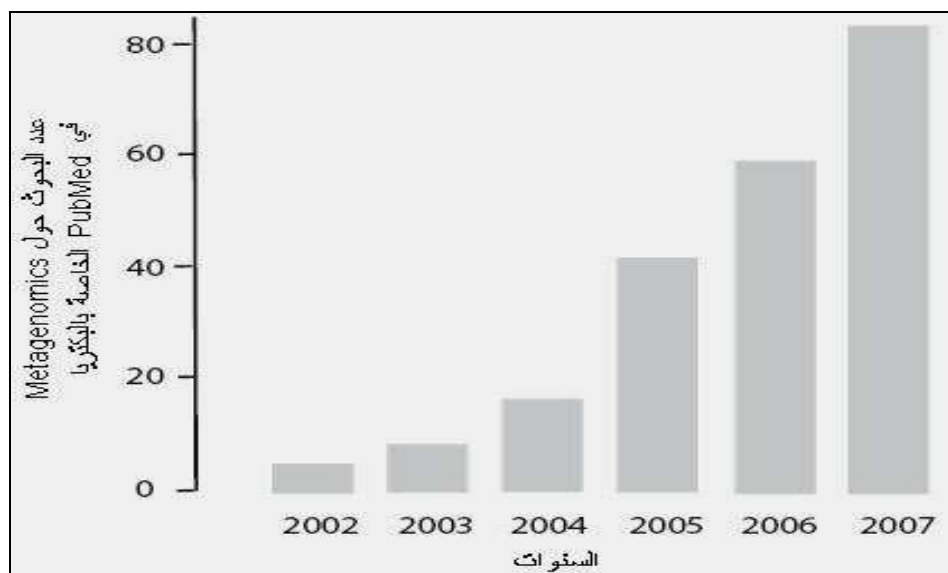
الطرق المستعملة عادة تمكن من الوصول إلى الأحياء التي لا يمكن أن تزرع بالطرق العادية . وتعتمد بعض الطرق على تحليل 16S rRNA . وبالرغم من أن الاعتماد على 16S rRNA أعطى صورة واضحة عن عالم الميكروبات وساعد في اكتشاف خطوط Lineages جديدة في مجتمع الميكروبات ، وأعطى وصفاً واضحاً للعلاقات التطورية لأعضاء المجتمعات إلا أنها أضفت القليل من الضوء على النواحي الوراثية والفلسجية والكيميائية الحيوية للأحياء . وهناك طرق أخرى منها طرق داخلية والتي تعتمد على تحليل تواليات الجينومات للتمييز بين التواليات المشفرة ORFs والتواليات غير المشفرة NORFs .

والطرق الخارجية وهي التي تبحث عن المناطق الثابتة خلال عملية التطور والتي يكون أساسها انه أثناء عمليات التطور فان الجينات الجديدة تنشأ من تضاعف وإعادة ترتيب الجينات والطفرات الحاصلة في الجينات الموجودة . والطرق الموجودة على الساحة في الوقت الحاضر تواجه بعض المعوقات منها الحاجة الى أتمتة الطرق لمواجهة الكم الكبير من البيانات ، والمشكلة الأخرى هو قصر التواليات في النماذج البيئية والتي تحجز عن إعطاء الصفات المميزة للتواليات وكذلك حصول إزاحة في الشفرات فضلا عن وجود شفرات وقف In-stop codons في التواليات . وفي هذا المجال طورت بعض البرامج لاستغلال المعلومات للتعامل مع القطع الصغيرة وان كانت غير كفوءة . وفي الوقت الحاضر تتم العملية بالبحث عن المتشابهات باستعمال برامج BLAST ويلاحظ ان الأنواع المتغلبة في النماذج البيئية الطبيعية تكون ممثلة بشكل اكبر في النماذج وهذه المجموعة لا تشكل عقبة ، وتكمن المشكلة في الأحياء قليلة التواجد اذ يمكن الحصول على متجاورات قصيرة فقط وبالتالي تكون التغطية قليلة فضلا عن ان هذه القطع يمكن ان تحوي على حالة إزاحة الشفرات ووجود شفرات الوقف داخلها لذلك يلاحظ ان البرامج العامة المستعملة لا يمكن ان تواكب هذه المشاكل .

ومن جهة ثانية يمكن اعتماد تواليات البروتينات وهنا يجب اعتماد عمليات الاستبدالات المتماثلة في الجينات المشفرة أكثر من المناطق غير المشفرة ، فزيادة المعدلات KS/KA أي غير المترادفة (KS) الى المترادفة يشير الى عمليات انتخاب موجب أو تطور معتدل وهذه يمكن ان تستغل في وضع بعض البرامج وان تستعمل معها قواعد بيانات ذات نوعية جيدة مثل Refseq وإبعاد البروتينات الإضافية المكررة .

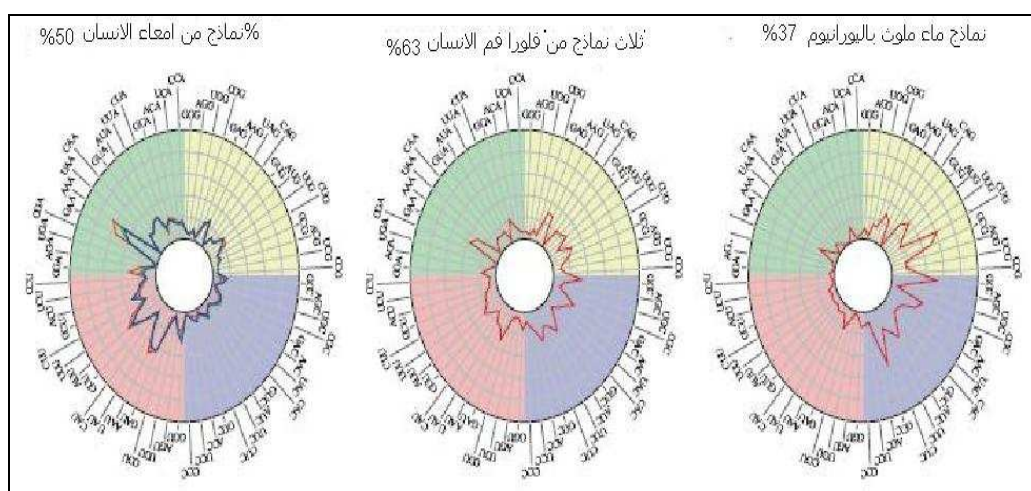
وعليه فان دراسة الجينومات البيئية (إضافة الى دراسة 16S rRNA) زادت من الإمكانات التي تسهل دراسة فلسجة وبيئة الأحياء المجهرية ، وان كانت تواجه بعض العقبات المذكورة آنفا فضلا عن عدم إمكانية إنتاج جينوم كامل من تحليل الجينوم البيئي وذلك لوجود الفجوات ووجود مكررات DNA التي تعيق تجميع الجينوم الكامل ، وأمكن مع هذا الكشف عن جينات جديدة ومنتجاتها. وعند استعمال الدراسات الجينومية البيئية لمجموعة معينة من الأحياء يمكن استعمال خيار التصفية (Filter) الذي توفره بعض البرامج وبذلك يمكن التعرف على مثلا النظام البيئي للبكتريا دون الحاجة الى زراعتها.

وتشير دراسة الجينوم البيئي او المجتمعي الى تفوق مجموعة Proteobacteria على Firmicutes و Actinobacteria والتي تصل الى 80% وعلى العموم فان هذه لا تشكل الا 10% من مجاميع البكتريا ، وقد يكون سبب ذلك ان قواعد البيانات تكون في الأغلب منحازة الى الأنواع الممثلة بشكل كبير ولذلك كانت عملية البدء بالعمل بـ 16S rRNA هي الخطوة الصحيحة للبدء على شرط الركون الى التهميشات الصحيحة لجزيئات rRNA ، وبغض النظر عن المثالب تعد عملية وضع الأشجار التطورية والعلاقات التصنيفية باستعمال 16S rRNA هي المفضلة وقد ازداد استعمال تحديد تواليات الجينومات البيئية بشكل كبير كما موضح في الشكل 5.6 :



شكل 5.6 التطور الزمني لدراسات جينوم المجتمعات Metagenomics

تأثير البيئة في الجينوم المجتمعي وتعامل المعلوماتية الحيوية معه
تؤثر البيئة في المحتوى الجينومي البيئي كما اثبت من النماذج البيئية المدروسة ، فالبيئة يمكن ان تؤثر في المكونات من القواعد النتروجينية . فمن المعلومات المتوفرة يلاحظ ان البيئة تؤثر في البكتريا الموجودة فيها وتكون لها لهجة وراثية Genetic dialect مشتركة ومن ذلك احتواءها على كميات من AT متشابه ، كما انها تؤثر في الشفرات المستعملة والذي لا يمكن تفسيره بالصدفة كما موضح في الشكل التالي (6.6) الذي يمثل ميل البكتريا المختلفة من تفضيل نوع من الشفرات على غيرها في بيئات مختلفة .



شكل 6.6 : تأثير الظروف البيئية في تفضيل استعمال الشفرات

فعند دخول قطع من DNA بالنقل الأفقي للأحياء فان مصير بقاء القطع في الخلايا المستلمة وفي الأفراد الناتجة منها يكون منوطا بمدى تأثيرها الايجابي والذي يشجع عادة فيما اذا كان محتوى AT واستعمال الشفرات في DNA القادم مشابها لذلك الموجود في الخلية المستلمة التي دخل اليها وإلا سيكون هناك كبح وإرباك لعملية التعبير عنه وترجمته . ونتائج الجينوم البيئي لا يمكن تمثيلها بأطلس نظرا لان تواليات الجينوم تكون غير كاملة ، وبرامج رسم الأطالس BLAST atlas تحتاج الى أطلس مرجعي ، وفي حالة رسم اطالس جينوم المجتمع Metagenomic atlas فهو يكون فقط للمقارنة بين جينومات لأحياء توجد في بيئة محددة . اما في حالة الحصول على جينوم بيئي متكامل فانه سيتم الحصول على كم هائل من المعلومات وفي مثل هذه الحالات يمكن اللجوء الى جينوم كامل كبير لأحد الأحياء التي تقطن البيئة واعتباره جينوم مرجعي ، فمثلا عند التعامل مع نماذج بيئة التربة فمن المتوقع ان تكون جينومات *Clostridium spp* هي الجينومات المرجعية كما في استعمال *Cl. botulinum AS* . وهناك قواعد بيانات خاصة بالجينوم البيئي او جينوم المجتمع .

الفوائد والجوانب التي يستعمل فيها الجينوم البيئي

من أهم تطبيقات دراسة الجينوم البيئي هو معرفة الخفايا التي تربط بين الأحياء والبيئة التي تشغلها كما في دراسة مكنون الإنسان الميكروبي Human microbiome الذي سيتم تناوله فيما بعد . ومن أهم التطبيقات ايضا هو معالجة البيئة التي تدمر جزءا كبيرا منها بالملوثات وأدى الى تغيرات كبيرة على المستوى العالمي كما في تغير المناخ . وتعتمد معالجة البيئة استعمال الدراسات الجينومية للبيئة لإلقاء الضوء على العمليات الأساسية التي بواسطتها تتأقلم الكائنات الحية وتتطور تجاه الجوانب الحيوية وغير الحيوية في بيئتها . وفهم كيفية التقبل والتفاعل مع التغيرات البيئية لها تطبيقات مباشرة في مجال المعالجة الحيوية Bioremediation . وقد رصدت الدول الطائل من الأموال لهذا المجال ولذا أنشأت في بعض الدول مراكز للمعلوماتية الحيوية الخاصة بالبيئة لتقديم الاستشارات وتدريب الكوادر وإدارة البيانات ، وهذه المراكز تعمل على إنشاء برامج تلاؤم أنواع أنظمة الحاسوب الموجودة وإجراء التجارب باستعمال تجارب المصفوفات Microarray experiments خاصة بالنماذج البيئية وقد صدرت كتب خاصة Catalogues تتضمن العديد من المعلومات حول الكائنات التي لها علاقة بالبيئة ضمن أرقام تسجيل Acc No. خاصة يمكن البحث عنها في قواعد البيانات العامة ، وتشمل omics data - للجينومات والبروتينات ونواحي أخرى فضلا عن النشرات والمراجع الموثقة .

والمعالجة الحيوية تعني الاستغلال الأمثل للأحياء خاصة المجهرية لإزالة الملوثات من البيئة وهذا يحتاج الى دمج كميات هائلة من البيانات من مراجع ومصادر مختلفة . ومن الأنظمة المعنية بهذا MetaRouter نظام لاحتواء وإدابة المعلومات المتباينة ذات العلاقة بالمعالجة الحيوية في نظام هيكلي يساعد في الاستعلام وتقديم أو إدخال البيانات وكذلك كشف البيانات الموجودة في النظام ، ويرتبط النظام داخليا وكذلك يرتبط خارجيا مع عدد من قواعد البيانات ، مع وسائل تمكن من استرداد البيانات ، ويتم التحديث فيه باستمرار. نظام MetaRouter

وإصداراته المختلفة مكتوب بلغة SQL التي تتناسب مهام البرنامج ويضم برامج ثانوية مثل PathFinder وغيرها وإمكانياته العامة مشابهة للقاعدة UM – BBD ويرتبط بها مباشرة. ولكن الملاحظ ان مصادر المعلوماتية الحيوية الخاصة لا تزال قليلة على الأقل لحد عام 2005 خاصة فيما اذا كانت العمليات تحتاج الى العوامل الخارجية وبالتالي فالعملية تدخل ضمن المنظومات الحيوية Systems biology ومؤخرا ازدادت المعلومات في هذا المجال بشكل كبير جدا . ومن الجهات المعنية بهذه النواحي هي

UM -BBD University of Minnesota Biocatalysis / Biodegradation database

والقاعدة تحوي في إصدارها نهاية عام 2011 على حوالي 200 مسار ايزي و 1389 تفاعل و 1283 من المركبات الكيماوية و 510 من الميكروبات والفعاليات التي تقوم بها فضلا عن احتوائها على عدد كبير من الإنزيمات تصل الى أكثر من 400 خاصة تلك المعنية بعملية الأكسجة Oxygenases . ومما يسهل استعمال قاعدة البيانات هذه ان فيها إمكانية للبحث Search machine ، فضلا عن وجود مواقع مناظرة لها Mirror في مواقع أخرى مثل EMBL وغيرها لتسهيل الوصول اليها ان لم يكن بالإمكان الوصول اليها مباشرة . والحقيقة ان القاعدة جيدة جدا فهي بالنسبة للمركبات الكيماوية توفر اسم المركب والمرادفات وكذلك SMILE code الذي يلزم في بعض معالجات الحاسوب باستعمال برامج خاصة فضلا عن توفير الصيغة الكيماوية والوزن الجزيئي وصورة للتركيب الكيماوي سواء العادي أو ثلاثي الأبعاد . وتوفر ايضا معلومات حول الصفات الفيزيائية – الكيماوية للمركبات مثل الكثافة ومعدل التبخر ونقطة الانصهار ونقطة الغليان ، والقابلية للذوبان في الماء وربط المركب بقواعد بيانات أخرى .

اما في حالة التفاعلات فتوفر القاعدة مواد أساس للتفاعل والنواتج والإنزيمات المساعدة فضلا عن ربط التفاعل بقواعد بيانات أخرى ، وفيما يخص الإنزيمات فهي تشير الى اسم الإنزيم و EC code ، والكائن الحوي على الجين الخاص بالإنزيم وربطه بقواعد البيانات الأخرى ، واهم المراجع التي تشتق منها المعلومات هي قاعدة ENZYME وقاعدة البروتينات السويسرية

المعلوماتية الحيوية والجوانب الصحية للإنسان

نظرا لأهمية الجوانب الصحية للإنسان فقد تداخلت جوانب عدة منها مع المعلوماتية الحيوية ، ولتعقد هذا التداخل سيتم التطرق الى عمليات التداخل التي قد تكون مباشرة كما في التعامل مع الأمراض ، وكذلك الأدوية وعمليات الايض والجوانب المناعية وكذلك التداخلات غير المباشرة كما في التأثيرات الضارة التي تبديها الأغذية المهندسة وراثيا وسيكون التعامل مع هذه الجوانب شبه مستقل وذلك لان الاستقلالية التامة غير ممكنة لان بعض الجوانب تشترك مع بعضها وفي بعض الأحيان يكون بدرجة كبيرة . والتجارب البايولوجية أنتجت وتنتج كم هائل من المعلومات والبيانات التي تحتاج الى إدارة وتحليل ثم الفهم . وفي مجالات الصحة والطب تساهم المعلوماتية الحيوية في العديد من الجوانب منها :

- كشف الأدوية الجديدة على وجه الخصوص وتحديد طرق تأثيرها .
- التشخيص .

- التعامل مع الأمراض .
 - إمكانية شخصنة الأدوية .
 - تشخيص الأنواع الثانوية للأمراض مثل السرطان وتحديد مدى بقاء المرض
 - تشخيص SNPs المؤدية للأمراض
- ويكون ذلك باستعمال المصفوفات والوسائل التي يمكن ان تقلل من كلفة تطوير الأدوية وذلك باستخلاص البيانات المفيدة من التجارب الصيدلانية التي تستعمل في الكشف عن الأدوية الجديدة ، والمحاولات جادة في نقل المعلوماتية الحيوية المعتمدة للأدوية الى المجال الصناعي .

المعلوماتية الحيوية والأمراض

التغطية في هذا الموضوع ستكون غير معمقة نظرا لوسع الموضوع . وتتمثل بالنقاط الآتية :

- التعامل مع الأمراض الوراثية ، فمن المعروف ان حصول الطفرات الوراثية يمكن ان يؤدي الى توليد الأمراض فهناك أمراض وراثية تنتج من تأثير عدد من الجينات وتؤدي الى أعراض متشابهة وهذا يشير الى ان هذه الجينات ذات علاقة وظيفية ومثل هذه العلاقات الوظيفية يمكن ان تستغل للمساعدة في إيجاد جينات المرض . ولكن بعض الأحيان تكون الآليات غير معروفة خاصة على المستوى الجزيئي . ولكن باستعمال المعلوماتية الحيوية ووسائلها أمكن تحديد السبب الجزيئي لعدد من الأمراض الوراثية وأمكن إنشاء مواقع ووسائل على الانترنت متاحة للباحثين الذين يدرسون الأمراض . كما ان تداخل البروتينات مع بعضها (بروتين – بروتين) تعد من أقوى المؤشرات على العلاقات الوظيفية بين الجينات ، ولذلك فان البروتينات المتداخلة تؤدي الى النمط المظهري نفسه عند حصول الطفرات . ولذا فان تداخلات البروتينات يمكن ان تعد البوابة للوصول الى الجينات المسؤولة أو المرشحة للمسؤولية عن بعض الأمراض الوراثية . وإضافة الى ذلك فان عمليات إحلال الحوامض الامينية مكان بعضها البعض يمكن ان يؤدي الى تغيير البروتينات وبالتالي إحداث الأمراض . والأخيرة يمكن حдسها باستعمال برامج خاصة وحسابات رياضية تمكن من حдس الفعاليات التي تتجم عن عمليات الإحلال . والمسجل في المراجع هو الميل لدراسة البروتينات المعروفة جيدا واستغلت من قبل الباحثين لحдس الجينات . ففي البداية تدرس البروتينات الموجودة في قاعدة البيانات ثم تطبق البرامج على البروتينات التي فيها طفرات وملاحظة ما يحدث نتيجة للطفرات وبما ان برامج الحاسوب غير متكاملة لذا كان الواجب تدخل البايولوجيين لتفسير النتائج ، وهذا يحتاج الى تدريب كوادر تكون ذات نظرة بايولوجية .

ان استعمال المعلوماتية الحيوية تكون ضرورة لمستقبل علاج السرطان ذات الأسباب المتعددة التي تكون في مقدمتها الجوانب الوراثية وتداخل ظروف البيئة مع المواد الوراثية . ولغرض العلاج لابد من تحديد نسق التعبير الجيني للأورام باستعمال مصفوفات DNA الدقيقة وغيرها من المصفوفات مثل مصفوفات البروتينات أو النسخ وغيرها التي تكون وسيلة فعالة في بعض الأحيان لغرض الاستهداف الصيدلاني للمعالجة وذلك لانها وسائل بعيدة عن الانحياز وهي من الطرق المهمة في الكشف عن جينات المرض الجديدة التي لم تحدد مسبقا . والأورام تحوي خليط من النسائل Subclones التي تحوي على جينات صامتة في حين الأخرى تحوي على جينات مطفرة ويعبر عنها بشكل كبير ، وهذه التباينية تصعب من الوصول الى تحديد الأهداف الجزيئية

، فالعديد من الطفرات أو الجينات ذات التعبير العالي قد لا تكون أهدافا جيدة نظرا لوجود النسائل المقاومة . ومن جهة ثانية يمكن الكشف عن المسارات الايضية لخلايا الورم التي تؤدي الى نموه وفي هذا المجال تساعد المعلوماتية الحيوية بشكل فاعل . فضلا عن إمكانية متابعة العلاج وذلك من إعادة تحديد جينوم خلايا الورم لمعرفة مدى تطور المرض بوجود العلاج .

وهناك الأمراض أخرى ذات الأسباب المتعددة والتي تشارك فيها عدد من الجينات والتي لا يمكن تحديد مواقعها أو حدسها مما يشير الى اشتراك أكثر من موقع وحدث التداخلات فيما بين الجزيئات أو مع البيئة أو الاثنين معا ومنها ارتفاع ضغط الدم والسكري وأمراض القلب الوعائية وغيرها . ولكثرة الأمراض كان لابد من تدخل المعلوماتية الحيوية لاحتوائها وان كان بشكل غير متكامل ، وبطبيعة الحال فان دراسة العلاقات الموضحة أعلاه وغيرها أدى الى تطوير المعلوماتية الحيوية السريرية Clinical bioinformatics التي تم بواسطتها ربط البيانات السريرية مع الدراسات الجينومية والتي ستساعد في تطوير وسائل التشخيص والتداوي والعناية بالمرضى . ومثل هذا التدخل يحتاج الى إيجاد قواعد بيانات . وقواعد البيانات الموجودة اليوم على الانترنت تختلف في توجهاتها وجودتها وفيما يلي بعض منها على سبيل المثال لا الحصر :

• (OMIM) Online Mendelian Inheritance in Man

توفر أرضية وافية للمعلومات حول جينات الإنسان ، وقد كانت بدايتها القاعدة MIM التي أنشأت في الستينيات من القرن المنصرم ونشرت حولها العديد من الكتب ، اما القاعدة الجديدة أو الواجهة الجديدة لها OMIM فقد أنشأت عام 1985 بتعاون من مؤسسات طبية عدة ووضعت عام 1995 على الانترنت ضمن الموقع الأمريكي NCBI ، لتكون مرجعا للأطباء والباحثين في مجال الاضطرابات الوراثية . تركز القاعدة على العلاقة بين النمط الجيني والنمط المظهري للأشخاص ، وكل مدخل فيه عدة روابط أخرى ، في آخر تحديث للقاعدة (عام 2011) تحوي القاعدة على 12000 جين خاصة بالاضطرابات الوراثية المندلية ، ويتم تحديث القاعدة على الصعيد اليومي .

• GeneCards

قاعدة يمكن البحث فيها عن الجينات البشرية التي تحوي معها بيانات حول المكونون الجينومي والبروتيني وكذلك مكنون النسخ للحالات السليمة والمرضية . والخاصية المميزة للقاعدة ان تركيبها يتجاوز التباين في صيغ البيانات وهذا يسهل أمور البحث . تحوي على ملخصات تصف الجينات لتساعد في فهم النواحي البيولوجية والاختلافات عند حصول المرض .

• قاعدة بيانات كيو تو للأمراض KEGG Disease

من المعروف ان قاعدة كيو تو من القواعد المهمة التي تهتم بعمليات الايض، وقد انبرت القاعدة لتضم قاعدة ثانوية خاصة بالأمراض وتحوي على قوائم للأمراض ذات الأسباب المتعددة مثل :

السرطان

أمراض الجهاز المناعي

الأمراض العصبية التحليلية

أمراض القلب الوعائية

أمراض الايض

فضلا عن احتوائها على بعض أمراض الإصابة والتي توضح فيها الشبكة الجزيئية للتداخلات لكل من الممرض والإنسان ، وتشير مخرجات البحث فيها الى الجينات المسؤولة في حالة معرفتها بشكل مميز .

• قواعد بيانات تخص miRNA

كما هو معروف (وسبق ذكره) ان جزيئات micro RNA غير المشفرة التي تتراوح أطوالها حوالي 22 نيوكلوไทيد تلعب دورا أساسيا في التنظيم خاصة السلبي وذلك بالارتباط والتكامل مع قطع من المناطق غير القابلة للترجمة 3'UTRs لجزيئات mRNA مؤديا الى انفلاق الأخيرة وتنشيط عملية الترجمة وهذه الجزيئات تلعب دورا مهما في العديد من الفعاليات الحيوية للخلايا مثل نمو وتكاثر الخلايا ، والتمايز للأنسجة والتطورات الجنينية والاستماتة . وعند حصول طفرات في هذه الجزيئات فانه يؤدي الى اضطراب وظائفها وبالتالي تؤدي الى عدة أمراض مثل

- السرطان
- أمراض القلب الوعائية
- انفصام الشخصية
- داء الصدفية الجلدي
- بعض الأمراض العقلية التي تؤثر في القدرات العقلية
- داء السكري
- التهاب الكبد المزمن
- السمنة

وغيرها من الأمراض

وأشارت الدراسات الى ان تداخل الجزيئات في هذه الأمراض يكون معقدا ولذلك كان توفير البيانات والتحليلات الوافية عن هذه الجزيئات يمكن ان يساهم في معرفة الآليات التي تشارك بها هذه الجزيئات في حالة المرض. وقد انشأت أكثر من قاعدة بالاعتماد على المراجع العلمية التي درست وتدرس هذه الحالات ، ومن هذه القواعد :

• Human microRNA disease Database

توفر القاعدة أسماء جزيئات miRNA ، وأسماء الأمراض ، والشواهد على اضطراب الوظائف فضلا عن إعطاء الرقم التعريفي في قاعدة المراجع PubMed ID للرجوع الى المراجع الأصلية ، كما ان كل جزيئة مرتبطة الى تهميش وافي عن جيناتها ، وتوفر صور لعمليات التعبير عن miRNA في العديد من الأنسجة . والقاعدة تحدث كل مدة وآخر تحديث كان في آذار 2011 ووفق تعليمات كادر القاعدة فإنها يمكن ان تستعمل أفضل مع متصفح الانترنت Firefox .

• قاعدة miR2Disease ، قاعدة تهتم بـ miRNA ذات نوعية جيدة يتم تشذيب المعلومات فيها بشكل عملي وتحديث بمدى نصف شهري . تهدف لإعطاء مصادر ومعلومات وافية حول عمليات اضطراب التنظيم التي تحدث بواسطة miRNA المشاركة في أمراض الإنسان المختلفة . وكل مدخل للأمراض يحوي على miRNA ID واسم المرض ووصف بسيط للمرض ونمط التعبير عن miRNA في حالة المرض وطريقة تحديد التعبير عنها . وتذكر ايضا الجينات

المستهدفة بتأثير miRNA ، والمراجع الخاصة . يكون البحث فيها اما باستعمال miRNA ID أو اسم المرض أو الجين المستهدف وتؤكد القاعدة عند استلام البيانات للإيداع فيها على تحديد العلاقة بين المرض و miRNA والشواهد عليها .

• ZF – HDD هذه القاعدة تعتمد على التشابه ووجود الجينات المتناظرة Orthologous genes بين السمكة المخططة Zebrafish (*Danio rerio*) واستعمالها كموديل حيواني لدراسة أمراض الإنسان ، فهناك تشابه في البروتينات بين الإنسان وهذا النوع من الأسماك يصل الى 65 % ، فضلا عن تشابه الجينات المتناظرة التي لها علاقة بالأمراض . لذلك كانت قاعدة مهمة في دراسة أمراض الإنسان .

وهناك العديد من قواعد البيانات المتعلقة بالأمراض وعدد منها تعطي معلومات وصفية لأمراض الإنسان لغرض القراءة والتتقيف ومنها :

• قاعدة بيانات الأمراض Disease database: وهذه القاعدة ليس للتشخيص أو تحديد الحالات السريرية وإنما هي للتتقيف وإعطاء الخلفية للقراء ، وتوفر فهارس بأمراض الإنسان التي تشمل الاضطرابات الصحية والأعراض ، والأمراض الولادية المستورثة وأمراض الإصابة والأحياء المسببة لها والأدوية والعلاجات وتوفر تعريف قاموسي للمفردات الموجودة فيها وفق (NLM) National Library of Medicine ، ولها روابط مع مواقع أخرى للمساعدة ، وتوفر البيانات بأكثر من لغة غير الانكليزية .

وتوجد قواعد البيانات للأمراض متخصصة بمجموعة من الأمراض فقد تم إنشاء قواعد بيانات خاصة بالأخطاء الولادية في مجال الابيض Inborn errors of metabolism وتبنى هذه من المعلومات والبيانات المتوفرة ويمكن ان توزع بياناتها وفق المناطق الجغرافية . وفي العادة وكما هو الحال مع العديد من قواعد البيانات تزود بوسائل وآليات تزيد من إمكانية الاستفادة من المعلومات المدرجة فيها .

المعلوماتية الحيوية والأدوية

ان تطور التقنيات الجينومية غطت معظم الجوانب الخاصة بالجينومات ، وبالمقابل تطورت إمكانيات تصميم الأدوية تجاه أهداف جزيئية وهذا ما وفر العديد من الفرص لاستعمال المعلوماتية الحيوية لجسر الهوة بين المعلومات البايولوجية والعلاج السريري .

ولا تقتصر مهمة المعلوماتية الحيوية في علاقتها مع الأدوية وإنشاء قواعد وبنوك للأدوية وإنما يجب ان يكون الإنشاء معتمدا على أسس صحيحة ، فبالإضافة الى التأكد من نوعية الأهداف للأدوية لابد من الأخذ بنظر الاعتبار تداخل الأدوية مع بعضها Drug-drug interaction (DDI) وهي عملية ضرورية لتطوير الأدوية وإمكانية استعمالها والتي تحصل عند استعمال أكثر من دواء في الآن ذاته وهناك عدة طرق لتوضيح العلاقات منها إيجاد مسارات الإشارات أو مسارات تخليق جزيئات الأدوية . وهذه الطرق تعتمد على المراجع ومنها يتم معرفة الإنزيمات وعوامل الانتساخ المشاركة مع الأخذ بنظر الاعتبار ان ايض الأدوية يحصل في الكبد بشكل رئيس ، وان نسبة معينة يتخلص الجسم منها ولكن اذا كانت عملية التخلص بطيئة فانها تؤدي الى السمية وعندها يتضخم تأثير الأدوية المضر . وعمليات حث وتنشيط الإنزيمات التي قد تكون مباشرة أو غير مباشرة لعمليات تنظيم الانتساخ وعملية تثبيط الإنزيمات تعد أهم أوجه تداخل

الأدوية (DDI) مما يؤدي الى فشل حالة التداوي ، في حين ان تنشيط الإنزيمات الذي يمثل الوجه الآخر لتداخل الأدوية يمكن ان يؤدي الى اضطراب التوازن وبالتالي عدم الاستفادة من الأدوية . وكما ذكر آنفا فان عوامل الانتساخ ايضا تعمل في ايض الأدوية وبتأثير غير مباشر . فضلا عن ذلك فان التأثير الآخر يكون من خلال الناقلات المسؤولة عن إدخال الأدوية الى الخلايا . كل هذه الجوانب يجب ان تؤخذ بنظر الاعتبار عند إنشاء القواعد وارتباطاتها والوسائل التي تقدمها .

ومن نتائج الجهود انشأت قاعدة Human Drug Metabolism Database لتكون بمثابة مصادر للمعلوماتية الحيوية والمعلوماتية الكيماوية وهي خاصة بايض الأدوية وتداخلاتها بشكل عام.

بنك الأدوية DrugBank

قاعدة بيانات للمعلوماتية الحيوية والمعلوماتية الكيماوية Cheminformatics ويعد مصدرا شاملا للكشف عن الأدوية عن طريق الحاسوب . وقد انشأ البنك كضرورة ملحة لان معظم المعلومات عن الأدوية لازالت في الكتب والمجلات المطبوعة . كما ان مستلمات الأدوية لازالت غير متاحة على شبكة الانترنت وان كانت موجودة فهي موزعة بين قواعد بيانات متفرقة ، وهذا يؤدي الى انعزال المعلوماتية الحيوية عن المعلوماتية الكيماوية ، لذلك صبت الجهود لجسر الهوة بين مختلف مجالات المعلوماتية وكانت في بداياتها موزعة بين Therapeutic target database (TTD) وقواعد بيانات خاصة بالأدوية مثل PharmGKB وموسوعات أخرى . لذلك يعد بنك الأدوية قاعدة مزدوجة للمعلوماتية الحيوية والمعلوماتية الكيماوية تربط بين قواعد البيانات ذات العلاقة والمراجع الطبية .

يحتوي البنك على الأدوية المرخصة من قبل FDA وكذلك الأدوية المنتجة عن طريق التقنيات الحيوية أي البروتينات والبيبتيدات المنتجة بطرق التقنيات الحيوية والهندسة الوراثية ، فضلا عن احتوائه على الأغذية الطبية Nutraceuticals والأدوية التجريبية . له إصدارات عدة كان آخرها الإصدار الثالث لسنة 2011 يعطي بيانات عن تفاصيل الأدوية من ناحية التطيب الصيدلاني مع ذكر الأهداف الدوائية من حيث التركيب والتوالي والمسار الايضي الذي توجد فيه .

ترتيب البيانات في بنك الأدوية

ترتب البيانات في البنك وفق بطاقة الدواء DrugCard لكل دواء يحوي العديد من حقول المعرفة تصل الى 80 حقل وقد زيد في الإصدار الأخير عدد كبير من البيانات الموضحة في الجدول 1.6

جدول 1.6: المفردات الجديدة التي أضيفت الى بطاقة الدواء DrugCard في بنك الأدوية

Chemical kingdom	ChEMBL link
Chemical class	Drug pathway
Chemical substructures	Drug pathway SMPDB ID
Drug manufacturers	Target actions (antagonist, agonist)
Drug packagers	Target priority

Drug prices	Target pharmacological effect (known/unknown/none)
Original patent date	Enzyme actions (inhibitor, inducer, substrate)
Original patent number	Drug metabolite structure
Last patent expiry date	Drug metabolite name
Last patent number	Drug metabolite HMDB ID
MMCD link	Drug metabolite reaction type (e.g. oxidation)
ChemSpider ID and link	Reaction K_m value
NDC ID and link	Reaction V_{max} value
DailyMed link	Metabolizing enzyme references
Drugs.com link	Metabolizing enzyme priority
OMIM link	Drug transporter name
CDPD link	Drug transporter actions (substrate, inhibitor, inducer)
TTD link	Route of elimination
STITCH link	Volume of distribution
BindingDB link	Clearance

تكون صيغ التراكيب الكيميائية SMILES اما البروتينات وتواليات DNA فتكون بصيغة FASTA

طرق البحث في البنك

توفر واجهة البنك وسائل للبحث Built-in-tools ويمكن البحث عن الأدوية فيه باستعمال صيغة SMILES للتركيب الكيميائي . ويمكن البحث عن الدواء باستعمال الرقم التعريفي DrugBank ID اذ تكون الأدوية مرتبة حسب الحروف . ويمكن البحث عن التراكيب الكيميائية من بوابة ChemQuery .

تطبيقات استعمال بنك الأدوية

فضلا عن البيانات التي يوفرها البنك فأن بالإمكان استغلاله في الكشف عن الأهداف الدوائية بالحاسوب ، فضلا عن إمكانية استعماله في تصميم الأدوية وإجراء المسح عليها والكشف عن ايض الأدوية أي بطريقة حدس ايض الدواء وكذلك إمكانية دراسة تداخل الأدوية .

ارتباطات البنك والقواعد الثانوية فيه

يرتبط البنك بروابط مباشرة مع قواعد البيانات ذات العلاقة مثل قاعدة كيوتو وقاعدة البروتينات السويسرية وبنك بيانات البروتينات وبنك الجينات وغيرها من القواعد المهمة التي تزيد على 30 قاعدة . اما القواعد الثانوية التي يحويها فهي :

- (T3DB) Toxin and toxin target database وتعد هذه من مصادر المعلوماتية المهمة الفريدة التي توفر البيانات الخاصة بالسموم وأهدافها . ففي الإصدار الأخير (عام 2011) تحوي القاعدة على أكثر من 2900 من السموم الموصوفة بمرادفات تصل الى 34000،

والسموم تشمل الملوثات والمبيدات والأدوية وسموم الأغذية. وكل هذه ترتبط بسجلات خاصة بها . ترتب السموم في بطاقات السم ToxCard الحاوية على ما يربو عن 50 حقل للمعلومات مثل قيم السمية ،التداخلات الجزيئية والخلوية للسم والتعليمات الطبية وغيرها . مثل هذه المعلومات تم استخلاصها من الآلاف المراجع وقواعد البيانات الأخرى . تركز القاعدة على آليات السمية والبروتينات المستهدفة لكل سم . ويكون هنالك تبادل بين السم والبروتين المستهدف اي يمكن البحث عن السم لإيجاد البروتين المستهدف وبالعكس . وبذلك تكون القاعدة مكانا مناسباً للبحث عن كل ما يتعلق بالسموم من حيث تراكيبها وتأثيراتها .

• **FOODB** وهي تعد أكبر واشمل مصدر لمكونات الأغذية وكيميائيتها وتداخلاتها البيولوجية . وكل مدخل في القاعدة يحوي على أكثر من حقل للبيانات تغطي النواحي التركيبية للغذاء ومعلومات فسلجية مستقاة من المراجع العلمية .

وتقع المعلومات حول مكونات الأغذية في عدة حقول تشمل التسمية ووصف المركب والتركيب الكيميائي واللون والنكهة والطعم وتأثيره الفسلجي وتأثيراته الصحية وتراكيزه في الأغذية المختلفة في حالة توفرها . ويمكن الوصول الى هذه البيانات من نافذة Food view أو Chem view .

• **Small molecule pathway database (SMPDB)** قاعدة بيانات لمسارات أيض الجزيئات الصغيرة في الإنسان ، تحوي على أكثر من 350 جزيئة صغيرة ، ثلثين منها تختص القاعدة بها دون غيرها من القواعد. القاعدة مصممة لتظهر النتائج بأشكال مختلفة وتكون طيبة الاستخدام في الكشف عن المسارات الايضية وفي دراسات الايض ودراسة مكنون النسخ والمكنون البروتيني والمنظومات الحيوية .

مما يسهل للقاعدة أداء هذه المهام هو وجود إمكانية البحث والارتباطات لمواقع أخرى وكذلك الارتباط الى مسارات أمراض الايض. المعلومات التي توفرها عن الجزيئات التي تحويها تشمل الأعضاء التي توجد فيها ، وكذلك المواقع داخل الخلية ومرافقات الإنزيمات ، ومواقع البروتينات ومواقع الايض التي لها علاقة بها . ترتبط بقواعد بيانات مهمة ذات العلاقة والقاعدة سهلة الاستخدام .

تداخلات أخرى مع النواحي الطبية

فضلا عن ما ذكر أعلاه من تداخل المعلوماتية الحيوية والجوانب الصحية هناك العديد من الطموحات التي لا تزال تنتظر التنفيذ . وفي هذا الجزء سيتم تناول بعض الجوانب البسيطة . فمثلا تعد عملية توزيع SNPs بين الجينومات غير متكاملة بالرغم من وجود العديد من الجينومات التي تم تحديد توالياتها . فمثلا يوجد حوالي 10-20 تغاير عام ١ جين ، وفي الإنسان تقدر الجينات بما لا يقل عن 30000 جين ، وتشير التقديرات الى وجود 500000 SNPs في الجينوم البشري ، وهذه تحتاج الى تصنيف ودقة وعمليات خزن وفهم لمعرفة الاضطرابات المحتملة منها ، وتغيرات SNPs يمكن ان تؤثر في عمليات التنظيم ، اذ ان تغير حامض أميني واحد في توالي البروتين يمكن ان يؤدي الى أمراض خطيرة ويتعقد فهم حالات تأثير SNPs بالـ Epistatis أي ان تأثير احد التغيرات يمكن ان يخفي تأثير تغاير آخر وهذه

تحتاج الى نماذج قد تصل الى 5000 نموذج لدراستها وقد تكون اكبر من القدرات البشرية لذا يتم اللجوء الى وسائل المعلوماتية الحيوية .

ومن جهة ثانية توجد هناك علاقات معقدة بين البيئة والوراثة تحتاج الى دراسات معمقة نظرا لما تضيفه هذه العلاقات من شيوع بعض الأمراض بين المجاميع المختلفة بين البشر . وكلها تساهم في تطوير الأدوية .

كما ان للصحة علاقة بعلم الوظائف التي تقف ورائها الإنزيمات ، فوسائل المعلوماتية الحيوية تشير الى ان التشابه العالي في تواليات الحوامض الامينية يمكن ان يرافقه تشابه في الوظائف ، وان البروتينات التي فيها اقل من 30 % من التماثل في توالياتها تكون مختلفة الوظيفة مثل هذه المعلومات لم يتم استغلالها بشكل كامل ولعل أهم الأسباب في هذا عدم وجود التوحيد في صيغ وملفات البيانات لتحديد مفردات علم الحياة وهذا قائم التأكيد عليه في مواقع أخرى من الكتاب وهو إيجاد Gene ontology . لتلافي مشكلة اختفاء المعلومات عند تحديث البيانات وظهور بيانات جديدة أي عدم انتظام البيانات في الإصدارات المختلفة .

وتهدف المعلوماتية الحيوية ضمن الآفاق المستقبلية طرح مشروع الإنسان الافتراضي Virtual human project (VHP) والذي يهدف الى تمثيل رقمي لوظائف الإنسان على المستوى الجزيئي للحياة . فضلا عن طرح مشروع المكنون الوظيفي Physiome project ويهدف الى الربط بين الوظائف والجزيئات وتداخل البيئة معها . ومن كل هذا يحاول المختصون إيجاد الإنسان الرقمي Digital human (DH) وهو مشروع يتعامل مع حل المشاكل وعلاقة البيئة بصيغ الأرقام لتستوعبها الحواسيب .

وتحتاج الجوانب الصحية الى قواعد بيانات لمنتجات هدم البروتينات أي قواعد تهتم بعمليات انقلاب البروتينات ومسارات تحطيمها ذات العلاقة بالأمراض كما في بعض السرطانات والأمراض التحليلية العصبية والاضطرابات الالتهابية ومثل هذه القواعد تشمل آليات التفكك ، والتحويلات بعد الترجمة ، ومستويات التركيب و التداخل مع البروتينات الأخرى .

تحتاج الجوانب الطبية الى دراسة DNA المايوتوكونديريا mtDNA و SNPs والطفرات الحاصلة فيه والتي لها علاقة بالعديد من الأمراض منها بعض السرطانات والعقم وداء السكري والصمم (الطرش) والأمراض العصبية التحليلية .

كل ما ذكر أعلاه من الاحتياجات والمجالات يراد بها الوصول الى ما يسمى بالطب الالكتروني e-Medicine أي البحث الطبي والعناية الصحية في عالم الانترنت وهذه تنمو بسرعة وبشكل فعال نتيجة لتطور الوسائل على كافة الأصعدة .

ويمكن ان تهدف هذه الى التطبب الشخصي والعناية الصحية الشخصية وهذه تعتمد بطبيعة الحال على الجينوم الشخصي وسجلات المريض وساعد في هذا Secure Health and SHAMAN Medical Access Networks التي تهدف الى تصميم الأدوية الشخصية وهذا يعني تحويل سجلات المريض وبياناتها الى صيغ رقمية . ويجب ان يشمل SHAMAN على بيانات عن الأحياء الممرضة وقواعد بيانات حول الموديلات الحيوانية لأمراض الإنسان . فضلا عن مشاركة جوانب أخرى مثل الوراثة الجزيئية والوراثة السريرية التي تساعد في ربط المعلومات التركيبية والوظيفية ضمن الأنماط السريرية Clinical phenotypes . كل هذه يمكن ان توصل الى التطبب الشخصي كما هو الحال في محاولة إيجاد التغذية الشخصية .

مكون الإنسان الميكروبي Human microbiome

الموضوع من المواضيع المتداخلة مع غيره مثل الايض والصحة . والمعروف ان جسم الإنسان يعج بالميكروبات ويصل تعدادها 10^{13} - 10^{14} من الأحياء الذي يصل الى عشرة أضعاف عدد خلايا جسم الإنسان مكونة من العديد من البكتريا الموجبة والسالبة لصبغة كرام والاراكيا والفطريات والفيروسات وكذلك الطفيليات التي تعيش داخل أو خارج الجسم وتحوي ملايين الجينات مقارنة بجينات الإنسان التي تتراوح أعدادها 20000 - 30000 . ودراسة الجينوم لبيئة الإنسان وطرق الحاسوب السريعة وفرت الفرصة المهمة لتقنيات بناء الشبكات الايضية وتتم بمساعدة تحليل المتغيرات المتعددة Multivariate التي تربط بين التغيرات في المسارات الايضية والتدرج في العوامل البيئية . والسؤال الأهم هو حول الكيفية لربط الاختلافات في مكونات المجتمع مع الوظائف المختلفة والتأثير في صحة الإنسان ومرضه ومثل هذه الدراسات تستعمل التوائم المتماثلة . اذ ان المكون الميكروبي (أو ما يمثل المجتمعات الميكروبية) لها التأثير الأكبر في توليد الأمراض وتوازنها مع الصحة .

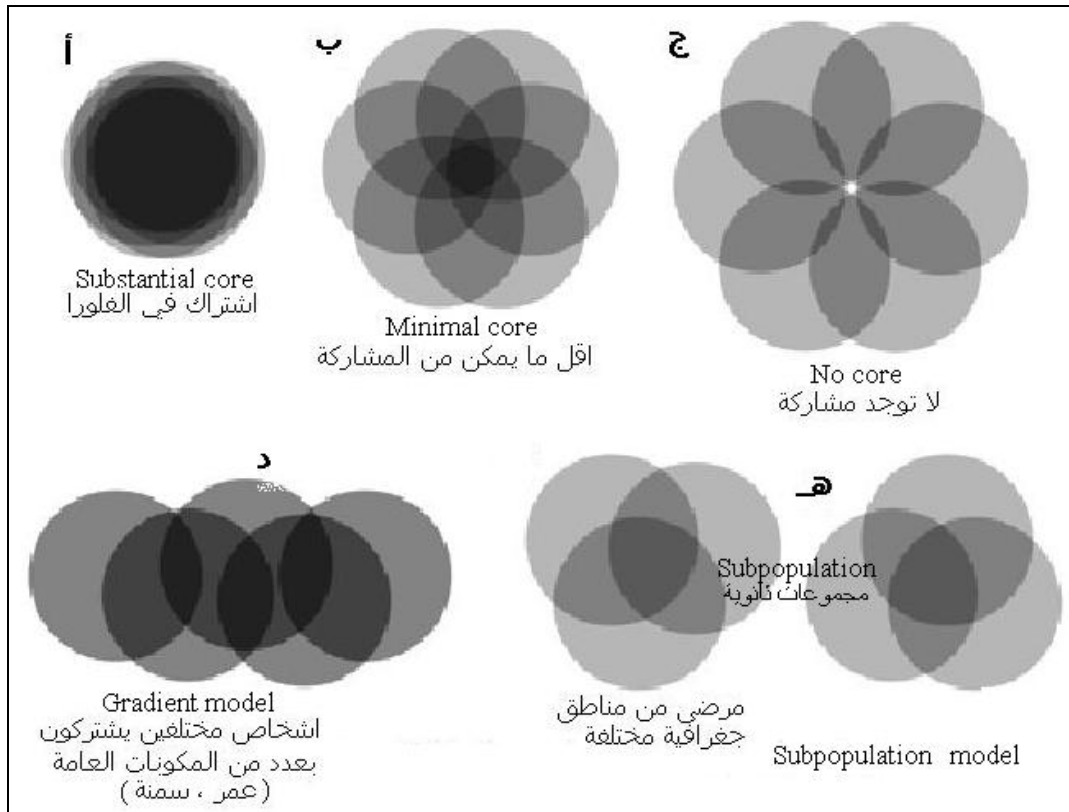
توزيع الأحياء المجهرية في جسم الإنسان

تتوزع الأحياء المجهرية على مناطق معينة من الجسم وهي الجلد والفم و الأنف والقناة الهضمية Gut والمهبل أي انها تعيش في وكذلك على جسم الإنسان وتصل الى عشرات الملايين كما ذكر أنفا مما دعا الى اعتبار الإنسان كائن خارق Superorganism أي انه نظام بيئي معقد وان جينومه أو مكونات جينومه تقع ضمن ما يسمى بجينوم المجتمع Human metagenome . وقد تم تحليل مئات الجينومات للميكروبات واكتشف ان هناك جينات وبروتينات جديدة لا يشفر لها بالجينوم البشري . وبذلك تبين ان الجينوم الأعم والأغلب هو للميكروبات كما أظهرت نتائج تحديد تواليات 16S rRNA في البكتريا والاراكيا و 18S rRNA لحقيقيات النواة ، تخدم في وظائف لا يقوم بها الجينوم البشري وإنما ما يحويه من الأنواع والأجناس المختلفة من الأحياء . والمكون الميكروبي يختلف من شخص لآخر حتى وان كانوا في البيئة نفسها فضلا عن الاختلافات التي سجلت للتوائم المتماثلة .

مشروع المكون الميكروبي البشري

على ضوء الاختلافات في الميكروبات وضع مشروع المكون الميكروبي Human microbiome project (HMP) وهو مشروع تعاوني عالمي لمعرفة بايولوجية الإنسان وتفاعله مع الأحياء المرافقة له لإدامة صحته ، بدأ عام 2007 يهدف الى وضع المقاييس والأسس الأخلاقية والقانونية والاجتماعية وتأثيرها في مجال البحث في مكون الإنسان الميكروبي فضلا عن البرامج والوسائل اللازمة للبحث . فمن المعروف ان البيئات المتطرفة مثل مابزل التعدين ألحامضي يقل فيها التنوع الميكروبي ولكن في بيئة مثل جسم الإنسان يزداد التنوع فيها توافقا مع البيئات الموضعية ، فالملحظ ان القناة الهضمية تعج بالعديد من الأنواع والأعداد الميكروبية وينخفض هذا التنوع في منطقة المهبل القريبة منها لعدة اعتبارات . والاختلافات الكبيرة بين الأشخاص عقدت هذا المشروع ، ويهدف المشروع الى إيجاد المكون الميكروبي المركزي Core microbiome التي يشترك فيه الجميع ولكن الاختلافات الكبيرة

لم تسمح بمثل هذا الاستنتاج مما أدى الى استعمال وحدات تصنيفية عالية مثل الشعب أو الاعتماد على الجينات الفعالة بدلا من البحث عن الأحياء على مستوى النوع من التصنيف . ولذلك كان من الضروري إجراء دراسات تربط بين البيئة والأحياء المجهرية والدراسات الجينومية وطرق حاسوب متقدمة وبمرافقة الدراسات المعتمدة على عمليات الزرع . ونتيجة كل هذه الدراسات تم الخلوصل الى استبعاد وجود مكنون ميكروبي مركزي كبير على مستوى النوع ولكن قد تكون واردة على مستوى الشعب Phyla . ونظرا لأهمية المكنون المركزي في تحديد حالة الصحة والمرض وعليه فاذا كان هناك ولو حد أدنى من المكنون المركزي أو عدم وجوده بالمرّة فيجب استعمال استراتيجيات بديلة للدراسات لانه في كل مرة سيكون هناك جينات جديدة عند فحص شخص جديد . والافتراضات المطروحة موضحة في الشكل 7.6 :



شكل 7.6 : فرضيات اشتراك المجاميع البشرية في المكنون الميكروبي

حيث تمثل الحالة الأولى (أ) وجود مكنون ميكروبي مشترك أو جينات مشتركة بين الأشخاص ، اما الحالة (ب) فهي تمثل الاشتراك الأدنى أي تكون هناك أحياء قليلة مشتركة بين الأشخاص . والحالة (ج) تشير الى عدم وجود مكنون مركزي والأشخاص لا يشتركون بأي شيء ، اما في الموديل المتدرج (د) فيشير الى وجود تدرج بين الأشخاص المشتركين بالتعاقب مثل أشخاص بأعمار مختلفة متدرجة أو أشخاص بدينين يشتركون بنسبة معينة . اما النموذج الأخير (هـ) فيمثل نموذج المجتمعات الثانوية Subpopulation model مثل أشخاص من مناطق جغرافية

مختلفة أو أشخاص لديهم أمراض مختلفة ولذا لا يكون شئ مشترك بين المجاميع الثانوية . وهذه التوزيعات تمثل الإستراتيجية لتحديد الأنواع المركزية لتحديد التواليات ثم استعمالها كأساسات للدراسات الجينومية وهذه الإستراتيجية تواجه بعض المشاكل . اما المشكلة الأخرى فهو حصول تغاير في كميات مكونات المجتمعات المتعلقة بالإنسان فمثلا يتغير مستوى وجود Firmicutes و Actinobacteria و Bacteroides التي لها علاقة وثيقة بالسمنة .

وقد يكون هناك محتوى مركزي ولكن لا يمكن إيضاحه فمثلا الاعتقاد السائد ان *E. coli* التي تعد بكتريا معوية تقليدية ولكن مجموعة γ -proteobacteria التي تضمها لا تشكل إلا اقل من 1 % من بكتريا القناة الهضمية . ولكن بما ان البكتريا قادرة على النمو في الأوساط الزراعية لذلك يمكن الكشف عنها بالرغم من انخفاض مستوياتها العددي . أي ان احتمالية ان الناس يمكن ان يشتركوا بالأنواع البكتيرية ولكن الاختلاف يكمن بالوفرة بحيث تؤثر في الصحة والمرض . وهذا يمكن الكشف عنه باستعمال طرق حساسة وحديثة مثل استعمال تحديد التواليات بالاعتماد على الفوسفات Pyrosequencing بدلا من استعمال طرق التحديد العادية اذ يمكن ان تكشف عن الأحياء ولكن تكون وفرتها اقل من 1 % أي المستويات قليلة بدقة معقولة ، والطريقة أسرع وارخص فضلا عن عدم الحاجة الى الكلونة والمكتبات الوسطية . وكذلك اعتماد المعلومات الموجودة في 16S rRNA فمثلا تحديد 200 قاعدة التي تشكل 12 % من بيانات 16S rRNA تكون كافية للعنقدة باستعمال برامج المعلوماتية الحيوية بدقة تصل الى دقة النتائج التي يمكن الحصول عليها من استعمال 70% من الطول الأصلي للجزيئات . وبما ان دراسات الجينوم المجتمعي تواجه صعوبة في قصر التواليات التي يمكن الحصول عليها من النماذج البيئية والتي لا تساعد في تحديد وظيفة الجين أو النوع من هذه المقروءات القصيرة ، ولكن وجد ان القطع 250-400 قاعدة تساعد بشكل أفضل على ان تكون هناك أكثر من 350 توالي للنموذج ، ويعتمد عدد التواليات على البيئة ، فمثلا البيئات قليلة المواد الغذائية مثل مياه البحر أشارت فحوص الإحصاء الى ان 17 نوع من التواليات يمكن ان تعطي عناقيد للميكروبات كافية للدراسات ، في حين يكون العدد 100 من التواليات في حالة القناة الهضمية في اللبائن .

ويعد المشروع الى التحليل النوعي أي إيجاد العلاقة بين أفراد مجتمع الأحياء وكذلك إيضاح النواحي الكمية أي تحديد تركيب المجتمع وتظهر على شكل وحدات تصنيفية Operational taxonomic units (OUTs) التي تستعمل في وضع الأشجار أو العلاقات التطورية باستعمال برنامج الجوار الأقرب (المراد الذكر) وفيه تضاف التواليات الى الوحدات التصنيفية OUTs اذا كانت مشابهة للتواليات الموجودة OUTs أصلا . أو استعمال طريقة الجوار الأبعد وفيه يضاف التوالي OUT اذا كان مشابهها لكل التواليات OUTs . ويمكن إضافة التواليات غير المصنفة الى الأشجار باستعمال برامج خاصة بعد تحديد الاصطفاف والتشابه باستعمال برامج BLAST كما في حالة البكتريا غير القابلة للزرع ، وتساعد في ذلك البيانات الإضافية حول النموذج باستعمال مفردات ملائمة . وعلى مستوى جينوم المجتمع يكون التوجه الأمثل هو تمثيل النتائج بشكل خارطة تشبه الخرائط الحرارية Heat map التي تشير الى وفرة كل وحدة تصنيفية في النماذج البيئية . كل هذه ستكون مهمة في جعل المعلومات المجمعة في مشروع HMP مصدرا مفيدة للدراسة .

العوامل المؤثرة في مكنون الإنسان الميكروبي

تبين مما ذكر أعلاه ان هنالك تشابها بين الأحياء التي تقطن الإنسان البعض منها مرضية في حين ان البعض الآخر ضروريا لصحته . وعند اضطراب التوازن فيما بينها تنتج الأمراض . وتساهم الظروف البيئية ضمن العوامل المؤثرة ومنها التهوية ووجود الأوكسجين ، فمثلا عند استئصال جزء من الأمعاء ileum وزراعة أجزاء أطول منه لغرض زيادة امتصاص الغذاء تتغير الفلورا بين المراحل فعندما تكون مفتوحة ومعرضة للأوكسجين تكون العصيات اللبينة وبعض البكتيريا المعوية هي السائدة وبعد الغلق تتحول الظروف الى ظروف لا هوائية وتعود البكتيريا اللاهوائية مثل Clostridia و Bacteroides للسيادة كما دل تحليل مواد الايض من النماذج .

ومن العوامل الأخرى هو توازن الطاقة التي يمكن الحصول عليها من المواد الغذائية فيلاحظ في حالة السمنة يحصل تغاير كبير على مستوى شعب الفلورا الميكروبية مما يؤدي الى تنوع الأحياء وتغير في الجينات والمسارات الايضية ، وتختلف عن حالة النحافة أي انها تعتمد على مؤشر كتلة الجسم (BMI) Body mass index الذي يقاس بـ كغم/م² والحد الفاصل للسمنة عندما يكون المؤشر (BMI ≥ 25 and < 30) وبعدها تعد سمنة وزيادة وزن ، ويمكن التوصل الى حالة السمنة بمساعدة مقياس CAzyme الإنزيم النشط في أيض الكربوهيدرات وتحديد تواليه باستعمال برنامج FASTAY.

تأثيرات الفلورا

بعد الولادة تستعمر الأحياء المجهرية أجزاء وأعضاء في الجسم وخاصة الأمعاء وتؤثر في المضيف . ان تحليل الدراسات الناتجة من دراسة جينوم المجتمع في أي موقع من الجسم مثل الأمعاء يمكن ان يوضح أي الأحياء القاطنة فيه والتي تتداخل مع الخلايا الطلائية للقناة الهضمية بطريقة متعادلة وإيجابية ولكن بعض الأحيان قد يكون تأثيرها سلبي كما في حث سرطانات القولون وغيرها من المواقع . ومن جهة ثانية فان المهام التي تؤديها الأحياء المستوطنة هي المساعدة في هضم المواد ، وكذلك تدريب الجهاز المناعي ، وفي البالغين فهي تضطلع في المساعدة في هضم السكريات النباتية المعقدة ، أي انها تقوم بمسارات ايضية لا يقوم بها جسم الإنسان .

مكنون القناة الهضمية الميكروبي Gut microbiome

تضم القناة الهضمية الأحياء التي يستفاد منها جسم الإنسان مقابل ان تؤدي فوائد للمضيف لغرض إقامة العلاقة السليمة . وتقع دراسات هذا الموضوع ضمن ما يسمى Metagenomic of human intestinal tract (Meta HIT) ويركز هذا الحقل على معرفة الفلورا الميكروبية في حالة السمنة وأمراض الجهاز الهضمي الأخرى ويكون التركيز على الجينات والبروتينات التي تنتجها . وهذا المجال يعد الجزء المكمل لمشروع HMP الذي يهدف الى تحديد التواليات لمئات من الجينومات البكتيرية ووضعها في قواعد بيانات.

تستعمل طرق تحديد تواليات جينات 16S rRNA في الكشف عنها نظرا لان 80 % من الأحياء غير قابلة للزرع . وأشارت الدراسات الى الاختلاف بين البالغين والأطفال وأوضحت دراسة الجينوم البيئي الى وجود حوالي 9.2 % من البكتيريا المتعايشة في الأمعاء و 4 %

أحياء مرضية و 2.7 % مجاميع أخرى وهذه النسب تختلف في الأطفال ، وفضلا عن الاختلافات الكبيرة بين الأشخاص توجد الاختلافات في مجموعة البالغين البدينين والنحفاء مما يشير الى العلاقة الوثيقة بين فلورا القناة الهضمية والسمنة . والملاحظ ان انخفاض الوزن في البدينين يؤدي الى تقارب في الفلورا لتصبح تشابه النحفاء ومن هذه يمكن مقارنة النتائج لمعرفة إمكانية استعمال الأدوية أو الأحياء العلاجية لتشكيل الفلورا الملائمة . وكذلك وجود الاختلافات في حالة الصحة والمرض التي تتمثل في العلاقة الوثيقة بين Inflammatory bowel disease (IBD) و فلورا الأمعاء ، اذ ان المصابين يكون التنوع عندهم منخفضا مقارنة بالأصحاء . وتظهر هذه الاختلافات بين التوائم المتماثلة ولكن بشكل اقل حدة .

وتستعمل طريقة تحديد تواليات 16S rRNA لدراسة وبائية الأمراض وكذلك الحساسية والسرطانات . ومجمل ما أشارت إليه الدراسات هو وجود حوالي 1800 جنس وعدد كبير من الأنواع وان أكثر مستوى تصنيفي يحصل فيه التباير هو السلالات . ووجد ان 98 % من الأنواع تعود الى أربعة أقسام من البكتيريا وهي Firmicutes (64%) و Bacteroides (23%) و Proteobacteria (8%) و Actinobacteria (3%) .

وعلى العموم فان لهذا الاختلاف العديد من الأسباب . فمثلا في الأطفال بعمر من الولادة الى سنتين يتأثر مكنونهم الميكروبي بطريقة الولادة فيما اذا كانت طبيعية خلال المهبل أو بطريقة العمليات القيصرية وكذلك طريقة الرضاعة فيما اذا كانت طبيعية أي رضاعة الثدي أو رضاعة صناعية . و فلورا الأمعاء في المواليد تتمثل بالعصيات اللبنية والمكورات العنقودية والمكورات المسبحية وكذلك البكتيريا المنشطرة Bifidobacteria . وإضافة الى تأثير العمر فهناك ظروف أخرى مؤثرة منها الغذاء ونمط الحياة والتوزيع الجغرافي أي ان هذه المجاميع الميكروبية تتشكل بناء على الضغوط الانتخابية. ومن المعروف ان تحديد التواليات بالاعتماد على جينات 16S rRNA لا تعطي فكرة عن الوظائف ولكن دراسة جينوم المجتمع يكون وسيلة أو طريقة فعالة لتحليل وتجميع المعلومات حول الجينات الموجودة في المجتمعات الميكروبية وبما ان الكثافة التشفيرية للبكتيريا عالية وتصل مستويات أعلى من 80 % لذلك فان التواليات Metagenomics sequence التي يتم الحصول عليها لابد ان تحوي على مناطق مسئولة عن الوظائف . وقد أعلنت البيانات حول جينوم المجتمع او الجينوم البيئي (Metagenome) لمكنون الميكروبات للإنسان والفيران ، ووجد ان هنالك العديد من الجينات البكتيرية هي المسئولة عن وظائف لا يقوم بها الجينوم البشري وانما تقوم بها الفلورا منها ما يشارك في أيض الكلايكان Glycan أو الحوامض الامينية أو المواد الداخلية وكذلك تخليق بعض الفيتامينات والمواد الأخرى التي تكون مهمة في بايولوجية الإنسان ، وهذا ما يولد العلاقة الوثيقة بين الإنسان والفلورا التي يؤويها وان الموقف هو كائن خارق مكون من الإنسان والفلورا الميكروبية . ولمعرفة هذه الوظائف بالنسبة لجينات الجينوم البيئي تستعمل الفيران الخالية من الجراثيم Gnotobiotic mice .

والفلورا الطبيعية في القناة الهضمية في الإنسان تقوم بشكل أساسي في أيض الكربوهيدرات المكونة وتنتج الحوامض الدهنية ومنها البيوترات Butyrates التي تعد مصدر الطاقة الرئيس لخلايا المضيف في تلك المناطق . فضلا عن ان البكتيريا أو الميكروبات المتعايشة

تكافح الممرضات وذلك بتحفيز الجهاز المناعي وكذلك الخلايا الطلائية مثل Paneth cells والخلايا الشجرية التي تكون المصدر الرئيس للمواد المضادة للميكروبات مثل إنزيمات Lysozyme و Phospholipase والبيبتيديات المضادة للميكروبات مثل α -defensins وغيرها .

وحصول اضطرابات في هذه المواد يؤدي الى الأمراض مثل Crohn's disease . وقد وجد ان بعض الببتيدات المضادة المذكورة أنفا يزداد التعبير عنها بعد الولادة ولكن ذلك يعتمد على طريقة الولادة .

ومن جهة ثانية فان من المؤثرات الرئيسة في تحديد المجاميع البكتيرية في أمعاء الإنسان هو أملاح الصفراء والإنزيمات الرئيسة التي تتعامل معها هو إنزيم تحليل أملاح الصفراء Bile salt hydrolase (BSH) ويعد الإنزيم البوابة للتعامل مع أملاح الصفراء . وتكون أملاح الغذاء بمثابة جزيئات إشارة لتنظيم عمليات أيضا وتؤثر في امتصاص الدهون وتوازن الكوليسترول والدفاع الموضعي في الأمعاء ، لذلك يعد الإنزيم البكتيري من المؤثرات القوية في أيض المضيف ، كما ان هذه الإنزيمات تساعد الأحياء لاستيطان القناة الهضمية والحفاظ على النظام البيئي الميكروبي فيها ، وباستعمال دراسات الجينوم البيئي وجد ان الإنزيم موجود في الأغلبية العظمى من أقسام البكتيريا والاركية المعوية .

ان المتوفر من الكميات الكبيرة من البيانات والتواليات التي تصل الى الملايين لتغطي آلاف الأنواع الموجودة في الأصحاء وغير الأصحاء تحتاج الى حواسيب متطورة Supercomputers ووسائل معلوماتية حيوية كفوءة للبحث والتحري في قواعد البيانات التي يجب ان تتشأ لتحتوي البيانات . وفي الوقت الحاضر تتم عمليات مقارنة تواليات DNA من أمعاء الإنسان وغيرها من النماذج باستعمال برامج BLASTP و tBLASTN ، وقاعدة البيانات الخاصة بالقناة الهضمية Human gutome database لازالت الى الوقت الحاضر (2011) تحت الإنشاء.

مكونون الجلد الميكروبي Skin microbiome

يدخل الجلد ضمن مشروع مكونون الإنسان الميكروبي ، وتهدف الدراسات في هذا تحديد المكونون المركزي وذلك لأنه مهم في فهم وتطوير العلاجات لأمراض الجلد مثل حب الشباب (Acne) والتهاب الجلد الوراثي (الأكزيما Eczema) وقد درست جينات البشرة ووجد ان الميكروبات التي لها علاقة بالأمراض المذكورة أعلاه هي بسبب *Staphylococcus aureus* و *Propionibacterium acne*

مكونون الفم الميكروبي Oral microbiome

في مجال مكونون الإنسان الميكروبي تمثل فلورا الفم أكثر الجوانب دراسة و يحوي التجويف الفموي على بكتيريا يكون 35% منها غير قابلة للزرع كما اتضح من تحديد تواليات جينات 16S rRNA وقد انشأت قاعدة بيانات خاصة هي Human oral microbiome database (HOMB) . وفيها معلومات تصنيفية وجينومات الأحياء الموجودة فيها التي يكون لها شفرة تعريفية خاصة بها Human oral taxon (HOT) ، وكل منها له روابط لعدد من مصادر

المعلومات ، فضلا عن وجود الكم الهائل من وسائل المعلوماتية الحيوية لتحليل تواليات الجينومات .

تحتوي القاعدة معلومات حول أهم الاضطرابات التي تصيب الفم مثل التسوس والتهاب اللثة وسرطانات الفم وغيرها . والبيانات الموجودة في القاعدة على درجة عالية من الجودة اذ انها مشدبة بشكل جيد . وتحتوي الأنماط التطورية Phylotype أي مجموعة من الأحياء تضم الى بعضها اذا كان هنالك 98.5 % تشابه كحد فاصل لوضعها في وحدات تصنيفية خاصة Taxa . والقاعدة تضم حوالي 619 من الوحدات التصنيفية بعد آخر تحديث لها (2011) وهي الوحدات التصنيفية للأنواع المعروفة . لكل وحدة تصنيفية HOT ID وهو التعريف لكل وحدة.

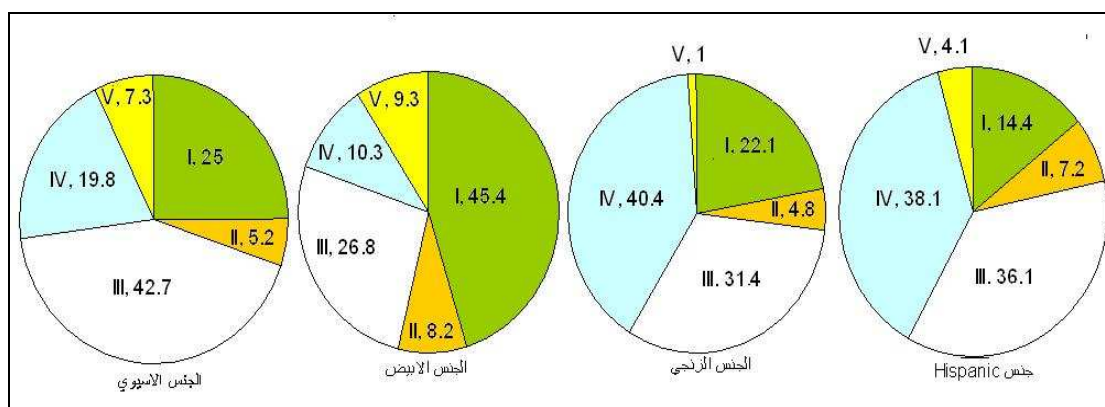
وتحتوي القاعدة على الأنواع التي لم تعرف لحد الآن تحت مسمى (Unnamed) . ويختلف التصنيف في القاعدة عن ذاك المستعملة في NCBI ، اذ ان التصنيف يعتمد على العلاقات التطورية كما ان الأحياء تعطي مرادفات لتسمية الأنواع . ونظرا لعدم وجود تصنيف للأنواع غير المسماة في NCBI فهذا كان وراء إنشاء HOMB التي تحدث كل أسبوعين . توفر القاعدة ارتباطات الى قاعدة المراجع ذات العلاقة ونظرا لارتباطها بالـ NCBI فيمكن البحث فيها عن النيوكليوتيدات والبروتينات بواسطة نظام البحث Entrez . والمعلومات التي يمكن الحصول عليها بالنسبة للأحياء القابلة للزرع تشير الى الأوساط الغذائية الخاصة للتنمية فضلا عن وصف متكامل .

الميزة الرئيسة للقاعدة انها تحوي مجمع شامل للتواليات المرجعية لجينات 16S rRNA للوحدات التي تضمها . فضلا عن احتوائها لبرامج BLAST ووسائل لإظهار الجينومات. وكذلك تحوي صفحة للبيانات الإضافية المتجددة لكل وحدة تصنيفية ، وتحتوي صفحة البيانات الإضافية على تفصيل للبايولوجي الجزيئي والتفصيل التصنيفي وملف التهميش وبرامج لتحديد القطيفات وبرامج للكشف عن المسارات الايضية المرتبطة بقاعدة كيو تو وقاعدة الإنزيمات ENZYME . والحقيقة تعد القاعدة الأساس لمواقع الجسم الخاصة مع التأكيد على فلورا الفم أي ان البنى التحتية هي للأحياء المجهرية الموجودة في الفم ولكنها يمكن ان تستعمل لمواقع أخرى في الجسم . والقاعدة تشمل أشجار تطورية مبنية على تواليات 16S rRNA ، ويمكن ان تستعمل كنموذج لإنشاء قواعد بيانات لمواقع أخرى مثل القناة الهضمية والجلد والمهبل .

مكونون المهبل الميكروبي Vaginal microbiome

فلورا المهبل الطبيعية تعد من الضروريات لحفظ الصحة الإنجابية وسير عملية الحمل بشكل طبيعي واضطرابات توازن الأحياء يمكن ان يؤدي الى العديد من المشاكل مثل الإصابات والعقم . والأحياء الشائعة هي العصيات اللبنية المنتجة لحمض اللاكتيك في حالة الصحة لذا يعد ارتفاع عدد العصيات اللبنية وانخفاض الرقم الهيدروجيني pH عن 4.5 مؤشرا صحيا ، ولكن مع هذا فهناك بعض حالات الصحة التي لا توجد فيها العصيات اللبنية وانما البدائل المنتجة للحامض وبذلك يكون انخفاض الرقم الهيدروجيني هو الدليل المرشح للتنبؤ بصحة المهبل . والبحث عن المكونون الميكروبي المركزي لم يظهر ان هنالك مثل هذا التجمع المركزي أو الأساسي على الأقل في النساء في سن الإنجاب وفقا لنتائج البحث عنه باستعمال البروتين الوصيف Chaperonin 60 . وهنالك عدد من الأسباب التي تؤثر في نوعية الأحياء التي

تقطن المهبل فالأمراض والإصابات تكون من الأسباب المؤثرة في ظهور الاختلافات وكذلك يؤثر نمط الحياة والعناية الصحية والسيطرة على الولادة وأكثر العوامل المؤثرة هو المضيف كما يتبين من تأثير العرق وهذا قد يعود الى الاختلاف في المناعة الفطرية Innate immunity وتطبع الجهاز المناعي والاختلاف في مستلمات الطبقة الطلائية المبطنة . وقد أشارت دراسات جينوم المجتمع الى ان الأحياء تتوزع على الأعراق المختلفة ضمن خمس مجاميع يكون نمطها خاص بكل عرق كما موضح في الشكل 8.6 :



شكل 8.6 : نمط توزيع مكنون المهبل على الأعراق المختلفة

اذ تمثل المجموعة الأولى I البكتريا *Lb . crispiths*

والمجموعة الثانية II *Lb . gasseri*,

والمجموعة الثالثة III *Lb . iners*

والمجموعة الرابعة IV مجاميع مختلفة

المجموعة الخامسة VI *Lb . jensenii*

تطبيقات دراسات المكنون الميكروبي والجينوم المجتمعي

ان الجينوم المجتمعي وكذلك مشروع مكنون الإنسان الميكروبي لها تطبيقات إضافة الى التي ذكرت أعلاه في مواقعها الخاصة منها

- الجينوم البيئي او المجتمعي يستعمل في تحديد صلاحية المياه ونوعيتها ، وفي هذا المجال يتم البحث عن واسمات يمكن ان تحدد مصادر التلوث مثل وجود مكنون المجترات الميكروبي Ruminant microbiome وغيرها. وبذلك فهو يمكن ان يستعمل في تسجيل نوعية المياه وبالتالي يسهل معالجتها .

- تهدف المشاريع المذكورة أعلاه الى التوسع لدراسة المجمعات الوراثية (الجينومات) في مجتمعات مختلفة من الكائنات وضم عناصر أخرى مثل الفيروسات والطفيليات وغيرها من الكائنات المجهرية .

المعلوماتية الحيوية والجهاز المناعي

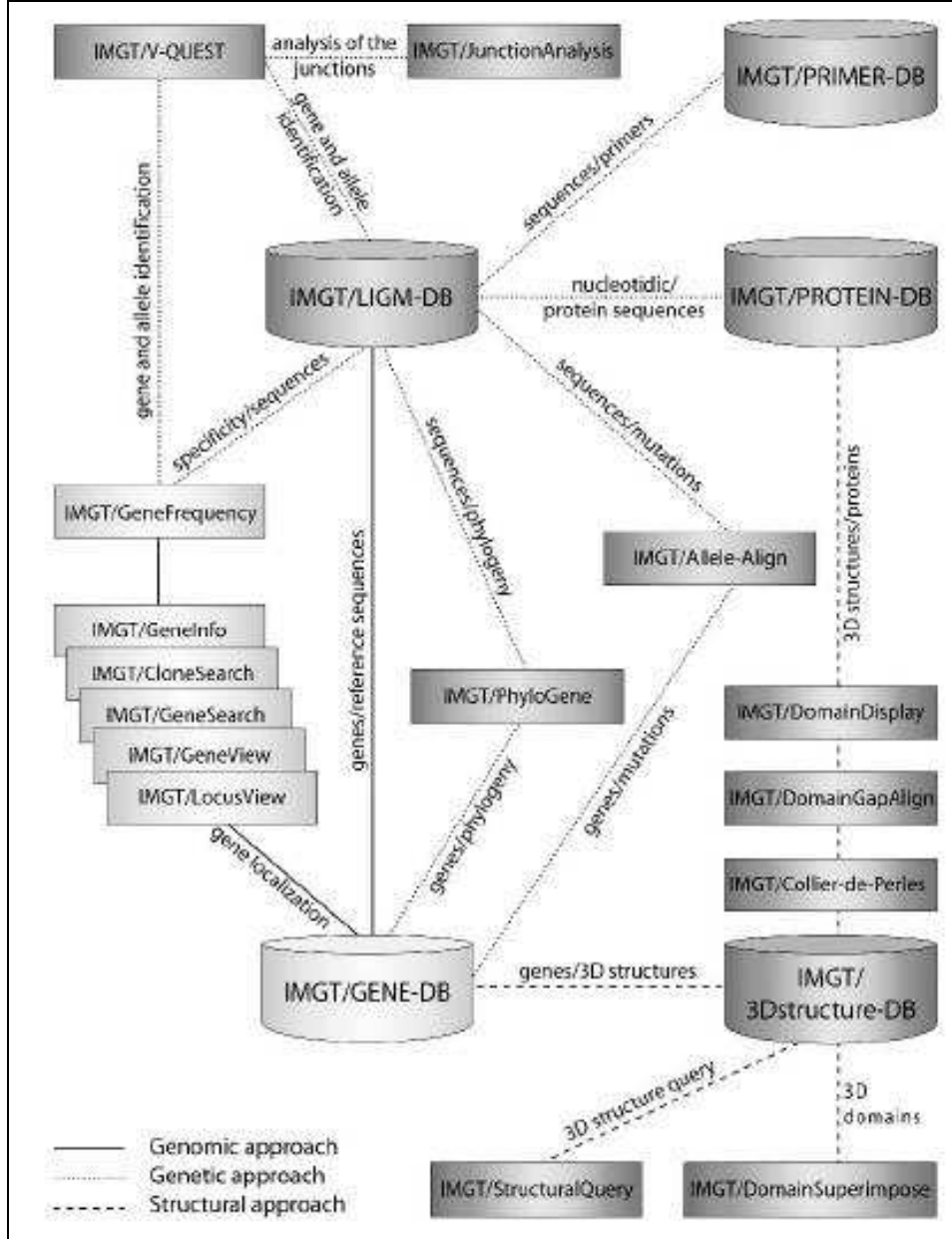
لا يخفى على احد أهمية الجهاز المناعي فضلا عن تعقيداته الواسعة التي تفوق القدرات البشرية في معرفتها ودراستها . ودخول المعلوماتية الحيوية في مجال المناعة كان لابد منه ودخلت تحت مسمى المعلوماتية المناعية أو علم المناعة الحاسوبي Computational immunology . وكانت المعلوماتية الحيوية طريقا لكثير من التطبيقات مثل إيجاد البرامج التي تحدد قابلية المواد على ان تكون مستضدات مثيرة للمناعة وتحديد مواصفاتها عاملا مهما في إنتاج اللقاحات Peptide vaccines ، فضلا عن ترشيح البعض منها لتكون أهدافا لإنتاج الأجسام المضادة تجاهها ، فمثلا البروتينات الغشائية أو الموجودة على سطوح الخلايا تكون هي الأفضل للاستهداف باللقاحات . وكذلك وفرت المعلوماتية الحيوية إمكانية تقليل القابلية المستضدية وذلك بإحداث الإحلال في الحوامض الامينية ضمن ما يسمى إزالة التمنيع Epitope direct deimmunization وبذلك لا يثير المستضد الخلايا التائية Th ويتم على ضوءه إخماد التفاعلات اللاحقة .

ويتعامل علم المناعة الحاسوبي بثلاث مستويات

- المستوى المجهرى Microscopic level وفيه تدرس إمكانية تخليق الجزيئات مثل DNA والبروتينات ودراسة إمكانية تفككها وتحديد مواقعها في الخلية وكذلك فعاليتها مثل التعبير الجيني ، وآليات تغير دورة الخلية وامتصاص المواد الغذائية الأساسية وتعطيل المستلزمات الخلوية ونقل الإشارات الكيماوية داخل الخلية .
 - المستوى الثاني Mesoscopic level ، وهنا يكون التعامل على مستوى الخلية ومجمعات الخلايا مثل صفات التجمع والتكتل كذلك التعاون بين الخلايا والتنافس والتناضح الدموي .
 - المستوى الثالث Macroscopic level وتكون الدراسات فيه على مستوى الأنسجة مثل هجرة الخلايا والاستجابات الميكانيكية والتداخل والتواصل بين الأنسجة .
- واغلب هذه المستويات تتخذ من فيروس HIV وتداخلاته موديلات للرياضيات والحسابات للجهاز المناعي.

قاعدة IMGT

تعد القاعدة أهم ما تمخض عنه دمج علم المناعة مع المعلوماتية الحيوية و IMGT مشتقة من ImMunoGeneTics انشأت القاعدة عام 1989 في جامعة Alberta والهيكل العام للقاعدة موضح في الشكل 9.6 :



شكل 9.6 : واجهة قاعدة IMGT الخاصة بعلم المناعة

وتعد القاعدة مجالا واسعا وشاملا بأغلب ما يتعلق بالجهاز المناعي ، ولها أنظمتها الخاصة في الترميز لمكوناتها سواء باستعمال الألوان الخاصة أو أنماط التعبير الأخرى . ويمكن إجمال شطر كبير لما تقدمه القاعدة من خلال القواعد التي ترتبط اليها ويكون IMGT web resources الواجهة للمعلوماتية المناعية وتدرج فيه قواعد البيانات التي يضمها الموقع ويوصل بروابط بهذه القواعد ومنها :

- قواعد للكلوبيولينات Igs
- قواعد بيانات لمستلمات الخلايا التائية T cells receptors (TRs)
- قواعد بيانات خاصة بجزيئات MHC (Major histocompatibility complexes)
- قاعدة بيانات خاصة بالمستضدات وهذه تتكون من قواعد بيانات خاصة بالبيتيدات التي ترتبط بـ MHC (pMHC) .

- المستضدات المثيرة لتخليق الأجسام المضادة.
- قاعدة بيانات الحواتم Epitome .
- قواعد بيانات خاصة بالمستضدات الناقصة Haptens .
- البيتيدات التي لها تأثيرات مضادة للميكروبات أو المضادات الحيوية البيتيدية
- قاعدة بيانات خاصة بنقص المناعة Immunodeficiency
- قاعدة بيانات خاصة بالجينات المسؤولة عن المناعة الفطرية .

وفي كل هذه القواعد تكون المعلومات شاملة حول موضوع القاعدة فمثلا الأجسام المضادة تتوفر لها تواليات الجينومات وتواليات البروتينات ، والتركيب الثلاثي للبروتين ، وتتكرر البيانات لكل نوع من أنواع الجزيئات المذكورة في قواعد البيانات مع الإضافات اللازمة لكل نوع من المفردات التي تضمها.

ترتبط قاعدة البيانات وتشارك مع مواقع المعلوماتية الحيوية مثل الارتباط مع المركز الأوروبي EBI – EMBL ويمكن ان تتم العمليات والمعالجات من خلال هذا الموقع ، وكذلك لها ارتباط وثيق بالموقع NCBI . توفر القاعدة برامج عدة لعمليات الاصطفاف مثل BLAST وبمختلف أغراضه وأنواعه لمختلف الجزيئات سواء البروتينات أو الحوامض النووية DNA و RNA . فضلا عن برامج لتحليل مفردات القاعدة مثل التعرف بالجينات مثل IgbLAST وغيرها وكذلك برامج لتبيان العلاقات التطورية . وبرامج لمقارنة تواليات الجينات والاليلات ، وتحديد السلاسل الخفيفة والثقيلة في تراكيب الأجسام المضادة. والتي يمكن ان تمتد الى الأجسام المضادة النانوية أو الصغيرة أو يطلق عليها Single domain antibodies . وتوفر أيضا البرامج الخاصة بالأجسام المضادة إمكانية هندستها على الحاسوب . وكذلك أطالس لبروتينات الإنسان ، وبيانات حول Integrins . وبيانات حول خلايا الأورام . كما ان هنالك جزءا خاص بالمرضات وما يتعلق بها وتداخلاتها مع المضيف . وتوفر القاعدة بيانات حول المستضدات الناقصة ولها قاعدة خاصة بها والتي تدرج فيها معلومات موسعة حول هذه الجزيئات وتعامل الأنظمة الحية معها.

[الأجسام المضادة النانوية : هي أجسام مضادة مكونة من قطعة واحدة Monomeric variable antibody domain وأوزانها الجزيئية 12 – 15 كيلو دالتون وتشابه الأجسام المضادة الكاملة من حيث الارتباط المتخصص المكونة من اثنين من السلاسل البروتينية الثقيلة واثنين من السلاسل الخفيفة وتنتج هذه البروتينات وتستعمل في علاج بعض أمراض القلب التاجية وبعض أمراض السرطانات والزايهمر . ويتكون الجسم المضاد النانوي من سلسلة بيتيدية مكونة من 110 حامض أميني تمثل واحدة من الدومينات المتغيرة V_H للسلسلة الثقيلة من الجسم المضاد أو من IgG العام . هذه القطعة لها الألفة نفسها تجاه المستضد الذي يتفاعل مع الجسم المضاد له ولكنها

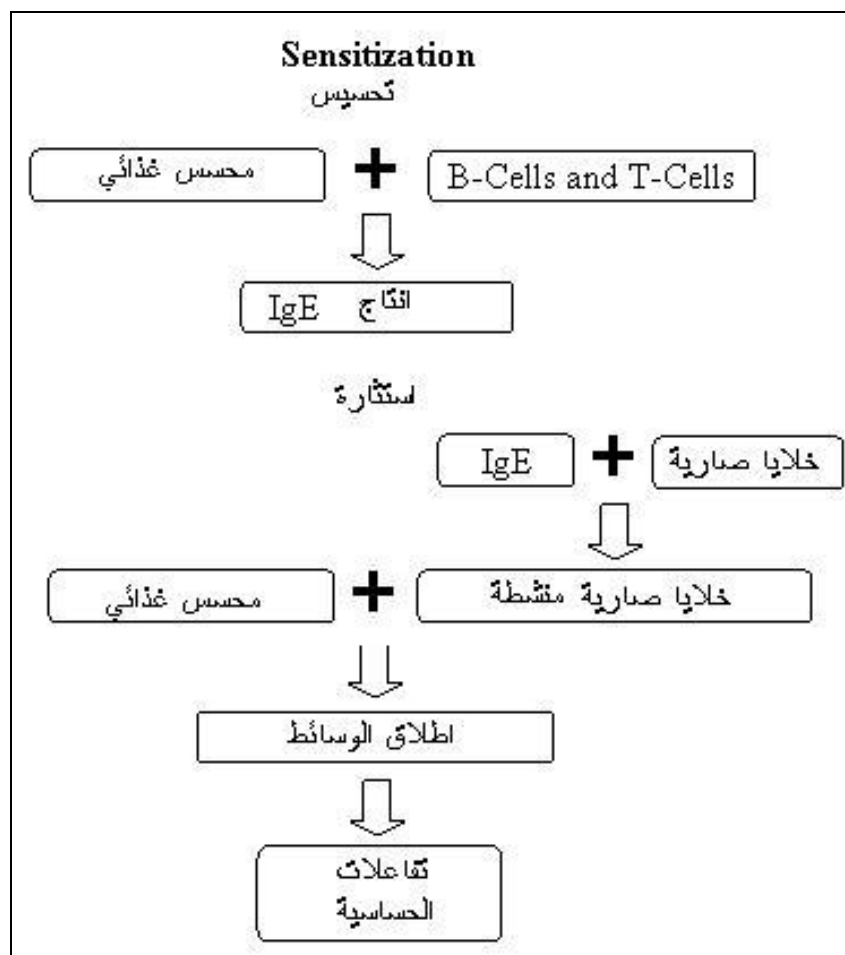
تكون مقاومة أكثر للحرارة والمنظفات وتراكم اليوريا العالية . وتختلف فيما بينها اعتمادا على المصدر المنتجة منه فالمنتجة من الأسماك وبعض الحيوانات الأخرى تكون أقل محبة للدهون وتذوب في الماء . وقلة وزنها الجزيئي يساعد في زيادة نفوذها في الأنسجة ، ولها عمر قصير في البلازما وتفرز مع البول . ونظرا لانعدام منطقة Fc فلا تظهر Complement triggered toxicity كما ان بعض منها ترتبط الى المناطق المخفية من المستضدات التي لا يمكن للأجسام المضادة العادية الارتباط اليها مثل المواقع الفعالة للإنزيمات. تستخدم فمويا ضد *E. coli* وغيرها من إصابات الجهاز الهضمي مثل سرطان القولون . وتستعمل في الشامبو ضد الأمراض الجلدية مثل الإصابة بـ *Malassezia furfur* نظرا لمقاومتها للمنظفات ، ونظرا لصغر حجمها فهي يمكن ان تخترق حواجز الأوعية الدموية ويمكن ان تستعمل في معالجة سرطانات الدماغ]

وتوفر القاعدة برامج لتوضيح التراكيب المجسمة للمفردات الموجودة فيها . البرامج المذكورة آنفا متوفرة لمستلمات الخلايا التائية TRs وكذلك لجزيئات MHC والبيتيدات المرتبطة بها pMHC وكذلك تتوفر البرامج الخاصة بالمحسسات Allergens مثل AllermatchTM لمقارنة التواليف وفي هذا المجال تتوفر المعلومات حول المحسسات في النباتات والحيوانات أو غيرها ، اذ يذكر الاسم العلمي للكائن الحاوي على المحسس أو الاسم العام له باللغة الانكليزية والفرنسية ، فضلا عن ذكر اسم البروتين. وهناك جداول توضح المحسسات الهوائية (الاستنشاقية) مع ذكر التسميات الخاصة بها وفق International Union of Immunological Societies (IUSI) وذكر للتفاعلات المشتركة بين المحسسات ، فضلا عن الأرقام الخاصة بينك بيانات البروتينات وفي حالة الإنزيمات يذكر رقم EC وكذلك الوزن الجزيئي وعدد الحوامض الامينية فيه. ويوجد تصنيف للمحسسات وفق مجاميع الأحياء التي تحويها (كما سيأتي تفصيله لاحقا).

تضم القاعدة بيانات حول الجهاز المناعي وقرعته للعديد من الحيوانات المدجنة Domesticated والاقتصادية مثل الأبقار والجمال والدجاج والكلاب والخيول والأرانب والأغنام وغيرها من الأحياء . ولكل من هذه الحيوانات يذكر جدول بالجينات والاليلات الخاصة بكل من الكلوبولينات المناعية ومستلمات الخلايا التائية . وكذلك تتوفر بيانات حول البروتينات الخاصة بالكلوبولينات المناعية ومستلمات الخلايا التائية. توفر برامج لإيجاد التراكيب الثانوية وكذلك التراكيب الثلاثية . وتحوي القاعدة على المعلومات التصنيفية لكل من الحيوانات وتواليف جينومها ، مع روابط للـ NCBI للتوصل الى البيانات المخزونة فيه والتي تخص الجهاز المناعي . وفي الحقيقة تعد قاعدة IMGT موسوعة كبيرة للحصول على البيانات التي تخص الجوانب المناعية ومتعلقاتها من الجوانب الأخرى مثل الجينومات والعلاقات التطورية .

الحساسية Allergy

الحساسية احد تفاعلات الجهاز المناعي وتكون بشكل سلسلة من التفاعلات تبدأ بتحفيز المحسس لمجموعة من الخلايا منها البائية والتائية والخلايا الصارية تؤدي في النهاية الى إنتاج المواد مثل الهستامين وجزيئات الالتهاب الأخرى من الخلايا الصارية Mast cells كما موضح في الشكل 10.6 :



شكل 10.6 : تفاعلات الحساسية

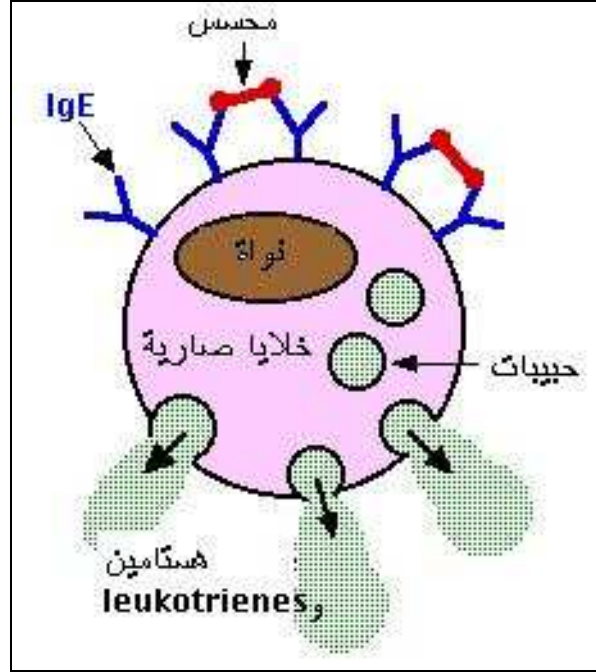
تفاعلات الحساسية ومسبباتها

الحساسية أكثر الأمراض انتشاراً في العالم وخاصة في المجتمعات الصناعية وتصيب حوالي ثلث السكان ويمكن أن تؤدي إلى الأمراض المزمنة . وللحساسية أنواع خمسة . منها المتأخر Delayed hypersensitivity (نوع الذي تتوسط فيه بعض الخلايا اللمفاوية التائية T- cells مثل الخلايا Th2 وبعض حوامل الأجسام المضادة ومنها IgE) وتشارك حوامل هذه الخلايا في أنواع محددة من المحسسات مثل الحساسية للبيض والحليب والجوزيات والنفاح وحبوب الطلع والحليب النباتي Latex . وتحدث بعد حوالي 8 ساعات من التعرض للمحسس وتسبب في الغالب اعتلالات معوية كما في حالة الرضع عند تناول الحليب أو فول الصويا مؤدية إلى Celiac disease ومن مسبباتها أيضاً الاضطرابات المعوية المسببة عن بروتين Gliadin الموجود في كلوتين الحبوب وتصيب حوالي 1:3000 وتصل إلى 1:300 بشكل يعتمد على المناطق الجغرافية . أما الحساسية الأنفية Immediate hypersensitivity وفيها يتوسط

الجسم المضاد IgE . وتكون الحساسية آنية لان الأعراض تظهر بعد دقائق أو ساعات قليلة بعد التعرض للمحسس.

وتحصل الحساسية الأنفية (باشتراك IgE) نتيجة للتعرض الى حبوب اللقاح أو سموم الحشرات وبعض الأغذية وتصل نسبتها في بعض الدول الى 10 - 25 % من الأشخاص . والأغذية المثيرة للحساسية تتمثل ببعض الفواكه والخضر الطازجة وتسبب في العادة متلازمة حساسية الفم Oral allergy syndrome (OAS) التي تمتد الى البلعوم والمرئ ، واغلب المحسسات مدرجة في قوائم Codex الخاصة بعلوم الأغذية ، وبعض المحسسات تكون حساسة لعمليات التسخين والهضم ، اما باقي المحسسات فتكون مقاومة للمعاملات الحرارية والهضم بالتربسين ولعل أهم أنواع الحساسية هي الحساسية الناتجة عن الأغذية (والتي سيأتي ذكرها لاحقا) وهي تفاعلات مناعية تجاه الأغذية أو بعض مكوناتها وأكثرها يتوسطه IgE أي انها تقع ضمن النوع الأول من فرط الحساسية Type I hypersensitivity وتحدث لدى الأشخاص الذين لديهم استعداد وراثي ، وتمتد أعراض الحساسية التي يتوسطها IgE من أعراض خفيفة مثل الحكة والعطاس والتهيج الى حالات مميتة . وتمتد التراكيز المؤثرة الى مستوى المليغرامات ولكن الأطفال يكونون حساسين حتى للتراكيز القليلة ويزداد التحمل بعد البلوغ . ويمكن ان تؤدي الحساسية تجاه المواد الغذائية الى حالات خطيرة مثل التأق Anaphylaxis الذي يحدث في 3 لكل 100,000 شخص ، كما ان البعض تحصل لهم حساسية تجاه الأغذية عند تناولهم أغذية طالها شئ من التلف . وتحصل الحساسية الغذائية في حوالي 3 % من البالغين وترتفع الى 6 - 8 % في الأطفال . و 90 % من الحساسية الغذائية تكون ناتجة عن عدد قليل من البروتينات الخاصة مثل بروتينات فستق الحقل Peanut ، حليب الأبقار ، بيض الدواجن ، الأسماك ، القشريات والحبوب مثل الحنطة وفول الصويا . ويمكن ان تثار الحساسية نتيجة للكربوهيدرات المرتبطة بالبروتينات ، وبعض المكونات الغذائية اذ لم تكن محسنة فانها يمكن ان تعمل بمثابة مستضد ناقص .

وتختلف طرق إثارة الحساسية فبعضها يكون بالتلامس أو الاستنشاق وأخرى بعد ابتلاع المواد المثيرة للحساسية مثل الأغذية . والبروتين المحسس يحث تخليق IgE ، ويحوي على حواتم ترتبط الى IgE والذي يرتبط الى الخلايا القاعدية Basophiles والخلايا الصارية ، ويمكن ان تحدث تفاعلات متداخلة تؤدي الى تفاعلات الحساسية المتمثلة بإطلاق بعض المواد مثل الهستامين . ويحتاج IgE الى موقعين للارتباط وقد حددت المسافة بينها بـ 5 - 24 نانومتر ، ولذلك فان المحسس يكون مؤثرا عندما يكون بتراكيز كافية وان تكون الآلفة للمحسس عالية بحيث يحصل ارتباط IgE بمدة حوالي 100 ثانية . ويكون ارتباط المحسس الى اذرع IgE ليقوم الأخير بالارتباط الى المستلم FcεRI على سطوح الخلايا الصارية وهذا الارتباط يكون السمة المميزة للنوع الأول من أنواع الحساسية ، فاذا كان للمحسس مزدوج متجانس فان توليده للحساسية سيزداد لانه يحتاج الى نوع واحد فقط من الأجسام المضادة (IgE) لإجراء عملية التفاعلات المتداخلة Cross-linking ، كما موضح في الشكل 11.6:



شكل 11.6 : تفاعل IgE مع المحسسات لتوليد الحساسية

آليات حصول الحساسية

لا بد عند التطرق الى موضوع الحساسية من المرور ولو بشكل مبسط حول عمليات حدوث الحساسية . فالحساسية كما ذكر آنفا انها وسيلة دفاعية وتحدد قابلية البروتينات (بشكل رئيس) من حيث كونها حائثة للحساسية أو غير مثيرة للحساسية بواسطة سلسلة من التداخلات المعقدة مع الجهاز المناعي وهذه السلسلة تشمل قبط المحسس ، ومعالجته ثم إطلاقه الوسائط وفي حالة الحساسية المتأخرة اذ عندما يصبح الغذاء بملامسة مع الطبقة المخاطية للأمعاء فانه ينشط خلايا الدم التي تنتج جزيئات الإنذار مؤدية الى إحداث الشرى وهذه تكون طريقة مختصرة في حالة الحساسية الأنفية .

في حالة الحساسية المتأخرة وكما ذكر تتوسط فيها الخلايا وخاصة Th2 التي تتوسط في حالة وعليه فان الاستجابات القوية ، والخلايا التائية تشارك في الحساسية لأنواع محددة من المواد مثل الحساسية للبيض والحليب والجوزيات وحبوب الطلع والتفاح والحليب النباتي Latex (المارة الذكر). وبما ان الجهاز المناعي مختلف بشكل كبير فان الأشخاص يمتلكون عددا كبيرا من تشكيلات HLA ، لذا فالأشخاص يظهرون تفاعلات مختلفة ضد المحسس الواحد . ولذلك كان البحث عن حواتم الخلايا التائية الخاصة بالحساسية عملية متعبة وغير مجدية وعلى الطرف الثاني فان حواتم الخلايا التائية المختلطة قد تقيد في التطبيق السريري وتصميم الأدوية المناعية للحساسية .

طرق فحص الحساسية

تحديد الحساسية يعتمد على

• قابلية البروتين على توليد الحساسية سواء بتوسط IgE أو استجابة خلايا التائية خاصة لمحسسات معينة .

• تحديد التفاعلات المتقاطعة أو المشتركة

• الأعراض السريرية .

وهناك العديد من الفحوص المستعملة لتحديد الحساسية ، منها الفحوص الجلدية المختلفة ويكون الحقن تحت الجلد وملاحظة التفاعلات كما موضح في الشكل 12.6 . أو الحقن تحت غشاء الخلب Intraperitoneal في الفيران أو أي حيوانات ملائمة وتؤخذ النتائج بنظر الاعتبار عندما يكون هناك تجانس بين النتائج الناتجة من طرق إدخال البروتين المختلفة ومقارنة النتائج مع محسسات قوية وأخرى ضعيفة . وهناك طرق أخرى تعتمد على تفاعلات الخلايا للمفاوية مثل هجرة الخلايا للمفاوية استجابة لوجود المواد المحسسة أو تكاثرها أو إطلاقها للمواد .



شكل 12.6 : تفاعلات الحساسية الجلدية

وفضلا عن ذلك فهناك طرق قديمة تعتمد على تحديد محبة البروتين للماء وكذلك نسبة الحوامض الامينية المحبة للماء والكارهة للماء أو الجمع بينهما ومقارنتها مع البروتينات المشكوك فيها فالتشابه يشير الى احتمالية ان البروتينات المفحوصة تكون محسسة . وكذلك الفحص والتحري عن النواحي الوظيفية وان كانت قليلة الأهمية بالنسبة للنواحي التركيبية . ويؤخذ بنظر الاعتبار إضافة السكريات الى البروتينات التي تقلل من تحلل البروتينات خاصة اذا كانت عملية إضافة السكريات ذات تغطية كبيرة مما يطيل بقاء البروتينات في الأمعاء .

وعلى العموم فإن إضافة السكريات قد تكون ذات تأثير مختلف فهي إما أن يكون الارتباط إلى الطرف الأميني N-link وهذه يمكن حدسها بدرجة عالية من الدقة أو يكون الارتباط الآخر O-link والتي يصعب حدسها وتكون غير موثقة . ولذلك في البروتينات السكرية يتم البحث عن حواتم الكلايكان وهذه تكون لها مستوى عالي من التفاعلات المتداخلة .

المعلوماتية الحيوية والحساسية

ازدادت المعلومات حول الحساسية نتيجة للتقدم الذي حصل في تقنيات دراسة البيولوجية الجزيئية للأحماض النووية والبروتينات لذلك احتاجت هذه المعلومات إلى تجميع واستغلال صحيح ، وهذا ما تقدمه المعلوماتية الحيوية لتقوم بإكمال العمل التجريبي المختبري ، وبالمجمل فإن البيانات تحتاج إلى عنقدة وجمع المحسسات في مجاميع معتمدة على العلاقة بين التركيب والوظيفة وكذلك وسائل لتحليل تراكيب حواتم الخلايا البائية والتائية ، فضلا عن الحاجة إلى وسائل دقيقة لتحديد وإيجاد المصفوفات اللازمة لدراسة المكون البروتيني ودراسة القابلية على توليد الحساسية Allergenicity باستعمال وسائل Machine learning tools للبحث والتصنيف وتشكيل العلاقات ، كذلك توفير وسائل للاسترداد وتحليل وتخزين البيانات .

قواعد البيانات

وفقا لما ذكر أعلاه كانت الحاجة ماسة إلى قواعد بيانات مدققة ومشذبة ، وأهم الهيئات المعنية بهذا الموضوع IUIS التي أعدت قوائم رسمية بالبروتينات المثيرة للحساسية . ومن القواعد التي انشأت في هذا المجال قاعدة المحسسات Allergen database و Allergome وغيرها من قواعد البيانات التي بلغ عددها بحدود عام 2003 تسع قواعد . والبروتينات أو المحسسات بشكل عام تحتاج إلى أن تخضع لبعض الفحوص والإجراءات قبل إيداعها في مثل هذه القواعد ومنها أن يكون لها القابلية للارتباط بـ IgE ومقاوم للهضم بالإنزيمات وغيرها . وبعد الإيداع يكون لكل محسس اسم رقم تسجيل خاص به . وقواعد البيانات لا تحوي على كل المحسسات بشكل كامل فالبعض حدد تواليه ولم يدخل إلى قواعد البيانات رسميا كما أن البعض منها تذكر في التهميشات وليس لها مدخل خاص ، ومثل هذا يحدث نتيجة لعدم التناغم والانسجام بين قواعد البيانات وكذلك عدم التدقيق والتشذيب الجيد عند جمعها ، ومن قواعد البروتينات المستعملة قاعدة البروتينات السويسرية و PIR و TrEMBL .

ويمكن إجمال مهام المعلوماتية الأخرى في هذا المجال بالنقاط الآتية :

- وصف متفاوت للصفات الوظيفية والتركيبية للمحسسات وهذه يكون لها روابط للاتصال بقواعد البيانات الرئيسة والمراجع الخاصة
- إجراء عمليات الاصطفاف الشاملة والموضعية باستعمال برامج FASTA أو BLAST للتعرف على مدى التشابه الكلي بين بروتينات الاستعلام والمحسسات الموجودة في قواعد البيانات . وقد حددت الجهات المسؤولة مثل FDA و WHO و Codex الخاصة بالأغذية أن وجود تشابه بنسبة أكبر من 35% في نافذة مكونة من 80 حامض أميني كحد فاصل ، إذ أن 80 حامض أميني يمكن أن تشكل دومين في البروتين ، وبعض الأحيان لا تتفع هذه الطريقة إذ أن التشابه العام قد يكون أقل أهمية من تشابه مناطق محدد ويكون من الأفضل البحث عن قطيقات خاصة في المحسسات واعتمادها في عمليات الاصطفاف ، ولذلك ادخل إجراء آخر

للبحث عن التماثل لـ 6 - 8 حوامض أمينية متجاورة التي لها القابلية للارتباط بالجسم المضاد IgE والأفضل هو الأخير (8) وذلك لأن الحوامض في المحسسات خاصة الغذائية تكون غير معروفة الحجم ، ولكن توصيات الجهات المعنية هو استعمال نافذة بحجم 6 حوامض أمينية متجاورة وذلك لأن التجارب أشارت الى ان حوامض IgE هي صغيرة ولكن أشارت دراسات أخرى ان استعمال نافذة بحجم 6 حوامض أمينية تحدث بنسب لا بأس بها بطريق الصدفة وهذا يؤدي الى محدودية استعمال هذا المؤشر لحداث القابلية على توليد الحساسية وذلك لان بعض حوامض IgE يمكن ان تصل في الحجم الى 4 - 23 حامض أميني وان كانت الأغلبية هي 8 حوامض أمينية ووجود 8 حوامض أمينية متجاورة في بروتينين يكون بعيدا عن الصدفة لذلك تكون له أهمية بايولوجية . وهذه الطرق تؤدي الى الحدس لقابلية توليد الحساسية وكذلك التفاعلات المتداخلة للمحسسات .

• تقوم المعلوماتية الحيوية بتحديد حوامض الخلايا التائية والبائية لتحديد مواصفات المحسس بالمقابل ، اذ ان وجود توالي محدد في الحاتمة يكون ضروريا للارتباط بالجسم المضاد (IgE) وحدث طفرات في هذه القطع يمكن ان يلغي قابلية الارتباط الى IgE وللتعرف على الحوامض المرشحة لهذه المهمة تستخدم طريقة الشبكات العصبية ونماذج ماركوف المخفية التي استعملت في توصيف الخلايا التائية في أمراض المناعة الذاتية والسرطان وأمراض الإصابة وكذلك الحساسية .

• تقوم وسائل المعلوماتية الحيوية بتحديد تراكيب المحسسات مثل التركيب الثانوي والثلاثي الأبعاد من التوالي الأولي للبروتين المحسس ، وكذلك تحديد وظائف الجزيئات الكلية أو دوميناتها الرئيسية والناجمة عن وجود 20 - 30 % من التماثل (وليس التشابه) . وتحديد هذه المستويات من التركيب يؤثر في قابلية توليد الحساسية من خلال التأثير في مواقع ارتباط IgE .

كل الانجازات التي تحدها المعلوماتية الحيوية في مجال الحساسية مثل القابلية على إثارة الحساسية (المفترضة) وتحديد التفاعلات المتداخلة يحتاج الى اختبارات واقعية مثل اختبارات الغفل المزدوج Double-Blind Placebo-Controlled Food Challenge (DBPCFC) وكذلك استعمال التفاعلات الجلدية مثل Skin prick test وغيرها من الفحوص .

المعلوماتية الحيوية وعلاج الحساسية

توفر المعلوماتية الحيوية إمكانية المساعدة في إيجاد العلاجات . فالعلاجات المستعملة في العموم هي تجنب الأغذية المثيرة للحساسية وهذه تكون لها جوانب سلبية تؤدي الى سوء التغذية ، والعلاجات الأخرى هو إعطاء مضادات الهستامين أو مركبات الكورتيزون Corticosteroids أو استعمال حقن الابينفيرين Epinephrine ، وكذلك إمكانية استعمال لقاحات DNA لحث أو تنشيط الاستجابة المناعية لذلك تهدف المعلوماتية في المساعدة في العلاج المناعي وذلك من :

• إعادة ترتيب حوامض IgE

• عملية إزالة التحسس Desensitization وذلك بإجراء تجارب على الحاسوب ويكون ذلك باستعمال مضادات للـ IgE وهي بروتينات أو ببتيدات مهندسة وراثيا . وعلى ارض الواقع تتوفر برامج تساعد في تجنب المحسسات كما في استعمال Allergen avoidance database tools وفيها يتم الاستعلام عن مادة معينة والنتيجة تظهر قوائم من المنتجات التي لا تحوي على المحسسات مثل منتجات العناية بالبشرة خاصة عند استعمال البرنامج الخاص بالتلامس Contact allergen avoidance program ، اذ يطبق البرنامج على المحسسات الموجودة في قاعدة البيانات التي يستند عليها لتجنب المواد المثيرة للحساسية ، وبطبيعة الحال يتم تطبيق البرنامج بعد عمل الفحص الجلدي للشخص وإيجاد المحسسات التي تؤذيه ، وكذلك تستعمل في تحديد محسسات الطحين المستعمل في المعجنات ومثل هذه تستند الى قياس IgE الكلي و IgE الخاص بالمحسسات .

الأغذية المهندسة وراثيا

افرز تطور الحضارة الصناعية وغيرها من الجوانب العديد من الجوانب السلبية للحياة . واحد هذه التطورات هو هندسة الأغذية وراثيا الذي فرض سواء رضي أو سخط المستهلكون لغرض سد النقص ولإغراض أخرى . وفي جميع الأحوال كانت نتائج هندسة الأغذية هي سلبية وان ارتضاها البعض على مضض ، وقد لوحظ في التسعينات ان إطلاق فول الصويا المهندسة وراثيا قد أدى الى ارتفاع الحساسية للصويا بنسبة 50% في بريطانيا لانها تحوي على بروتينات مشابهة للمحسسات . وفي الحالات الطبيعية تكون الحساسية لفستق الحقل هي الأوسع انتشارا وأكثرها ضراوة (90 % من الحساسية الغذائية) ثم يليه فول الصويا ثم الحليب ثم البيض والذي يؤدي تناولها الى ظهور الأعراض في دقائق وتؤدي الى الربو والشرى Ucaria واضطرابات الجهاز الهضمي وتؤدي بعض الأحيان الى الصدمة Anaphylactic shock ومثل هذه الأعراض تظهر من لدغات الحشرات مثل النحل والزنابير . وهذه تتضخم بعد عمليات الهندسة الوراثية مما يجعل تقبلها صعبا ، وعلى العموم فان إثارة الحساسية بواسطة الأغذية سواء كانت طبيعية أو مهندسة تكون اقل من المحسسات الاستنشاقية مثل الغبار ومخلفات الحيوانات .

التحوير الوراثي يؤدي الى اضطرابات كثيرة أهمها الحساسية ويكون ذلك بإدخال المحسسات الى الجسم أو إدخال بروتينات غريبة على الجسم ليقوم الجسم بتفعيل الدفاعات المناعية ضدها ، ومن جهة ثانية فان الهندسة الوراثية يمكن ان تغير من مستوى أو طبيعة المحسسات الموجودة أصلا في النباتات

مواصفات المحسسات الغذائية وتأثيراتها

المحسسات الغذائية يكون تأثيرها بتوسط IgE أي انها من النوع الأول من الحساسية واغلبها يكون فيها اثنين أو أكثر من مواقع الارتباط للجسم المضاد وهناك حوالي 1 : 100,000 من المكونات الطبيعية للأغذية تثير الحساسية ، ولذلك فعند وجود حاتمة واحدة لا تجعل البروتين مثيرا للحساسية ، ولكن في الغالب يتعامل الجسم معها على انها مواد غريبة ومثيرة للتفاعلات المناعية والمحسسات الغذائية وكذلك الهوائية Aeroallergens تكون بمثابة بروتينات حامضية ونقطة التعادل الكهربائية pI حوالي 5.7 في أكثر من 80 % من البروتينات المحسنة ، اما

غير المحسنة 20 % منها حامضية ولكن pI لها 5.8 . والمستضدات الببتيدية لبعض البروتينات يمكن ان تكون لها علاقة مع التراكيب الدقيقة للبروتينات مثل β -turn وكذلك صفة محبة الماء وكذلك حركة البروتينات ، ويكون للحساسية علاقة بصفات أخرى مثل الارتباط الى الدهون ، والشحنات الأيونية والتحويلات بعد الترجمة ولذلك فان تأثير الرقم الهيدروجيني pH وشحنات الايونات على سطح البروتينات في المحسسات الحامضية أكثر مقارنة بغيرها من المحسسات ويمكن ان تكون محدثات أو مؤشرات تستعمل في حدس الحساسية وتستعمل في دراسات الحاسوب ، فعند زيادة الشحنات السالبة على سطوح البروتينات تزداد صفاتها المستضدية وذلك لان الشحنات السالبة للكهربائية المستقرة لها تأثير مباشر في الخلايا الطلائية وكذلك قنوات الايونات الموجودة في الخلايا الطلائية للجهاز التنفسي في الإنسان ، أي ان الصفات الفيزيائية للبروتينات المتكاملة اثر في تحديد القابلية على توليد الحساسية وهذه يجب ان تؤخذ بنظر الاعتبار عند إجراء دراسات الحاسوب ، وعليه فان التغيرات في تركيب المستضد وتوالياته من الحوامض الامينية الأثر الكبير في تغير صفة إثارة الحساسية فمثلا Tropomyosin وهو من محسسات الروبيان Shrimp وجد انه عند تغيير أو استبدال حامض أميني واحد يمكن ان يفقد البروتين قابليته على توليد الحساسية نظرا لعدم قابليته للارتباط بـ IgE . فالمحسسات تكون أكثر سالبية حيث يكون جهد الكهربائية المستقرة (EP) Electrostatic potential لها أكثر سالبية مقارنة بالبروتينات غير المثيرة للحساسية وهذا يتفق مع ان الخلايا التائية تميز مولدات المناعة بالاعتماد على شحنتها. وتؤثر الأغذية المحسنة وكذلك الأغذية المهندسة وراثيا في إثارة الحكة والتقيؤ والإسهال وربما الموت بالصدمة وعلى المدى الأبعد يمكن ان تؤثر في أشكال النطف وتطور الأجنة المتعثر الذي يؤدي الى الموت . وقد قامت IUIS بوضع قوائم خاصة بالمحسسات ومواصفاتها والمؤسسة لها طريقة خاصة بالتسمية ، فأول ثلاث حروف تعرف الجنس والحرف الأخير هو للنوع ، اما الأرقام فتشير الى ترتيب اكتشاف المحسس كما في Bet v1 المحسس الموجود في نبات البتولا *Betula verrucosa* المكتشف أولا ، اما نظائر المحسس فلها أرقام أخرى : الأول والثاني للمحسس النظير ، والثالث والرابع هي متغايرات النظير كما في الآتي Bet v 1.0101 . ولذا يفضل المستهلكون الأغذية العضوية التي لا تحوي على مشتقات الأغذية المهندسة وراثيا والتي يمكن ان تدخل في مفردات غذائية كثيرا للتخلص من الحساسية والأضرار الأخرى .

المحاصيل المهندسة

أكثر النباتات والمحاصيل التي تعرضت للتغييرات الوراثية هي فول الصويا والذرة والبنور الزيتية مثل السلجم وبنور القطن ، وفي كثير من الأحيان تطرح للأسواق دون تعليم وهي المسببة لأعلى مشاركات في الحساسية كما ذكر أعلاه . وقد أشارت التقارير ان تناول الصويا المهندسة يؤدي الى متلازمة القولون المتهيج ، ومشاكل في الهضم وإعياء مزمن وصداع والشعور بالنعاس ، فضلا عن تفاعلات جلدية مثل ظهور البثور والاكزيما حتى عند إدخال مشتقات النباتات المهندسة وراثيا وليس استعمالها المباشر كما في دخولها في بعض خلطات الأغذية للأطفال يؤدي الى إثارة تفاعلات حساسية لديهم ، وقد وجد ان البروتينات المدخلة الى المحاصيل لها تشابه مع المحسسات وكذلك قد تكون لها

تفاعلات متداخلة والقابلية على إثارة المناعة خاصة تخليق IgE . والهندسة الوراثية يمكن ان تحول المحسسات المخفية Hidden allergens وقد وجد ان الصويا المهندسة تحوي على حواتم للـ IgE غير موجودة في الصويا الطبيعية .

والأغذية المهندسة وراثيا لها تأثيرات أخرى غير تلك المتعلقة بحواتم البروتينات ، فعند نقل DNA الذي يخلق المحسس يمكن ان يلتقط من قبل بعض البكتيريا القاطنة في الأمعاء وتستمر في إنتاجه وتستمر الأعراض حتى عند توقف تناول الأغذية المهندسة لان المحسسات سيستمر إنتاجها في القناة الهضمية ، وفي الوقت ذاته فان المحسسات المنتجة بهذه الطريقة قد تكون أكثر ضراوة نتيجة لعدم طوبها الصحيح أو عدم إجراء التحويلات الأصلية عليه لان البكتيريا لا تستطيع إجراء بعض التحويلات التي تجري في خلايا النبات .

فضلا عن ان نواتج بعض الجينات يمكن ان تؤدي الى تخليق مساعدات المحسسات Allergenic adjuvant التي تؤثر في مكونات الغذاء التي كانت قليلة القابلية على إثارة الحساسية .

كما ان التغيرات التي تجري على DNA النباتات يمكن ان تؤدي الى إنتاج محسسات جديدة غير معروفة ، وكذلك فان RNA المنتسخ من المناطق المجاورة للإقحام يمكن ان تنتسخ وتولد محسسات غير معروفة .

وهندسة النباتات لمقاومة مبيدات الأدغال مثل Glyphosate التي تتراكم في النباتات وعند تناولها يمكن ان تعرقل الهضم وتؤدي الى الحساسية وكذلك فالأمر لا يتعلق بالمبيد الذي يمكن ان يتراكم وانما سيتعدى الى نواتج تكسيره والتي يمكن ان تثير الاضطرابات ، فعند إطعام الفيران لفول الصويا المهندس وراثيا أدى ذلك الى تقليل إنزيمات البنكرياس وهذا يؤدي الى إطالة بقاء المواد المحسسة لمدة أطول في القناة الهضمية ، وعند عدم وجود إنزيمات تتعامل مع البروتين المدخل أو المبيد فانها ستبقى لمدة طويلة وبالتالي تزداد المشاكل ، كما ان انخفاض قابلية هضم البروتينات نتيجة لتأثير المبيدات فانه يمكن ان يسبب الحساسية لبروتينات أخرى . فعند إدخال مثبطات إنزيم الاميلز-ألfa في البزاليا لمنع الديدان من التغذية عليها يمكن ان يؤدي الى فشل تحليل النشا وتتداخل مع عمليات الهضم .

ومن المنتجات والمواد المنتجة بالتقنيات الحيوية قد لا تكون مفردات غذائية مثل الأغذية الطبية و مواد التجميل وغيرها التي تحتاج الى الفحص لضمان أمان استعمالها قبل الوصول الى المستهلك وانها لا تولد الحساسية .

كما ان هناك مسألة استعمال البروتينات المستعارة Surrogate proteins المنتجة في البكتيريا التي لا تستطيع إضافة السكريات مما يؤدي الى تغير مستضدية البروتينات وهذه تكون غير مقبولة في الفحوص التي تجري على الأغذية المحورة وراثيا . ومن الأمثلة على المحاصيل المهندسة الصويا التي صممت لتحوي على بروتين الجوز البرازيلي الغني بالميتايونين لتحسين نوعية بروتيناتها من الأحماض الامينية الحاوية على الكبريت التي تتصف النباتات بصورة عامة لافتقارها لياه ولكنها سحبت من الأسواق نظرا لتوليدها الحساسية ، ويتفاهم الأذى نظرا لان الطبخ والمعاملات الحرارية يمكن ان تزيد من أذى المواد المهندسة وراثيا .

ومن الجدير بالذكر ان عمليات الهندسة التي تستهدف الجزء المأكول من النبات مثل البذور في حالة الذرة وفول الصويا وغيرها يمكن ان يمتد أذاها الى أجزاء أخرى من النبات ، فأجزاء

النبات الأخرى المشتقة من محاصيل مهندسة وراثيا تسبب الاضطرابات عند الحيوانات عند استعمالها كعلف فقد وجد ان البعض منها يموت في مدة 5 - 7 أيام أي انها تولد السمية كما أوضحت الدراسات التشريحية للأغنام النافقة . فضلا عن ذلك فان حبوب الطلع للنباتات المهندسة وراثيا مثل الذرة تسبب الربو وصعوبة في التنفس وأعراض الحساسية الأخرى عند استنشاقها .

ولكن في مجال الهندسة الوراثية والحساسية يمكن ان تكون هناك جوانب ايجابية فبعض الأغذية المنتجة في هذا المجال يمكن إزالة بعض الجينات أو قطع منها المتعلقة بتوليد الحساسية أي يكون المنتج خاليا من المحسسات كما في حالة فستق الحقل والروبين المنتج في أمريكا وفي اليابان تزال بعض المحسسات من الرز وبالتالي يمكن للمتحمسين تناول هذه الأغذية .

أسباب هندسة المحاصيل وراثيا

هناك عدة أسباب دعت الى هندسة النباتات الإستراتيجية وذلك لتحسين صفاتها مثل :

- مقاومة الأمراض بأنواعها سواء الإصابات الفطرية والبكتيرية أو الإصابة بالديدان
- مقاومة الحشرات
- مقاومة المبيدات التي تستعمل لإبادة الأدغال Herbal pesticides
- تحسين النوعية للمحصول مثل تحسين القيمة الغذائية كما في الرز الذهبي أو تحسين الطعم أو النسجة أو زيادة مدة الصلاحية Shelf life .

وبعض هذه المنتجات المهندسة وصلت الى مرحلة الاستهلاك البشري ولكن جاءت المشاكل معها . ومن أكثر الأغراض التي تهندس النباتات لأجلها هو مقاومة الحشرات والتي تتمثل باستعمال الجينات المسؤولة عن قتل الحشرات بواسطة سموم Cry الموجودة في البكتريا *Bacillus thuringiensis* واستعملت في الذرة وسوق منها 1507 maize و Mon 863 وهي تحوي جينات Bt toxins مختلفة وقد وجد ان كل من السموم Mon 810 و Bt 176 وهي تحوي جينات Bt toxins مختلفة وقد وجد ان كل من السموم Cry2Ab و Cry3Aa و Cry3Bb موجودة في المحاصيل المسوقة . فالسم Cry1Ab يضيف المقاومة للإصابة بالحشرات حرشفية الأجنحة Lepidoptera مثل الفراشات و Cry3Bb1 يضيف المقاومة ضد حشرات غمدية الأجنحة Coleoptera مثل الخنافس وهو المعروف بـ Mon 863 وتشير الدراسات الى ان الأخير يكون محسسا ضعيفا ، وقد أشارت التجارب الى ان إطعام هذا المنتج يثير الحساسية كما اتضح من استعمال الفحوص المناعية مثل زيادة الخلايا القاعدية واللمفاوية وكذلك زيادة العدد الكلي لخلايا الدم البيض .

والسموم المذكورة مسجل وجودها في كل أنواع الذرة المهندسة وراثيا ووجد ان هناك بروتينات الضروب المهندسة تشابه حوالي 17 من المحسسات المعروفة والثابتة جدا لعمليات الهضم ، فضلا عن ذلك فان هذه السموم تثير الحساسية عند استعمالها بشكل مباشر (أي بدون إدخالها للمحاصيل عند هندستها)، والتركيز في حالة الرش يكون اقل الآلاف المرات مما في حالة إدخالها الى صميم النباتات .

في مجال تحسين النوعية الغذائية للمحاصيل فقد ذكر أنفا دس جينات البروتينات الغنية بالميثايونين من الجوز البرازيلي المعروف بصفته المثيرة للحساسية في الصويا التي ولدت الحساسية مما أدى الى عدم تسويقها .

وعند هندسة المحاصيل لتكون مقاومة لمبيدات الأدغال ودس الجين المسئول عن Glyphosate oxidoreductase الذي يضيف المقاومة للمبيد ، وجد ان بروتين الإنزيم يحوي على ستة حوامض أمينية متجاورة هي LAEEAP مشابهة للمحسس الموجودة في Tropomyosin المنتشرة في الأغذية البحرية وطفيليات الأسماك وتكون جزءاً من امتداد لتسع حوامض أمينية متجاورة موجودة في بروتينات الروبيان التي ترتبط الى IgE في الأشخاص الحساسين للبروتينات . وهناك قطيقات كثيرة من الحوامض الامينية التي اكتشفت في الأغذية المهندسة وراثيا وذات صلة وثيقة بالمحسسات .

فحوص الحساسية للأغذية المهندسة وراثيا

المحاصيل الزراعية المهندسة وراثيا يجب ان تخضع لعدة فحوص قبل طرحها للأسواق وذلك للمتطلبات الأخلاقية والاجتماعية ووجود التحديات الكبيرة سواء في مجال البيئة أو صحة الإنسان أو صحة الحيوان . اذ وجد ان بعض البروتينات المنتجة من أحياء مهندسة وراثيا تكون مشابهة لمحسسات معروفة . لذلك تم التأكيد على فحوص إثارة الحساسية نظرا لأهمية هذه الناحية بالنسبة للمستهلك . والحساسية بداية هي وسيلة دفاعية للجسم ولها مظاهر مختلفة يمكن ان تعتمد في تصميم الفحوص اللازمة . وقد وضعت الجهات المختصة FAO/WHO شروط للفحوص الواجب إجراؤها اعتمادا على أصل البروتينات باعتبارها الأساس المثير للحساسية بامتلاكها لحواتم المحسسات وتقران الفحوص للمحسسات مع محسسات موثقة وكذلك مع بروتينات لا تثير الحساسية . فهناك أسس للفحوص الآتية من مصادر مختلفة مثل نباتات ذوات الفلقة أو الفلقتين ، والفطريات وكذلك المصادر الحيوانية مثل الفكريات واللافكريات ومصادر أخرى مثل البكتريا. وتعد الأسس الموضوعة من قبل الجهات المختصة (FAO/WHO) عام 2000 هي الأساس للرجوع إليها وان كانت تشهد تحديثات وتغيرات وفق الإصدارات والإرشادات المختلفة التي تعدل وفق حالة الغذاء وتشير التعليمات الى وجوب تحديث قواعد البيانات الخاصة بالمحسسات باستمرار .

مواصفات المحسسات المعتمدة في الفحوص

- فحوص الحساسية مختلفة في آليتها وطريقة عملها وتعتمد على طبيعة المحسسات ومنها
- المحسسات تكون مقاومة للهضم بالببسين والحوامض الخفيفة، وفي هذه الحالة يجب الأخذ بنظر الاعتبار احتواء الأغذية على مثبطات البروتيازات ، لذلك تستعمل البروتينات النقية لان مقاومة المحسسات للهضم والظروف الحامضية يجعلها تمر بدون تغير في محيط المعدة لتصل الى الأمعاء وتمارس فعاليتها . في هذا الفحص يوضع البروتين في محيط حامض حاوي على الإنزيمات ثم يفحص ثبوته ، لذلك تجرى الفحوص في ظروف مشابهة لظروف المعدة . وهذا الفحص لا تصلح له معظم برامج المعلوماتية الحيوية الخاصة بتحديد مناطق فلق البروتينات بالإنزيمات وانما يجب ان يجرى الفحص مختبريا .
- فحص وقياس نسب الحوامض الامينية المحبة للماء والأخرى الكارهة له لتحديد المستضدية وذلك بقياس محبة الماء ١ كارهة الماء لامتداد معين من توالي الببتيدات . فزيادة الحوامض الامينية المحبة للماء تكون لها صفات مستضدية عالية، وكذلك تكون الحوامض الامينية المنضمة للمجموعة الحامضية أكثر إثارة لتفاعلات الحساسية وهذه يمكن ان تجرى على الحاسوب باستعمال الجداول الخاصة بكراهية ومحبة الماء .

• قابلية الارتباط الى المصول وفيها يقاس مدى الارتباط الى IgE سواء داخل أو خارج الجسم الحي لمصول على الأقل لخمس من الناس المتحسسين . وكذلك توجد فحوص أخرى تجرى خارج الجسم الحي تشمل تحديد مدى تكاثر الخلايا اللمفاوية وهجرتها وانطلاق المواد منها وغيرها .

• الفحوص السريرية ، وفيها تستعمل الموديلاات الحيوانية ولكنها قليلة نظرا لقلّة الحيوانات الملائمة ما عدا الجرذ النرويجي Brown Norway rat . وكذلك أجراء فحص التحسيس الفموي Oral sensitization والحقيقة انه لا يعول على مثل هذه الفحوص التجريبية . والتوجهات الحديثة لا تدعم هذه الفحوص كونها قادرة على إثارة الحساسية . وذلك لان الحيوانات التي تأكل النباتات المحورة تذبح بعد مدة قصيرة ولا تترك لمدة طويلة لملاحظة التأثيرات المتأخرة .

• الاعتماد على عمليات الاصطفاف والتشابه التي تتم بوسائل المعلوماتية الحيوية مع محسسات موثقة . وهو ما يتناوله الجزء التالي من المناقشة .

تداخل المعلوماتية الحيوية في تحديد الحساسية للأغذية المهندسة وراثيا

نظرا لتوسع وسائل المعلوماتية الحيوية وازدياد البيانات الخاصة بالحساسية وبشكل خاص للأغذية المهندسة وراثيا من جهة ومن جهة ثانية كان لابد من الموافقة والتصريح بالتسويق للمواد المهندسة وراثيا لتحديد صلاحيتها للاستهلاك وقد وضعت شروط صارمة من قبل Codex حول الأغذية قبل التصريح بتسويقها ، لذلك كان لابد من الربط بينها ، وقد تبنت عمليات الربط المؤسسات المسؤولة عن ذلك مثل FAO و WHO لإنشاء شبكة مشتركة لتوحيد عمل المختبرات وتسهيل الممارسات السريرية لتحديد الحساسية وخاصة للأغذية المهندسة وراثيا . وكانت جهود هذه المؤسسات تتوزع على نواحي شتى وتهتم بالبروتينات مثل بروتينات البذور ولكن باقي أجزاء النباتات يمكن ان تكون مصادرا "خطرة كما ذكر أنفا وعلى العموم فان الجوانب التي اهتمت بها الجهات المعنية هي :

• إنشاء قواعد بيانات خاصة بالمحسسات وتحديد تواليات الحوامض الامينية فيها بصيغ FASTA من البروتينات الناضجة دون ببتيديات الإشارة . واهم قواعد البيانات الخاصة بهذا المجال Allermatch التي تستخدم من قبل FAO/WHO وكذلك Allergen database ، وتستقي هذه القواعد بياناتها من قواعد البروتينات العامة مثل قاعدة البروتينات السويسرية و WHO-IUIS وكما ذكر أنفا فان البروتينات في هذه القواعد هي البروتينات الناضجة بدون تواليات الإشارة ، وقد طورت وسائل خاصة لاسترداد البيانات منها ، فضلا عن كونها قواعد لا توجد فيها المكررات والإضافات اي انها قواعد مشذبة . وعلى العموم فان قواعد البيانات العامة والخاصة قد لا تلبي متطلبات المعلوماتية الحيوية في هذا المجال ، لذا يتطلب الأمر إنشاء قواعد مرجعية خاصة Reference allergen databases تكون جهد الإمكان حاوية على تواليات للمحسسات المعروفة والنادرة . وهذا يتفق مع تنامي الشعور بوجوب تحري الأمان بشكل معقد وكبير وهذا يحتاج الى بنى تحتية جيدة كما في إنشاء القواعد المذكورة أعلاه فضلا عن استمرار عمليات التحديث والإدماة والإدارة الجيدة واستخدام الوسائل الفاعلة للاسترداد والتحليل .

• التوصية بإجراء عمليات الاصطفاف للبروتينات الجديدة مع البروتينات الموجودة في قواعد البيانات لتحديد القطيقات أو الاصطفاف الشامل على ان تكون قيم E value واطئة وتبدأ من 0.001 فما دون .

• عنقدة البيانات الناتجة للبروتينات بعد استعمال مجموعة برامج Clustal مثل ClustalW

• التوصية بدراسة كل حالة على حدة وعدم الاعتماد على المراجع وفحوص النباتات الأخرى وعمليات الاصطفاف الموصى بها تكون :

❖ اصطفاف كلي لبروتينات الاستعلام مع بروتينات قواعد البيانات التي تشير الى الأخذ بنظر الاعتبار درجة التشابه 50 % من طول توالي الى البروتين الكامل ، وذلك للاحتمالية الكبيرة لحدوث تفاعلات الحساسية أو على الأقل حصول التفاعلات المتداخلة .

❖ صف البروتينات باستعمال نافذة بطول 80 حامض أميني وذلك لان هذه تمثل حجم معظم الدومينات البروتينية في البروتينات الرباعية التركيب واعتماد درجة التشابه اكبر من 35 % كمؤشر خطر .

❖ استعمال صف وإيجاد توالي من الحوامض الامينية بطول 6 - 8 حوامض أمينية متجاورة على ان تقاس قيمة محبة الماء لكل حامض أميني Hydrophilicity value التي تكون ثابتة لبعض الحوامض الامينية والتي تمثل الحواتم التي يمكن ان ترتبط الى IgE ، وتحسب قيم محبة الماء على مخططات محبة الماء Hydrophilicity plot وتؤخذ أعلى نقطة على انها تمثل موقع الارتباط مع الجسم المضاد وفي حالة التطابق تكون المخرجات موجبة لامتداد التوالي . وفي البداية اقترح ان وجود قطيقات بطول 6 حامض أميني تكون منذرة بالخطر ولكن اقترح زيادة العدد الى 8 حوامض أمينية لتقليل النتائج الموجبة الكاذبة ، والتأكيد على عدم استعمال تواليات 4 أو 5 من الحوامض الامينية لانها ستزيد في النتائج الموجبة الكاذبة . وعلى العموم فان وجود اثنين من هذه القطيقات للبروتين تكون أساسية لجعله مثبثا للحساسية ، اذ ان وجود واحدة لا يثير الحساسية نظرا لان IgE يحتاج الى اثنين من الحواتم كي ينشط ويرتبط الى الخلايا الصارية وتحفيزها وإطلاق المواد مثل الهستامين (كما سبق ذكره) .

الاصطفاف الشامل: وتبدأ عمليات الاصطفاف بتحضير مصفوفات من 80 حامض أميني من البروتين الذي يتم التعبير عنه في الأغذية المحورة . وتتم عمليات الاصطفاف باستعمال برامج BLAST أو FASTA وتوصي الجهات المختصة باستعمال برنامج FASTA مع استعمال مؤشرات البرنامج الأصلية Default values فيما يخص جزاءات الفجوات (الجزء 10- لفتح الفجوة و 2- لامتداد الفجوة) واستعمال المصفوفة BLOSUM 50 و $Ktup = 2$. والبرنامج يبدأ بصف الحوامض الامينية من النهاية الامينية الى النهاية الكربوكسيلية للبروتينات ويمكن ان تستعمل للتواليات القصيرة . والحقيقة ان إجراء الاصطفاف الشامل هو ليس جزء من توصيات FAO/WHO ولكنها تعد وسيلة إضافية مفيدة . والتوصيات تشير الى وجوب استعمال برنامج قليل الأخطاء مثل FASTA ولكن مع هذا يلاحظ بعض الأخطاء منها ان بروتين الاستعلام يمكن ان يشخص على انه محسس قوي ولكنه في الحقيقة ليس كذلك وهذا يقع ضمن نتائج الموجبة الكاذبة وهذا ما يحصل مع التواليات القصيرة ، والخطأ الآخر هو إشارة

النتائج الى ان البروتين غير محسس ولكنه في الحقيقة غير ذلك والسبب في ذلك يعود الى ان البروتينات المحسنة لم تدخل كلها في قواعد البيانات وان بروتينات الاستعلام الجديدة لا يوجد ما يشابهها في القواعد ولتلافي هذه الحالة أدخلت Orphan entries في قاعدة Allermatch allergen database وهي ليست لها علاقة بالمداخل الأخرى في القاعدة وانما تكون مستقلة وذلك لتقليل معدل النتائج السالبة الكاذبة . ان ظهور النتائج الأخيرة يمكن ان يأتي من قصر تواليات بروتينات الاستعلام ، فمثلا توالي بطول 28 حامض أميني يمكن ان يعطي تشابه يصل الى 100 % وهذا التشابه يعطي تشابه اقل من 35 % عند تحويله الى نافذة بطول 80 حامض أميني ، وعليه فان حجم القواعد يمكن ان تكون من العوامل المؤثرة في النتائج التي يتم الحصول عليها ، فصغر حجم القاعدة يمكن ان يزيد النتائج السالبة الكاذبة ويحصل العكس عند كبر حجم القاعدة .

والاختلافات تعود الى ان القواعد الكبيرة يمكن ان تحوي على أعداد كبيرة من مناظرات المحسسات Isoallergens (وهي المحسسات التي توجد فيها اختلافات صغيرة أو بسيطة من تواليات الحوامض الامينية) . كما ان النتائج السالبة الكاذبة قد تكون ناتجة عن وجود ببتيديات الإشارة التي قد تكون موجودة في بروتينات القاعدة نتيجة لضعف عمليات التشذيب والتدقيق ، ولذلك فان النتائج الموجبة الكاذبة يمكن ان تتناسب أيضا مع حجم قاعدة البيانات نتيجة لعمليات الاصطاف العشوائية . وعليه كان من الأفضل استعمال مؤشر اصطاف 80 حامض أميني مع اصطاف القطيفات 6 - 8 حامض أميني سوية خاصة مع القواعد الكبيرة مثل Allermatch وتوثيق النتائج مع المراجع العلمية وتدعيمها بالفحوص التأكيدية التي سيأتي ذكرها لاحقا .

الاصطاف باستعمال نافذة بطول 80 حامض أميني

وتكون هذه الحالة مكملة ومؤكدة لعمليات الاصطاف الشامل التي تؤكد عليها المؤسسات المعنية ، وقد كان الأساس اعتماد الحد الفاصل 35 % تشابه ولكنها بعد التعديل أصبحت أكثر من 35 %

نظرة عامة للطرق المستعملة

فضلا عن دراسة التشابه في تواليات الحوامض الامينية فهناك جوانب أخرى يجب ان تؤخذ بنظر الاعتبار اذ ان حدس المواقع بالحاسوب قد يكون غير متكامل ، وذلك لان حدس المحسسات يمكن ان يعتمد على الخصائص التركيبية ، فصفة الارتباط الى الدهون قد يكون لها ارتباط بحث الحساسية وهذه يمكن ان تستغل في التعرف على المحسسات ، وكذلك ارتباط البروتينات بالكربوهيدرات مثل الكلايكان الذي يمثل مواقع مهمة للارتباط بـ IgE في البروتينات السكرية المحسنة .

وعليه فان النتائج التي تظهر وتشير الى ان البروتين قد يثير الحساسية أو لا تكفي وانما يجب ان تأخذ بنظر الاعتبار معها مواصفات أخرى واتخاذ خطوات أخرى ، منها حجم جزيئات البروتين وثبوته وقابلية ذوبانه ونقطة التعادل الكهربائي ، كل هذه الصفات تدرس للبروتين المهندس ويقارن مع المحسسات المعروفة . وإضافة الى ذلك النظر الى التوزيع المجسم للحواتم وفيما اذا كانت مستقيمة ، ومنظمة Codex الخاصة بالأغذية تفضل استعمال التوجيهين

، وعليه فإن استعمال 80 حامض أميني في الاصطفافات يمكن ان يغطي التوجيهين أي وجود Linear epitopes و 3D-Structures .

إضافة الى استعمال FASTA لصف التواليات الشاملة والبحث عن التشابه بنافذة 80 حامض أميني ، فإن البحث عن تواليات الحوامض الامينية المتجاورة يمكن ان يتم باستعمال برنامج MEME لتحديد القطيقات على ان تكون E value بحدود 10^{-8} وذلك لزيادة الدقة وتكون كدليل للحساسية كما يمكن استعمال طريقة Wavelet transform لاستخلاص القطيقات من المحسس لحسد الحساسية . وهناك برامج أخرى تستعمل التراكيب الثانوية مثل الحلزونات وصفائح بيتا والاستدارات في تراكيب البروتين ولكن الأفضل ربط البرامج مع بعضها .

الفحوص التأكيدية

ان النتائج التي يتم الحصول عليها من وسائل المعلوماتية الحيوية لا يمكن ان تعتمد لوحدها فيما اذا كان البروتين يثير الحساسية أم لا ، وذلك لان ليس كل أنواع الحساسية يتوسطها IgE الذي يشكل حدسه جزءا مهما من توصيات الجهات المختصة ، ولغيرها من الأسباب فان نتائج المعلوماتية الحيوية يمكن ان تحدد خارطة طريق أو توجه يحتاج الى فحوص للتأكد ومنها :

- إجراء فحص الهضم بالإنزيمات في أوساط حامضية
 - قياس ارتباط البروتين (الجديد) الى IgE من مصول أشخاص معروف ان لديهم حساسية على ان يكون في الأقل 5 أشخاص وان كان من المستحسن استعمال 25 .
 - إجراء الفحوص الجلدية مثل فحص وخز الجلد أو غيرها
 - استعمال النماذج الحيوانية
- وغيرها من الفحوص التي ان كانت في الاتجاه الذي يشير الى إثارة الحساسية فعندها لا يعتمد أو يسوق الكائن المهندس وراثيا.

المصادر العربية

- الراوي، خاشع محمود .1987. "المدخل الى تحليل الانحدار"، كلية الزراعة / جامعة الموصل.
- الشخيلي، علي عبد الحافظ إبراهيم . 2004. "التنبؤ بالتركيب الثانوي للبروتينات باستخدام طرائق Chou-Fasman والتماثل والشبكات العصبية الاصطناعية" أطروحة دكتوراه، قسم الإحصاء / كلية الإدارة والاقتصاد / جامعة بغداد.
- Abd-Elsalam, K. 2003. Bioinformatic tools and guideline for PCR primer design. Afr. Biotechnology .2: 91-95.
- Abrahams, J., Berg, M., Batenburg, E .and Pleii, C .1990. Prediction of RNA secondary structure, including pseudo – knotting by computer simulation .Nuc. Acids . Res .18:3035-3044.
- Alexandrov , N. 1993. Structural argument for N-terminal initiation of protein folding .Protein Sci. 2:1989-1991.
- Anderson, N. 2004 .A Qualitative Assessment of Bioinformatics Tasks, Tools, and Workflow within an Academic Biological Research Laboratory. MSc. thesis, University School of Medicine .Oregon, USA.
- Andersson, A., Bernander, R. and Nilsson, P.2005. Dual – genome primer design for construction of DNA microarrays . Bioinformatics, 21:325-332.
- Angellotti, M. , Bhuiyan ,S. , Chen, G. and Wan , X. 2007 . CodonO: Codon usage bias analysis within and across genomes. Nuc .Acids Res. 35: W 132-W136.
- Arakawa, K., Tamaki, S., Kono, N., Kido, N., Ikegami, K. Ogawa, R. and Tomita, M. 2009 . Genome Projector : Zoomable genome map with multiple views. BMC Bioinformatics .10: 31 – 40.
- Audic, S. and Claverie, J. 1997. Detection of eukaryotic promoters using Markov transition matrices. Computers Chem.21:233-227.
- Audic, S. and Claverie, J. 1998. Visualizing the competitive recognition of TATA-boxes in Vertebrate promoters .Trends Genet. 14: 10-11.
- Behrouz , F.2007 . Data Communications and Networking .4th edition .MaGraw-Hill Comp. Inc. USA .
- Bendana , y. and Holmes, I . 2008. Colorstock, SScolor, Raton: RNA alignment visualization tools. Bioinformatics, 24:579 – 580.
- Bendtsen , J. Binnewies , T., Hallin , P. and Ussery, D.2005. Genome update: Prediction of membrane proteins in prokaryotic genomes .Microbiology, 151: 2119 – 2121.

-
- Bendtsen ,J. , Binnewies ,T. , Hallin ,P., Sicheritz-Ponten and Ussery , D. 2005.Genome update :Prediction of secreted proteins in 225 bacterial proteomes. Microbiology 151:1725-1727.
 - Bendtsen , J. , Nielsen, H., Heijne ,G. and Brunak, S.2004.Improved prediction of signal peptides: SignalP 3.0 J. Mol. Biol. 340:783-795.
 - Berg, J., Tymoczko ,J. & Stryer, .2002.Biochemistry.5th Edition.WH.Freeman and Company, New York .
 - Bernhart, S ., Hofacker, I., Will, S., Gruber, A. and Staalder, P.2008. RNAalifold :improved consensus structure prediction for RNA alignments .BMC Bioinfo. 9:474-486.
 - Binnewies ,T., Hallin, P., Henrik, H., Ussery, D.2005.Genome update: Proteome comparisons.151:1-4.
 - Bromberg ,Y., Yachdav, G. and Rost, B.2008. SNAP predicts effect of mutations on protein function . Bioinformatics 24:2397-2398.
 - Brown, N., Sander, C. and Bork P.1998.Frame: detection of genomic sequencing errors. Bioinformatics 14:367-371.
 - Brusic, V., Millot, M., Petrovsky , N., Gendel, S., Gigonazac ,O. and Stelman, S. 2003. Allergen database. Allergy 58:1093-1100.
 - Brusic, V., Petrovsky, N., Gendel, S., Millot, M., Gigonazac ,O. and Stelman, S . 2003.Computational tools for the study of allergens. Allergy 58:1083-1092.
 - Butte, A. 2008 . Translational bioinformatics: Coming of age. J. Am. Med Inform.Assoc.15:709-714 .
 - Bryan , M . 1994 . Multivariate Statistical Methods APRIMER . 2nd edition , Chapman and Hall/CRC, Boca Raton , USA .
 - Catanho, M., Mascarenhes, D., Degrave, W. and Miranda, A .2006. GenoMycDB: a database for comparative analysis of mycobacterial genes and genomes .Genet. Mol. Res. 5:115-126.
 - Chamber ,J. , Cleveland, W., Kleiner, B., and Tukey, P. 1983. Graphical Methods for Data Analysis. Belmont, CA: Wadsworth.
 - Chap , L . 2003 . Introductory Biostatistics Methods.John Wiley and sons ,Inc.
 - Chen, T., Yu, W., Izard, J. Baranov, O., Lakshmanal, A. and Dewhirst ,F. 2010. The human oral microbiome database: a web accessible resource for investigating and microbe taxonomic and genomic information . Database 2010:bago13.
 - Chen, X., Li, Ma, B. and Tromp , J. 2002. DNA Compress: fast and effective DNA sequence compression. Bioinfo Appl. Note, 18: 1696-1698.

-
- Choi ,C, Munch ,R., Leupold , S., Klein, K., Siegel , I., Thielen , B. *et al.*2007 SYSTOMONAS –an integrated database for systems biology analysis of Pseudomonas Nuc. Acids Res .35:D 533- 537.
 - Coutinho, P. 2005 . ZF- human disease :A database for zebra fish model of human genetic disease. BMC Bioinfo. 6:P 7
 - Crooks, G, Hon, G., Chandonia , J. and Brenner, S. 2004 . WebLogo: A sequence logo generator. Genome Res. 14 :1188-1190.
 - Delamarche, C. 2000. Color and graphic display (CGD): Programs for multiple sequence alignment analysis in spreadsheet software. BioTechniques 29:100-107.
 - Dieffenbach ,C., Lowe ,T. and Dveksler ,G. 1993 . General concepts for PCR primer design. Genome Res. 3:530- 537.
 - Dieffenbach, C. Lowe, T., Dveksler ,G.1995.G General Concepts for PCR design .*In*" PCR primer ,A Laboratory Manual "C. Dieffenbach and G. Dveksler (Eds).Cold Spring Harbor Laboratory, New York. PP133-155.
 - Dillon, R., Webster ,G., Weightman, A .and Charnley A . 2010. Diversity of out microbiota increases with aging and starvation in the desert locust . Annt .van Leew. 97: 69-77.
 - Dlakic, M., Ussery , D. and Brunak ,S, 2004. DNA Bendability and Nucleosome Positioning in Transcriptional Regulation .*In*" DNA Conformation and Transcription" T. Ohyama (ed). Eurekah .Company.
 - Edwards, J. and Palsson , B. 1998. How will informatics influence metabolic engineering? Biotechnol. Bioeng.58:162-169.
 - Edwards , J. and Palsson, B. 2000. Metabolic flux balance analysis and the *In silico* analysis of *Escherichia coli* K-12 gene deletions . BMC Bioinfo. 1:1-10
 - Ellis , L., Hershberger ,C., Bryan, E. and Wackett, L. 2001 .The University of Minnesota biocatalysis / biodegradation database : emphasizing enzymes . Nuc. Acids Res. 29 : 340-343.
 - El- Maarri, O., 2003 . Methods : DNA methylation. Adv. Exp. Med. Biol. 544:197-204
 - Erhardt, P.2003. A human drug metabolism database : Potential roles in the quantitative predictions of drug metabolism and metabolism- related drug –drug interactions. Curr. Drug Metab. 4:411-422.
 - FAO and WHO .2003. Codex Principles and Guidelines on Foods Derived from Biotechnology . FAO, Rome.
 - Fersht, A. and Daggett ,V. 2002. Protein folding and unfolding at atomic resolution .Cell 108:1-20.
 - Field , D., Tiwari ,B. and Snape, J.2005. Bioinformatics and data management support for environmental genomics . PLoS Biology 3:1352- 1353.

-
- Forno, E., Onderdonk, A., McCracken, J., Litonjua, A., Laskey, D., Delaney, M. *et al.* 2008 . Diversity of the gut microbiota and eczema in early life. Clin. Mol. Allergy 6: 11-19 .
 - Forouzan, Behrouz A. . 2007. "Data Communication and Networking", Fourth Edition, the McGraw – Hill Companies
 - Fourment, M. and Gillings, M. 2008 A. comparison of common programming languages used in bioinformatics. BMC Bioinformatics 9:82-90
 - Freund, John E. . 1979 . "Modern Elementary Statistics", Fifth Edition, Prentice / Hall International, Inc., London.
 - Frolkis, A., Knox, C. Lim, E., Jewison, T., Law, V., Hau, D. 2010.SMPDB; The small molecule pathway database .Nuc. Acids Res. D480-487.
 - Fuchs, R. and Astwood, J. 1996. Allergenicity assessment of foods derived from genetically modified plants. Food Technol . 50: 83-88
 - Fujimori, S., Washio, T. and Tomita, M.2005. GC- compositional strand bias around transcription start sites in plant and fungi . BMC Genomics 6 :26-36
 - Fulekar, M. and Sharma, J. 2008 . Bioinformatics applied in bioremediation .Innov .Romanian Food Biotechnol. 2:28-36.
 - Gangadharaiyah, D. 2010. Patterns of dipeptide usage for gene prediction. MSc thesis. Wright state University.
 - Gill, S., Pop, M., DeBoy, R., Eckburg, P., Turnbaugh, P. Samuel, B. *et al* . 2006 . Metagenomic analysis of the human distal gut microbiome. Science 312:1355-1359
 - Gogarten, J. and Hilario, E. 2006 .Inteins, Introns, and homing endonucleases : recent revelations about the life cycle of parasitic genetic elements. BMC Evol. Biol. 6:1-5.
 - Goodman, R., Hefle, S., Taylor, S. and van Ree, R . 2005. Assessing genetically modified crops to minimize the risk of increased food allergy : a review. Int .Arch. Allergy Immunol. 137:151-152.
 - Gottesman, S. 2004. The small RNA regulators of *Escherichia coli* : Roles and mechanism. Ann Rev. Microbiol . 58:303-328
 - Goyal, S., Lillian, T., Blumberg, S., Meiners, J., Meyhfer, E. and Perkins, N. 2007. Intrinsic curvature, of DNA influences LasR- mediated looping Biophys . J. 93:4342-4359
 - Grafahrend –Belau, E., Weise, S., Koschutski, D., Scholz, U., Junker, B. and Schreiber, F. 2010 MetaCrop :a detailed database of crop plant metabolism . Nuc. Acids Res. 36 : D 954-D 958.
 - Grigoriev, A. 1998. Analyzing genomes with cumulative skew diagrams . Nuc. Acids Res. 26:2286-2290.

-
- Gruber, A., Lorenz , R., Bernhart, S., Neuba, R. and Hofacker, I. 2008 . The Vienna RMA website. Nuc. Acid Res. 36:1-5
 - Hallin ,P., Binnewies ,T. and Ussery .D.2004 . Genome update : chromosome atlases . Microbiology, 150:3091-3093.
 - Hair, Joseph F., Anderson, Rolph E., Tathan, Ronald L. 1987. "Multivariate Data Analysis with Readings", Second Edition, Macmillan Publishing Company.
 - Hallin P., Coenye , T., Binnewies ,T., Jarmer, H. Staerfeldt , H. and Ussery , D.2004 . Genome update: correlation of bacterial genomic properties. Microbiology 150:3899-3903
 - Hallin, P., Nielsen, N., Devine, K., Binnewies, T., Willenbrok , H. and Ussery, D. 2005. Genome update : base skews in 200 bacterial chromosomes, 151:633-636.
 - Hallin, P., Staerfeldt, H., Rotenberg, E., Binnewies ,T., Benham ,C. and Ussery, D.2009. Genewiz browser: An interactive tool for visualizing sequenced chromosomes. Stand. Genomic. Sci. 1:204-215
 - Hallin ,P. and Ussery, D. 2004.CBS genome atlas database :a dynamic storage for bioinformatic result and sequence data. Bioinformatics ,12:3682-3686.
 - Ham , M . 2000 . Principles of Neurocomputing for Sciences and Engineering . McGraw / Hill Comp. Inc. USA .
 - Hamady , M. and Knight, R. 2009. Microbial community profiling for human microbiome projects: Tools technique; and challenges. Genome Res. 19:1141-1152
 - Hamilton, G.2007. A Visualization tool for interpreting BLAST search results. MSC dissertation University of Glasgow, UK.
 - Hancock, T., Takigawa, I. and Mamistuka, H. 2010. Mining metabolic pathways through gene expression . Bioinformatics, 26:2128-2135.
 - Handelsam, J. 2004 . Metagenomics :Application of genomics to Uncultured microorganisms. Mol. Biol .Rev 68: 669-685.
 - Hattori, M. and Taylor, T. 2009 .The human intestinal microbiome: A new frontier of human biology. DNA Res. 16:1-12
 - Havlak, P., Chen, R., Durbin , K., Egan, A., Ren, Y., Song, X. *et al.* 2004. The atlas genome assembly system. Genome Res. 14:721-732.
 - Hofacker, I., Fekete, M. and Stadler, P. 2002. Secondary structure prediction for aligned RNA sequences. J. Mol. Biol. 314: 1059-1066
 - Hofacker, I. and Stadler, P. 2006. Memory efficient folding algorithms for circular RMA secondary structures. Bioinformatics 22:1172-1176.
 - Hutzler, M., Messing, D. and Wienkers, L. 2005. Predicting drug- drug interactions in drug discovery Curr. Opin . Drug Discov. Devel. 8:51-58

-
- Ivanciuc ,O., Garcia, T., Torres , M., Catherine, H. and Braun, W. 2009 .The property distance index PD predicts peptide that cross- react with IgE antibodies. Mol. Immunol. 46: 873-883 .
 - Jacek , Z .1992 . Introduction to Artificial Neural Systems . 2nd edition .West Publishing Comp . USA .
 - Jameson, B. and Wolf, H. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosci. 4:181-186 .
 - John , F. 1979 . Modern Elementary Statistics .5th edition .Prentice /Hall International , Inc. London .
 - Jones ,B., Begley, M., Hill ,C. Gahan, C and Marchesi, J,. 2008. Functional and comparative metagenomic analysis bile salt hydrolase activity in the human gut microbiome . Proc. Nalt. Acad. Sci. 105: 13580-13585.
 - Jensen, L., Friits. C. and Ussery, D. 1999 . Three Views of microbial genomes. Res. Microbiol. 150:773-777.
 - Kanhere, A and Bansal ,M. 2003. An assessment of three dinucleotide parameters to predict DNA curvature by quantitative comparison with experimental data. Nuc. Acids Res. 15:2647-2658.
 - Kapushesky , M., Emam , I., Holloway, E. , Kurnosov, p., Zorin, A., Malone, J. 2010. Gene expression atlas at the European Bioinformatics Institute. Nuc. Acids Res. D690-D698.
 - Kawagi, H. and Hayashizaki, Y. 2008. Exploration of small RNAs. PLoS Genet . 4:e22-e27.
 - Kill, K., Ferchaud, J., David, C., Binnewies, T., Wu, H., Sicheritz- Ponten , T. 2005 . Genome update : distribution of two- component transduction systems in 250 bacterial genomes. Microbiology 151:3447-3452.
 - Kill, K ., Binnewies ,T., Sicheritz- Ponten , T., Willenbrok , H., Hallin, P., Wassenaar, T., and Ussery, D. 2005. Genome update : sigma factors in 240 bacterial genomes. Microbiology 151:3147-3150.
 - King, R., Stansfield ,W. and Mulligan, P. 2006 . A Dictionary of Genetics .7th ed. Oxford Univ. Press: Oxford & New York.
 - Kleter, G. and Peijnenburg ,A . 2002. Screening of transgenic proteins expressed in transgenic food crops for presence of short amino acid sequences identical to potential , IgE- binding linear epitopes of allergens. BMC Struct. Biol. 2:8-18.
 - Knox, C. Law, V., Jewison, T., Liu, P., Ly. S., Forlkis, S. *et al* .2011. Drugbank 3.0: a comprehensive resource for 'omics' research on drugs. Nuc. Acids Res. 39:D1035-D1041.

-
- Kraus , R., Murray ,E., Wiley , S., Zink, N., Loritz, K., Gelembiuk, G. and Mertz, J. 1996 . Experimentally determined weight matrix definitions of the initiator and TBP binding site elements of promoters .Nuc Acids Res . 24:1531-1539.
 - Krause, L., Diaz, N., Bartels , Edwards, R., Puhler, A., Rohwer, F. *et al* .2006. Finding novel genes in bacterial communities isolated from the environment . Bioinformatics 22:281-289.
 - Kozobay – Avraham, L., Hosid, S. and Bolshoy , A . 2006. Involvement of DNA curvature in intergenic regions of prokaryotes .Nuc. Acids Res . 34:2316-2327.
 - Kumar , L., Futschik , M. and Herzel , H. 2006 . DNA motifs and sequence periodicities. *In Silico Bio*. 6 : 71-78.
 - Kummerfeld , S. and Teichmann, S. 2006. DBD: a transcription factor prediction database, Nuc. Acids Res. 34:D74-D81.
 - Latheef , S. 2010. A database for inborn errors of metabolism in the Indian state of Andhra Pradesh. Bioinformation 4: 276-277.
 - Le T. C. 2003 . "Introductory Biostatistics", John Wiley and sons, Inc.
 - Lee, C., Grasso ,C. and Sharlow , M. 2002. Multiple sequence alignment using partial order graphs. Bioinformatics . 18:452-464.
 - Lee, C. and Wang , Q . 2005. Bioinformatics analysis of alternative splicing .Brief. Bioinfo. 6: 23-33.
 - Lefebure, T. and Stenhope , M. 2007 . Evolution of the core and pan –genome of *Streptococcus* : positive selection , recombination and genome composition . Genome Biol. 8: R71.1-R71.17.
 - Lefranc, M. 2001. IMGT. The international ImMunoGeneTics database. Nuc. Acids Res .29:207-209.
 - Lefranc , M., Giudicelli, V., Busin , C., Bodmer , J., Muller , W., Bontrop, R. *et al* . 1998. IMGT. The international ImMunoGeneTics database. Nuc. Acids Res. 26:297:303.
 - Li, X. and Pan , X. 2001. New method for accurate prediction of solvent accessibility from protein sequence . Proteins 42:1-5.
 - Lim, S., Tan, T. and Tong , S. 2010.Computational epigenetics: The new scientific paradigm . Bioinformation , 4:331-337
 - Lipman, D., Altschul , S. and Kecicioglu , J. 1989. A tool for multiple sequence alignment. Proc. Natl Acad . Sci. 86:4412-4415 .
 - Lipps , G. 2008. Plasmids : Current Research and future Trends . Caister Acad. Press: Germany.

-
- Liu. Z. , Wu, L., Wang, Y., Zhang , T and Chen , L., 2010. Prediction of protein- RNA binding sites by a random forest method with combined features. *Bioinformatics*, 26:1616-1622.
 - Maidak, B., Cole, J., Lilburn, T., Parker , C., Saxman , P., Farris, R. *et al* 2001 . The RDP-II (Ribosomal Database project).*Nuc. Acids Res.* 29:173-174.
 - Malandain , H. 2004. Basic immunology , allergen prediction, and bioinformatics. *Allergy* 59:1011-1012.
 - Manly, Bryan F.J. (1994), "Multivariate Statistical Methods APRIMER", Second Edition, Chapman and Hall.
 - Mano, A., Tuller, T., Beja, O. and Pinter, R. 2010 . Comparative classification of species and the study of pathway evolution bases on the alignment of metabolic pathway . *BMC Bioinfo.* 11: S38, 1-10.
 - Mari , A., Palazzo, S., Zennaro, R. *et al* . 2006. Bioinformatics applied to allergy : Allergen database, from collecting sequences information to data integration. The allergome platform as a model. **244:97-100.**
 - Mathe, C., Sagot , M., Schiex , T, and Rouze , P. 2002. Current methods of prediction, their strengths and weakness. *Nuc. Acids Res .* 30:4103-4117
 - Mathews , D., Moss, W. and Turner, D. 2010. Folding and finding RNA secondary structure . *Cold Spring Harb. Prospect. Biol.* A003665:1-16.
 - Matthiesen ,R.(Ed). 2010. *Bioinformatics Methods in Clinical Research* Humana Press.
 - Mathews ,D., Sabina J., Zucker, M. and Turner, H. 1999. Expanded Sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure .*J. Mol. Biol.* 288: 911-940.
 - Mattick , J. 2004. RNA regulation : a new genetics ? *Nature Rev/ Genetics.* 5: 316-323
 - McCaskill, J . 1990. The equilibrium partition function and pair binding probabilities for RNA secondary structure . *Biopolymers* 24:1105-1119
 - McEntyre, J. and Ostell , J. *The NCBI Handbook.* 2002. The NCBI Handbook . Bethesda (MD) . National center for Biotechnology Information .US .
 - Medini , D., Donati, C., Tettelin, H., Masignani , V. and Rappuoli , R. 2005. The microbial pan- genome . *Curr. Opin. Genet.* 15:589-594
 - Meinhart , A., Cramer , P . 2004. Recognition of RNA polymerase II carboxy – terminal domain by 3'-RNA-processing factors . *Nature* 30(6996):223-226 .
 - Merelli, E., Armano, G., Cannata , N., Corradini , F., d'Inverno , M., Lord , P . *et al* . 2004 Agents in bioinformatics computational and systems biology . *Brief . Bioinfo* 8:45-59.

-
- Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S. and Kanehisa, M. 2010. Pathperd : an enzyme – catalyzed metabolic pathway prediction server' Nuc. Acids Res. 38:W138-W143
 - Necsulea, A and Lobry, J. 2007 .A new method for assessing the effect of replication on DNA base composition asymmetry . Mol. Biol. Evol. 24:2169-2179 .
 - Neil, T. 2002 . Applied Multivariate Analysis .Springer – Verlag, Inc. . New York
 - Ng, P. and Henikoff, S. 2003. SIFT: predicting amino acid changes that affect protein function. Nuc. Acids Res. 31: 3812- 3814.
 - Ng, P. and Henikoff, S. 2006. Predicting the effects of amino acid substitutions of protein function . Annu . Rev .Genomic Hum. genet .7: 61-80.
 - Nikola, K. K. . 1996. "Foundations of Neural Networks Fuzzy Systemes and Knowledge Engineering", Second Edition, the MIT press Cambridge, Massachusetts, London, England
 - Nikolova, D. and Toncheva, D. 2008 . RNA interference - regulations and application in oncology. J . Cancer Mol. 4:67-77.
 - Notredam, C. Higgins, D. and Heringa, J. 2000. "T- Coffee" A novel method for fast and accurate multiple sequence alignment. J. Mol . Biol . 302: 205-217 .
 - Ojangu, E., Tover, A., Teras, R. and Kivisaar, M. 2000. Effects of combination of different -10 hexamers and downstream sequences on stationary – phase – specific sigma factor δ^s - dependent transcription in *Pseudomonas putida*. J. Bacteriol. 182:6707-6713.
 - Oti, M., Snel, B., Huynen, M. and Brunner, H. 2006. Predicting disease genes using protein – protein interactions . J. Med. Genet . 43:691-698.
 - Pace, J. and Feschotte, C. 2007. The evolutionary history of human DNA transposons : evidence for intense activity in the primate lineage , Genome Res. 17: 422-432.
 - Passarge, E. 2001. Color Atlas of Genetics . Thieme . Stuttgart, New York.
 - Pazos, F., Gujas, D., Valencia, A., and Delonzeno, V. 2005. MetaRouter : bioinformatics for bioremediation . Nuc. Acids Res. 33:D588-D592.
 - Pedersen, A., Jensen, L., Brunak, S., Staerfeldt, H. and Ussery, D. 2000. A. and structural atlas for *Escherichia coli* . J. Mol . Biol. 299: 907-930.
 - Pedersen, J., Meyer, I., Forsberg, R., Simmonds, P. and Hein, H. 2004. A comparative method for finding and folding RNA secondary structures within protein-coding genes. Nuc. Acids Res 32:4925-4936.
 - Pesany, B. and Zhang, Z. 2004. Bioinformatics and cancer target discovery. Drug . Discov. Today 15:795-802.
 - Pinter, R., Roklenko, O., Yeager-Lotem, E. and Ziv-Ukelson, M. 2005. Alignment of metabolic pathways. Bioinformatics, 21:3401-3408.

-
- Poaxevanis, A. and Quellette, B. (Eds) .2001. Bioinformatics : A practical Guide to the Analysis of Genes and Proteins . 2nd Edition. John Wiley & Sons, INC. publication: New York Chichester.
 - Qin, S., He, Y. and Pan, X. 2005. Predicting protein secondary structure and solvent accessibility with an improved multiple linear regression method . *Proteins* 61:473-480.
 - Qureshi, M., Rotenberg, E., Staerfeldt, H., Hansson , L. and Ussery , D. 2010. The genome atlas resource. *CSBio* 115:141-150.
 - Raasch , P., Schmitz, U., Patenge, N., Vera, J., Kreikemeyer, B. and Walkenhaure, O. 2010. Non- coding RNA detection methods combined to improve usability, reproducibility and precision. *BMC Bioinfo .II*:1-12.
 - Ravel , J., Gajer, Abdo, Z., Schneider, M., Koenig, S., McCulle, S., Karlebach, S. *et al.* 2011 .Vaginal microbiome of reproductive –age women . *Proc. Natl. Acad. Sci.* 108 Suppl. 4680-4687.
 - Rocha, E. 2004. The replication –related organization of bacterial genomes . *Microbiology* 150:1609-1627.
 - Rocha, E. and Danchin , A . 2003. Gene essentiality determines chromosome organization in bacteria. *Nuc. Acids Res.* 31:6570-6577.
 - Rocha, E., Touchon, M. and Feil, E. 2006. Similar Compositional biases are caused by very different mutational effects. *Genome Res.* 16:1537-1547.
 - Rouvinen, J., Janis, J., Laukkanen, M., Jylha, S., Niemi, M., Paivinen, T. *et al.* 2010. Transient dimmers of allergens. *PLoS ONE* 15:e9037-e9045.
 - Roux, K. 1995. Optimization and troubleshooting in PCR. *Genome Res.* 4:5185-194.
 - Russell, R. 2002. Meeting review ; Bioinformatics and medicine .form molecules to humans, Virtual and real .*Comp. Funct. Genom.* 3:270-276.
 - Ruvinsky, A. and Ward, W. 2006. A gradient in the distribution of introns in eukaryotic genes *J. Mol. Evol.* 63:136-141.
 - Rychlik, W., Spencer, W, and Rhoads, R. 1990. Optimization of the annealing temperature form DNA amplification *in vitro* . *Nuc. Acids Res.* 18:6409-6412.
 - Sakharkar, M., Passetti, F., de Souza, J., Long, M. and de Souza, S. 2002. ExInt: an exon intron database. *Nuc. Acids Res.* 30:191-194.
 - Sandhu, K. and Acharya, K. 2005. ExPrimer : to design primers from exon- exon junctions. *Bioinfamatics*, 21:2091-2092.
 - Sanghani, S., Zakrzewska, K. Harey, S. and Lavery , R. 1996. Molecular modeling of (A₄T₄NN)_n and (T₄A₄NN)_N: sequence element responsible for curvature. *Nuc. Acids Res.* 24: 1632-1637.

-
- Saxonov ,S., Daizadeh, I., Fedorov, A. and Gilbert, W. 2000. EID: the exon – intron database –an exhaustive database of protein –coding intron –containing genes . Nuc. Acids Res .28:185-190.
 - Schonbach, C., Ranganathan, S., Brusica , V. (Eds). 2008.Immunoinformatics . Springer , New York .
 - Scipioni, A., Anselmi, C., Zuccheri, G., Samori, B. and Santis , P. 2002. Sequence – dependent DNA curvature and flexibility from scanning force microscopy images. Biophys. J. 83: 2408-2418.
 - Selkov. E., Grechkin, Y., Mikhailova, N. and Selkov , E. 1998. MPW: the metabolic pathways database. Nuc. Acids Res .26: 43-45.
 - Shabalina , S. and Spiridonov, N. 2004. The mammalian transcriptome and the function of non- coding DNA sequences. Genome Biol. 5: 105-112.
 - Shen, Z., Qu, W., Wang , W., La, Y., Wu, Y., Li, Z. *et al.* 2010. MP primer: a program for reliable multiplex PCR primer design. BMC Bioinformatics 11:143-149.
 - Shimoni , Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia , S., Biham , O. and Margalit , H. 2007. Regulation of gene expression by small non- coding RNAs: a quantitative view , Mol. Sys. Biol. 3:2-9.
 - Silva, M., Grossi, M., Chrispeels, M., Togawa , R. and Neshich, G. 2000. Analysis of structural and physico - chemical parameters involved in the specificity of binding between – amylases and their inhibitors. Protein Eng. 13:167-177.
 - Sim, J., Kim, S. and Lee, J. 2005. Prediction of protein solvent accessibility using fuzzy K- nearest neighbor method. Bioinformatics 21:2844-2849.
 - Simon R. 2005 . Bioinformatics in cancer therapeutics hype or hope? Nature Clin. Pract. Oncol. 2:223,
 - Singh, S., Taneja, B., Salvi , S. and Agrawal ,A. 2009. Physical properties of intact proteins may predict allergenicity or lack thereof . PLoS ONE 4:e6273.
 - Singh, V. and Kumar, A., 2001. PCR primer design , Mol , Biol. Today, 2:27-32.
 - Skovgaard , M., Jensen ,L., Brunak, S., Ussery , D. and Krogh, A. 2001. On the total number of genes and their length distribution in complete microbial genomes. Trends Genet 17:425-428.
 - Spriggs, R., Murakami , Y., Nakamura , H. and Jones , S. 2009. Protein function annotation from sequences : prediction of residues interacting with RNA. Bioinformatics 25:1492-1497.
 - Stanke, M. 2003. Gene Prediction with a Hidden Markov Model. Ph. Dissertation. University of Gottingen.
 - Stevens, R., Goble, C., Baker, P. and Brass. A. 2001. A classification of tasks in bioinformatics . Bioinformatics 17:180-188.

-
- Stothard, P., Domselaar, G., Shrivastava, S., Guo, A., o' Nill, B., Criz, J. et al 2003. BacMap : an interactive picture atlas of annotated bacterial genes , Nuc. Acids Res. 33:D317-D320.
 - Swingley, W., Blankenship, R. and Raymond, J. 2008. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families . Mol. Biol. Evol. 25:643-654.
 - Takada, K., Hanawa, K., Lee, S., Himeno, H. and Muto, A. 2000. The structure and function of tmRNA. Nuc. Acids Res. Suppl. No.2: 65-66.
 - Thompson, J., Higgins, D. and Gibson, T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position- specific gap penalties and weight matrix choice . Nuc. Acids Res. 22: 4673-4680.
 - Thompson, J., Plewniak, F. and Poch, O. 1997. BALiBASE:: A benchmark alignment database for the evaluation of multiple alignment programs . Bioinformatics 15:87-88.
 - Thompson, J., Plewniak, F. and Poch, O. 1999. A comprehensive comparison of multiple sequence alignment programs . Nuc. Acids. Res. 27: 12682-12690.
 - Thomas, K., Bannon, G., Hefle, S., Herouet, C., Holsaple, M., Ladics, G., MacIntosh, S. and Privalle, L. 2005. *In Silico* methods for evaluating human allergenicity to novel proteins . International Bioinformatics Workshop Meeting Reports . 23-24 Feb. 2005.
 - Timm H. N. 2002 . "Applied Multivariate Analysis", Springer – Verlag New York , Inc.,.
 - Tonges, U. , Perrey, S., Stoye, J. and Dress, A. 1996. A general method for fast multiple sequence alignment. Gene 172:GC33-GC41.
 - Turnbaugh, P., Hamady, M., Yatsunenko, T., Cantarel, B., Duncan A., Ley, *et al.* 2009. A core gut microbiome in obese and lean twins . Nature, 457:480-484.
 - Ussery, D. 2004. Genome update: chromosome atlases. Microbiology 150:3091-3039.
 - Ussery, D., Binnewies, T., Gouveia- Oliveira, R., Jarmer, H. and Hallin, P. 2004. Genome update : DNA repeats in bacterial genomes. Microbiology , 150:3519-3521.
 - Ussery, D. and Hallin, P. 2004. Genome update : Length distributions of sequenced prokaryotic genomes. Microbiology , 150: 513-516.
 - Ussery, D., Hallin, P., Lagesen, K. and Wassenaar, T. 2004. Genome update : tRNAs in sequenced microbial genomes. Microbiology 150:1603-1606.
 - Ussery, D., Jensen, M., Poulsen, T. and Hallin, P. 2004. Genome update: alignment of bacterial chromosomes . Microbiology 150:2491-2493.

-
- Ussery, D., Tindbaek , N. and Hallin, P. 2004. Genome update : promoter profiles Microbiology 150: 2791-2793.
 - Ussery , D., Wassenaar, T. and Borini, S. 2009. Computing for Comparative Microbial Genomics. Springer , New York.
 - Vaknin , K, Goren, A. and Ast , G. 2009. TEs or not TES ? That is the evolutionary question. J. Biol. 8:83.1-83.3.
 - Vaucheret , H. 2006. Post –transcriptional small RNA pathways in plants : mechanisms and regulation . Genes Dev. 20:759-771.
 - Wang , TT., Kaloper , M. and Benham , J. 2006. SIDDBASE: a database containing the stress- induced DNA duplex destabilization (SIDD) profile of complete microbial genomes .Nuc. Acids Res. 34:D373-D378.
 - Wang, H., Noordewier, M. and Benham, C. 2004. Stress- induced DNA duplex destabilization (SIDD) in the *E coli* genome : SIDD sites are closely associated with promoters. Genome Res. 14: 1575-1584.
 - Wang, Z., Chen ,Y. and Li, Y. 2004. A brief review of computational gene prediction methods . Geno. Prot. Bioinfo. 2:216-221.
 - Wassenaar, T., Binnewies , T., Hallin , P. and Ussery, D. 2010 74 Tools for comparison of bacterial genomes. In "Handbook of Hydrocarbons and Lipid Microbiology ". Springer- Verlag. Berlin , Heidelberg .
 - Wells, W., Viola, P., Atsumi, H., Nakajima, S. and Kikinis, R. 1996. Multi- modal volume registration by maximization of mutual information Med. Image. Anal. 1:35-51.
 - Wessler , S. 2001. Plant transposable elements. Plant Physiol. 125: 149-151.
 - Wessler, S. 2006. Eukaryotic transposable elements and genomes evolution special feature . Proc. Natl. Acad. Sci. 103:17600-17601.
 - Willenbrok , H., Binnewies , T., Hallin, P. and Ussery, D. 2005. Genome update : 2D clustering of bacterial genomes . Microbiology 151: 33-336.
 - Wishart , D. , Knox, C. Guo, A., Eisner, R., Young, N. Gautam, B. *et al* 2009 HMDB : a knowledgebase for the human metabolome . Nuc. Acids Res. 37:D603-D610.
 - Wishart , D., Knox, C., Guo, A., Shrivastava, S., Hssanali, M., Stothard, P. *et al*. 2011. Drug Bank: a comprehensive resource for *in silico* drug discovery and exploration . Nuc. Acids Res. 34:D668-D672.
 - Wittig , U. and Beuckelaer , A. 2001. Analysis and comparison of metabolic pathway databases . Brief Bioinfo . 2: 126-142.
 - Wu., C., Huang, H., Yeh, L., Banker, W. 2003. Protein family classification and functional annotation. Comp. Biol. Chem. 27:37-47.

-
- Yin, Y., Laing, C., Leontis, N. and Schlick, T. 2008. Annotation of tertiary interactions in RNA structures reveals variations and correlations –RNA 14:2465-2477.
 - Xiong, J. 2007. Essential Bioinformatics Cambridge University press. UK.
 - Yakouchuk , P. Protozanova , E. and Frank –Kamenets Kij ,M. 2006. Base- stacking and base-pairing contributions into thermal stability of the DNA double helix .Nuc .Acids Res .34:564-574.
 - Ye, Y. and Tang , H. 2008. An ORFome assembly approach to metagenomics sequence analysis. Comput. Syst. Bioinformatics Conf. 7:3-13.
 - Yiannias, J. and El- Azhary, R. 2000. Contact allergen avoidance program : a topical skin care product database . Am .J. contact Dermatitis . II: 243-247.
 - Zhang ,H., Yu, H., Ren, J., Qu, X . 2006. Reversible B/ZDNA transition under the low salt condition and non-B- from polyA polyT selectivity by a cubane- like europium – L- aspartic acid complex . Biophys . J. 90: 3203-3207.
 - Zhang, M. 2002. Computational prediction of eukaryotic protein –coding genes. Nature Rev. / Genetics 3: 398-700
 - Zhang ,Y. 2008. Progress and challenges in protein structure prediction . Curr . Opin. Struct. Biol. 18: 342-348.
 - Zuker, M. and Stiegler , P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information . Nuc. Acids Res. 9: 133-148.
 - Zurada M. Jacek, "Introduction to Artificial Neural Systems", second edition, West Publishing Company, 1992.

فهرس عربي

إزالة التمنيع.....	711	ا
اثر المصفوفة.....	276	ا
استبدال الحامض الأميني.....	518	أ
استبدالات.....	434, 435, 518	أ
استرداد.....	11, 12, 43, 44, 45, 49, 693	ا
استعلامات معقدة.....	44	ا
استعمال الشفرات.....	23, 24, 27, 38, 66, 90, 91,	ا
احتمالية الاستبدال العشوائي.....	384	ا
أحداث الانفراج الوراثي.....	436	ا
أحصاء الاختبار.....	315	أ
اختبار الفرضية.....	316, 319, 320	أ
اختزال الجينوم.....	415, 520	ا
اختلاف الشفرات الوراثية.....	429	ا
اختناق المجتمع.....	512	ا
إدارة قواعد البيانات.....	30, 42	ا
إدلاف الجينات.....	424, 556	ا
ارتباط الاكسونات.....	541	ا
أشجار العلاقات التطورية.....	384, 387, 388, 405,	ا
أشجار ثنائية التفرع.....	419	ا
أشكال الحياة القديمة.....	597	ا
أصبع الزنك.....	197, 208, 632, 633	ا
إزاحة الأطر.....	103, 480, 569	ا
إزالة المثلث.....	233	ا
إزالة الانترونات.....	104, 149, 167	ا
إزالة التحسس.....	721	ا
استعمال متغايرات عشوائية.....	659	ا
استقلالية الاحتمالات.....	291	ا
استعلامات معقدة.....	44	ا
استعمال الشفرات.....	23, 24, 27, 38, 66, 90, 91,	ا
احتمالية الاستبدال العشوائي.....	384	ا
أحداث الانفراج الوراثي.....	436	ا
أشجار العلاقات التطورية.....	384, 387, 388, 405,	ا
أشجار ثنائية التفرع.....	419	ا
أشكال الحياة القديمة.....	597	ا
أصبع الزنك.....	197, 208, 632, 633	ا
إزاحة الأطر.....	103, 480, 569	ا
إزالة المثلث.....	233	ا
إزالة الانترونات.....	104, 149, 167	ا
إزالة التحسس.....	721	ا
إصدارات.....	62, 66, 67, 264, 393, 403, 464,	ا
إشارات توائم الارجنين.....	637	ا
إسكات الجين.....	222, 223, 228, 233	ا
إسطوانات بيتا.....	638, 642, 643	ا
أستلة.....	209, 226, 235, 237, 242	ا
اختبار الفرضية.....	316, 319, 320	ا
اختزال الجينوم.....	415, 520	ا
اختلاف الشفرات الوراثية.....	429	ا
اختناق المجتمع.....	512	ا
إدارة قواعد البيانات.....	30, 42	ا
إدلاف الجينات.....	424, 556	ا
ارتباط الاكسونات.....	541	ا
أشجار العلاقات التطورية.....	384, 387, 388, 405,	ا
أشجار ثنائية التفرع.....	419	ا
أشكال الحياة القديمة.....	597	ا
أصبع الزنك.....	197, 208, 632, 633	ا
إزاحة الأطر.....	103, 480, 569	ا
إزالة المثلث.....	233	ا
إزالة الانترونات.....	104, 149, 167	ا
إزالة التحسس.....	721	ا
إصدارات.....	62, 66, 67, 264, 393, 403, 464,	ا
إشارات توائم الارجنين.....	637	ا
إسكات الجين.....	222, 223, 228, 233	ا
إسطوانات بيتا.....	638, 642, 643	ا
أستلة.....	209, 226, 235, 237, 242	ا

اصطفاف التراكيب 409
اصطفاف التواليات المتعددة 401

أصل التضاعف 109, 118, 577, 585
أصناف الاكسونات 500
أصناف البلازميدات 113

إضافة البايوتين 230, 236, 543
إضافة السكريات 128, 212, 261
إضافة الدهون 635

أطالس 21, 34, 61, 556, 576, 577, 583, 585, 586, 587, 588, 590, 592, 593, 594, 595, 713
أطالس التعبير 590
أطالس الجينومات 21, 34, 556, 576, 577, 583, 585, 587, 595
أطالس الزعزعة 592
أطالس العلاقات التطورية 595
أطالس المقارنة 593, 594, 595
أطر القراءة 3, 6, 11, 12, 42, 103, 104, 130, 171, 393, 479, 480, 482, 483, 496
أطر القراءة الستة 393

أطر القراءة المتعكسة 482

أطر القراءة المفتوحة 3, 6, 11, 12, 479, 480, 482, 483
أطراف الكروموسومات... 124, 126, 127, 128

أطلس القواعد 591, 592
أطلس الكروماتين 213, 591
أكسنة 134, 169, 170

الأثير 270
الأجسام البيروكسيمية 69
الأجسام المضادة 260, 633, 635, 643, 711, 713, 715, 716
الأجسام المضادة النانوية 713
الأجسام النانوية 260
الإجهاد 189
الإجهاد ألتأكسدي 128, 225
الإجهاد الجينومي 129
الاحتمال أعظم 452, 460
الاحتمالات المستقلة والمشروطة 288
الاحتمالية البعدية 460
الاحتمالية السابقة 460
الاحتمالية المشروطة 291
الأحجية الرباعية 458, 459
الإحصاء الحيوي 596
الإحصاء الوصفي 284
الاختزال 68
الاختيار العشوائي 287
الألوية 1, 6, 7, 11, 61, 127, 190, 202, 232, 477, 580, 652, 654, 681, 694, 695, 698, 699, 700, 702, 707, 717
الأرضية النووية 97
الاركيما 16, 57, 68, 97, 100, 107, 154, 163, 165, 652, 666
الاستبدال 113, 380, 383, 384, 385, 386, 387, 390, 391, 412, 431, 433, 435, 457, 458, 460, 599
الاستبدالات المترادفة 429
الاستعلام 31, 41, 45, 48, 51, 62, 66, 341, 388, 389, 390, 393, 394, 395, 397,

398, 408, 482, 506, 524, 586, 597,
600, 605, 613, 614, 615, 617, 628,
635, 655, 660, 673, 686, 693, 719,
721, 727
403.....الاستقراء الرجعي
610.....الاستقطاب المزدوج
227, 230, 235, 236.....الأسئلة
127, 224, 238, 518.....الاستماتة
118, 131, 213, 220, 223, 224, الإسكات
231, 232, 236, 237, 239, 240, 242,
491
346, 348.....الاشتباك العصبي
569.....الأشرطة المنزلقة
35, 156, 191, 199, 409, 579, الأشعة السينية
584, 601, 605, 614, 619, 639, 661
106, 107, 138, 247.....الأسعة فوق البنفسجية
280.....الأشكال التربيعية
6, 17, 21, 107, 337, 341, 342, الاصطفاف
343, 344, 370, 374, 375, 378, 379,
380, 381, 382, 384, 387, 388, 390,
391, 393, 394, 395, 396, 397, 401,
402, 403, 404, 405, 406, 407, 409,
410, 411, 412, 429, 430, 431, 432,
464, 480, 481, 492, 526, 550, 598,
599, 605, 611, 615, 617, 623, 628,
631, 647, 648, 658, 660, 661, 672,
673, 677, 705, 713, 719, 727, 728
374, 375, 382, 387, 402, الاصطفاف الشامل
403, 405, 430, 481, 727, 728
404, 406, 407.....الاصطفاف المتدرج
374, 375, 382, 390, الاصطفاف الموضعي
403, 406
382.....الاصطفاف شبه الشامل
405.....الاصطفاف شبه المتدرج
658.....الاصطفافات الجاهزة
388, 405, 482.....الاصطفافات الموضعية

546, 548, 696.....الاضطرابات الوراثية
224.....الإطاحة بالجين
43, 51, 205, 468, 516, 568.....الإعلان
725.....الأغذية البحرية
699, 723.....الأغذية الطبية
694, 721, 722, 725.....الأغذية المهندسة وراثيا
69, 70, 78, 177, 184, 203, الأغذية الخلوية
596, 627, 633, 640
113, 115, 118, 233, 420.....الاقتران
168.....الاقتصاد الجيني
1, 46, 68.....الأقسام
708.....الأكزيما
202.....الاكستين
694.....الأكسجة
68, 104.....الأكسدة
507, 516.....الأكسونات الثانوية
504, 510.....الأكسونات الحقيقية
510.....الأكسونات الصحيحة
510.....الأكسونات المحدوسة
509, 510.....الأكسونات المفقودة
220, 468, 493, 500, 613.....الالتباس
89, 562, 579, 582.....الالتفاف الفائق
105.....الألكلة
238.....الليل
375, 449, 450, 451.....الامتالية
190.....الأمراض البروتينية
714.....الأمراض الجلدية
518.....الأمراض المنذلية
46, 77, 101, 140, 227, الأمراض الوراثية
518, 695
706, 707, 708, 718, 723, 725.....الأعضاء
429.....الانتخاب السلبي
239, 416, 460, 494, 512, الانتخاب الطبيعي
563
141.....الانتخاب المتراخي

الانتروبي	21, 185, 186, 187, 200, 518,
	536, 664, 669, 673
الانتروونات	53, 63, 64, 65, 92, 93, 95, 96,
	97, 98, 100, 101, 104, 121, 122, 134,
	136, 137, 148, 161, 166, 167, 168,
	217, 218, 220, 224, 259, 430, 466,
	467, 483, 495, 497, 498, 500, 502,
	503, 511, 516, 524, 573, 598, 671
الانتينيات	202.....
الانجراف المعتدل	653.....
الانجراف الوراثي	512.....
الانحراف	68, 121, 283, 301, 305, 306,
	307, 360, 399, 450, 461, 566, 578,
	584, 622
الانحراف الأقل	450.....
الانحراف المعياري	283, 301, 305, 306, 307,
	360, 399, 566, 578, 584
الانحراف المعياري والتباين	305, 306.....
الانحناء	87, 562, 584.....
الانحنائية	584.....
الإنزيمات	14, 35, 39, 40, 52, 58, 62, 63,
	78, 86, 144, 178, 184, 186, 204, 206,
	231, 233, 234, 237, 243, 253, 254,
	256, 257, 432, 518, 530, 534, 543,
	576, 580, 599, 606, 628, 652, 677,
	683, 684, 686, 688, 694, 698, 701,
	702, 708, 709, 714, 725
الإنزيمات القاطعة	234.....
الإنزيمات الهاضمة للبروتينات	599.....
الإنسان الرقمي	702.....
الأنظمة الثنائية	203, 643.....
الأنظمة الحيوية	5, 63, 68, 72, 74, 127, 156,
	172, 173, 181, 200, 213, 219, 243,
	369, 370, 435, 463, 472, 528, 582,
	597, 689
الانفراجات	435.....
الانقسام الخيطي	70, 71, 146, 224.....
الأنماط السريية	702.....
الأنواع الطبيعية	494.....
الأواصر البيئية	178.....
الأواصر الهيدروجينية	142, 147, 172, 181,
	182, 185, 187, 188, 189, 193, 194,
	195, 197, 200, 582, 610, 663, 665
الأوزان الابتدائية	350.....
الايض الثانوي	112, 113, 683, 686.....
البايولوجي الحاسوبي	680.....
البدء	3, 95, 97, 98, 119, 122, 151, 153,
	169, 208, 213, 227, 265, 294, 314,
	369, 396, 480, 481, 482, 483, 487,
	491, 493, 494, 507, 511, 516, 521,
	524, 525, 534, 538, 541, 544, 547,
	548, 550, 551, 579, 583, 660, 691
البدء الخاطئ	548.....
البدء العشوائي	550.....
البدار	390, 464.....
البرامج	4, 8, 9, 13, 14, 15, 16, 17, 20, 23,
	24, 26, 27, 28, 42, 45, 47, 48, 54, 62,
	66, 67, 89, 154, 183, 206, 222, 260,
	261, 263, 363, 380, 390, 393, 397,
	402, 403, 406, 407, 409, 410, 411,
	412, 445, 461, 462, 465, 468, 473,
	475, 477, 480, 481, 483, 486, 488,
	490, 492, 493, 494, 495, 496, 497,
	501, 502, 506, 507, 508, 510, 511,
	513, 515, 516, 517, 519, 524, 525,
	526, 527, 528, 538, 541, 542, 543,
	548, 549, 550, 551, 552, 562, 563,
	572, 574, 577, 581, 586, 587, 591,
	594, 602, 605, 608, 611, 613, 616,
	618, 619, 620, 622, 626, 630, 631,
	633, 634, 635, 636, 639, 641, 642,
	643, 646, 647, 652, 654, 655, 657,
	658, 659, 660, 661, 662, 663, 665,
	666, 667, 669, 670, 671, 672, 678,

683, 686, 690, 691, 695, 703, 711,
713, 714, 729
البرامج الملونة..... 672.....
البرمجة الدائيمية، 375, 378, 379, 381,
382, 389, 390, 393, 399, 402, 404,
405, 406, 481, 508, 614, 617, 648,
649, 653, 655, 658, 659, 660, 685
البروتين المدمر 236, 237.....
البروتين الوصيف 709.....
البروتينات الأساسية..... 638, 639.....
البروتينات الافتراضية..... 55, 602, 651.....
البروتينات الدهنية 204, 206, 637.....
البروتينات الرايبوزومية..... 161.....
البروتينات السكرية..... 69, 206, 719, 728.....
البروتينات السكرية المحسنة 728.....
البروتينات الغشائية، 119, 170, 188, 189, 190,
201, 203, 206, 261, 583, 638, 639,
640, 641, 711
البروتينات الكروية، 191, 197, 199, 201, 206,
599, 613, 622, 623
البروتينات اللاهستونية..... 211, 229, 236, 237.....
البروتينات المستعرضة للأغشية..... 19.....
البروتينات المطموعة 237.....
البروتينات المفروزة 203, 636, 637.....
البروتينات الملتفة- الملتوية 625.....
البرولين، 174, 175, 178, 181, 198, 237,
383, 429, 610, 611
البريميدات..... 82, 214, 254, 383, 481, 501.....
البريون 190.....
البصمة الوراثية..... 529, 679.....
البقع المتوافقة..... 619.....
البكتريا الحزونية..... 113.....
البكتريا الزراعية 115.....
البكتريا المنشطرة..... 707.....

البلازميدات، 22, 23, 68, 107, 112, 113, 114,
115, 116, 117, 118, 162, 233, 243,
258, 469, 471, 526, 561, 576, 589
البلازميدات الاندماجية..... 115, 116.....
البلازميدات المحركة 114.....
البلازميدات المكوكية..... 116.....
البلازميدات غير الاقترانية..... 114.....
البلاستيدات 46, 69, 203, 263.....
البوداي، 11, 29, 91, 145, 245, 248, 527,
528, 530, 532, 534, 535, 536, 537,
538, 539, 540, 541, 542, 543, 544,
545, 546, 548, 549, 550, 551, 562
البوداي العامة..... 543.....
البيئة الكروموسومية..... 419.....
البيانات الإضافية..... 24, 33, 34, 43, 44, 49.....
البيانات الإضافية المشتقة 34.....
البيانات الجزئية 416, 437, 574.....
البيانات العشوائية 461.....
البيانات المميزة..... 461.....
البيوترات 707.....
التأشب، 113, 117, 127, 134, 141, 142, 143,
257
التباين..... 283, 306, 307, 320, 436, 696.....
التحسس 39, 208.....
التحليل الخطي للتمييز 508.....
التحليل العنقودي 356, 358.....
التحوصل الليفي 190.....
التحويرات بعد الترجمة..... 165, 184, 234.....
التخليق الضوئي..... 68, 69, 463.....
التدريب، 342, 350, 351, 352, 509, 634, 666
الترباط الكامل 364, 365.....
الترباط الوحيد 363, 364, 365.....
التراكيب البروتينية فائقة الالتفاف 624.....
التراكيب الثانوية، 11, 66, 95, 157, 158, 179,
197, 198, 200, 201, 253, 378, 402,
411, 517, 540, 605, 610, 611, 612,

613, 617, 624, 633, 634, 648, 651,
653, 654, 655, 656, 658, 659, 660,
661, 662, 663, 664, 670, 714, 729
التركيب الثانوية الفائقة.....624
التركيب الثانوية لبروتين 66
التركيب الثلاثية 75, 200, 201, 614, 617,
647, 653, 661, 662, 714
التركيب الصليبية 521, 539, 540, 550
التركيب المجسمة 714
التركيب ثلاثية الأبعاد 7, 161, 653, 658
التربسين 17, 138, 262
الترجمة 11, 35, 48, 54, 92, 94, 95, 102,
121, 122, 147, 148, 149, 150, 151,
153, 161, 164, 165, 166, 170, 171,
172, 177, 183, 186, 192, 206, 213,
215, 218, 219, 220, 221, 226, 238,
261, 263, 481, 484, 487, 496, 511,
516, 521, 525, 550, 562, 604, 610,
633, 634, 654, 697, 702, 722
الترجمة المخالفة 164
الترجيحات 5, 349
الترحيل الكهربائي 124, 245, 247, 248, 251,
253, 260, 261, 438, 534, 578, 579,
643
التركيب البنيوي 271
التركيب البنيوي للشبكة 271
التركيب الثلاثي 156, 188, 199, 200, 201,
209, 488, 491, 581, 600, 605, 606,
611, 613, 614, 616, 619, 620, 621,
626, 647, 661
التركيب الحلقي 272, 667
التركيب الداخلي 599
التركيب الرباعي 191, 201, 213, 606, 619,
620, 621
التركيب الشبكي 271
التركيب الشريطي 646

التركيب ألنجمي 271, 272
الترويسة 49
التسوس 709
التشابات 509
التصفية 16, 377, 691
التصنيع الحيوي 7
التصنيف الهرمي 44, 205
التصنيفات 52
التضخيم الكمي 546, 547
التضخيم الكمي الحقيقي 547
التضخيم المتعدد 548
التضخيم الموضعي 546
التضخيم غير المتناظر 546
التطابق 10, 30, 67, 221, 342, 343, 344,
371, 372, 376, 377, 378, 380, 382,
383, 389, 390, 391, 395, 397, 399,
400, 410, 411, 432, 436, 450, 464,
480, 506, 535, 538, 539, 540, 543,
544, 547, 599, 629, 630, 648, 649,
654, 727
التطابق الصارم 629
التطبيب الشخصي 702
التطبع البيئي 130
التطفر النسبي 385
التطفير 104, 118, 429, 543, 614, 664
التطفير الموجه 543
التطفير الموضعي الموجه 614
التطور الأقل 451, 452
التطور الجزيئي 415, 426, 428, 660
التعاقبات 298
التعاونية 490, 491, 659
التعبير البديل 202
التعبير ألبروتيني 100, 172, 261, 262, 619
التعبير الجيني 6, 7, 41, 93, 95, 97, 100,
123, 129, 134, 147, 156, 164, 172,
208, 210, 216, 218, 220, 221, 223,

224, 225, 226, 233, 234, 237, 259,
262, 466, 473, 475, 476, 477, 484,
489, 491, 492, 505, 511, 513, 562,
577, 582, 583, 590, 663, 677, 711
التعريف التصنيفي..... 57
التغاير المشترك..... 653
التغيرات التطوعية..... 113
التفاعلية..... 265, 644
التفرعات الطويلة..... 452
التقنيات الحيوية 699, 636, 584, 202, 63, 9, 7
التكثف..... 186, 190, 261, 362, 363
التماثل الجذري 383, 381, 371, 340, 118,
395, 401, 409, 418, 419, 420, 475,
479, 497, 502, 507, 508, 511, 517,
535, 543, 597, 612, 620, 647, 651,
662, 663, 690
التماثل الجذري البعيد..... 647, 651, 663
التناظر الدخيل..... 419
التنبؤ 98, 65, 63, 62, 39, 25, 18, 12, 11,
121, 183, 189, 213, 215, 227, 262,
284, 353, 370, 409, 412, 426, 452,
462, 463, 479, 483, 484, 486, 488,
489, 490, 491, 495, 500, 502, 503,
505, 506, 508, 509, 515, 518, 522,
523, 525, 540, 592, 598, 604, 605,
614, 618, 622, 627, 635, 642, 646,
647, 678, 679
التنبؤ بالجينات..... 98, 506, 523
التنظيم 140, 100, 95, 94, 93, 86, 70, 39,
156, 164, 168, 192, 197, 207, 208,
209, 210, 211, 213, 214, 215, 216,
217, 218, 221, 225, 226, 227, 233,
234, 236, 239, 463, 465, 484, 486,
491, 492, 493, 498, 506, 517, 522,
523, 526, 527, 577, 579, 582, 583,
592, 627, 652, 665, 680, 697, 701
التميط الجيني..... 162

التميط النووي..... 137, 139, 257
التنوع الحيوي..... 413, 416, 689
التهجين 474, 473, 258, 243, 217, 141,
475, 476, 538, 636, 690
التهميش 205, 111, 47, 52, 35, 34, 11,
567, 572, 709
التهميش اليدوي..... 34
التواءات..... 579
التوائم 703, 542, 539, 538, 535, 528, 37,
707
التوائم العادية..... 37
التوابع..... 107, 123, 139
التوافق..... 465, 508, 545, 613
التوافقية..... 114, 680
التوافق..... 287, 459
التواليات البروتينية 617, 429, 386, 173, 35, 4,
التواليات المرجعية..... 43
التواليات المشفرة 471, 123, 107, 103, 89, 2,
493, 495, 690
التوزيع التجريبي..... 629
التوزيع الطبيعي 324, 318, 317, 301, 300,
325, 388, 477
التوزيع القياسي الطبيعي..... 301
التوزيع الهندسي..... 298
التنبؤات..... 50
الترموديناميكية..... 535, 653, 662
التماللات القطبية..... 185
التماللات المهمة..... 615
الجرذ..... 726
الجرذ النرويجي..... 726
الجزر الثنائية..... 231
الجزر الجيومية..... 556
الجزيرات الحيوية العملاقة..... 1, 5, 35, 172
الجزيرات اللاقطبية..... 187
الجزيرات النانوية..... 172

الجسم المضغوط..... 107
 الجسميات البروتينية..... 184
 الجسور الكبرى..... 184, 200, 201, 206, 213, 261, 619, 623, 633, 634
 الجسم النووي..... 211, 228, 229, 230, 235, 490, 491
 الجمعية الأمريكية..... 56
 الجهاز العصبي..... 137, 232, 345, 347, 509
 الجوار الأقرب..... 364, 659
 الجين المحنوس..... 510
 الجينات الأساسية..... 110, 112
 الجينات الحرجة..... 679
 الجينات الضرورية الظرفية..... 679
 الجينات القافزة..... 224, 416
 الجينات الكابحة للأورام..... 211, 224
 الجينات الكاذبة..... 92, 133, 224, 497, 511, 512, 513
 الجينات الكاذبة الفعالة..... 513
 الجينات الكاذبة المعالجة..... 133, 512
 الجينات الكاذبة المعطلة..... 512
 الجينات الكاذبة غير المعالجة..... 512
 الجينات المتعددة الرسائل..... 515
 الجينات المتناظرة..... 464, 515, 517, 520, 543, 562, 602, 698
 الجينات المشفرة..... 53, 89, 96, 108, 115, 134, 135, 216, 218, 224, 424, 463, 465, 483, 486, 489, 515, 517, 519, 566, 601, 633, 652, 657, 691
 الجينات المهمشة..... 604
 الجينات غير المثالية..... 463
 الجينوم الأدنى..... 519
 الجينوم الأقل..... 464
 الجينوم البشري..... 3, 38, 98, 130, 135, 136, 137, 140, 207, 220, 228, 231, 232, 265, 511, 513, 529, 535, 567, 573, 701, 703, 707

الجينوم البيئي..... 689, 690, 691, 693, 706, 707, 708, 710
 الجينوم ألتجوازي..... 689
 الجينوم المجتمعي..... 689, 692, 705, 710
 الجينوم المركزي..... 110, 111, 112, 115, 463, 553, 557
 الجينوم المركزي الثابت..... 553
 الجينوم غير الأساسي..... 112
 الجينوم غير الضروري..... 111
 الجينومات..... 2, 3, 4, 6, 9, 10, 11, 23, 25, 36, 37, 38, 62, 88, 91, 92, 96, 99, 109, 110, 112, 113, 118, 122, 130, 133, 134, 148, 154, 162, 168, 169, 204, 205, 207, 227, 252, 258, 378, 413, 415, 430, 462, 463, 464, 465, 467, 469, 470, 471, 472, 478, 479, 491, 495, 498, 517, 519, 520, 523, 524, 529, 530, 540, 543, 544, 550, 552, 553, 554, 555, 556, 557, 559, 560, 562, 563, 565, 566, 568, 569, 570, 571, 573, 574, 575, 576, 578, 585, 586, 589, 592, 594, 595, 601, 604, 619, 657, 662, 664, 678, 681, 688, 689, 690, 691, 693, 701, 703, 706, 709, 710, 713, 714
 الجينومات البيئية..... 689, 690, 691
 الحاسوب الشخصية..... 16
 الحرارية الدايناميكية..... 188
 الحساسية..... 48, 190, 341, 377, 389, 390, 409, 509, 510, 584, 586, 618, 672, 707, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729
 الحساسية الأنفية..... 715, 716, 717
 الحشرات..... 140, 228, 231, 238, 239, 470, 716, 721, 724
 الحفظ الحيوي..... 115

الحلزونيات, 82, 192, 193, 194, 197, 199,	الخمائر المتبرعمة 71
206, 214, 581, 582, 610, 612, 613,	الخوارزمية, 14, 356, 382, 403, 436, 459,
624, 625, 626, 641, 642, 646, 658,	460, 493, 657
659, 663, 729	الخيطة 97, 166, 167, 483
الحوام, 643, 720, 727.....	الخيطة البديلة, 100, 134, 149, 166, 168,
الحواسيب, 8, 14, 27, 264, 269, 424, 618,	169, 170, 172, 218, 224, 480, 495,
702	504, 505, 506, 507, 512, 516, 529
الحواسيب الشخصية 264.....	الخيطة المتباعدة 169.....
الحوامض الامينية الحامضية..... 21, 208.....	الدراسات الجينومية التركيبية 6, 463.....
الحوامض الامينية القاعدية 21.....	الدراسات الجينومية الوظيفية 465, 478.....
الحوامض الامينية القطبية 21.....	الدراسات الحاسوبية 12, 44.....
الحوامض الامينية الكارهة للماء... 21, 175, 208...	الدرجات العالية..... 391.....
الحوامض الامينية المحبة للماء 175, 718.....	الدليل الإحصائي 315, 394.....
الحوامض الامينية المشحونة 175, 599, 610.....	الدوريات 13, 562.....
الحوامض الامينية ذات الصفات الخاصة 175.....	الدورية 207, 581.....
الحيوانات المدجنة 141, 714.....	الدومينات, 40, 54, 170, 186, 203, 205, 208,
الخرائط اللاجينية 541.....	239, 387, 401, 479, 491, 493, 518,
الخصائص, 49, 51, 52, 83, 110, 289, 302,	597, 602, 615, 620, 624, 627, 630,
346, 356, 371, 372, 383, 525, 728	631, 650, 654, 713, 727
الخصائص الفيزيوكيميائية 371, 383.....	الرايبوزوم المتعثر 154.....
الخط التطوري 417, 418.....	الرص الحلزوني 582.....
الخلايا الجنينية 71, 232.....	الرقم الهيدروجيني, 43, 73, 175, 178, 260,
الخلايا الصارية 714, 716, 727.....	582, 709, 722
الخلايا القاعدية 716, 724.....	الروابط, 35, 45, 48, 57, 58, 60, 63, 271,
الخلايا المناعية 71, 127.....	272, 601
الخلايا حقيقية النواة, 51, 53, 64, 70, 89, 90,	الزهايمر 37, 137, 528.....
93, 95, 96, 97, 101, 104, 119, 121,	الساعة الجزيئية, 412, 424, 425, 426, 440,
123, 126, 129, 135, 142, 144, 145,	441, 442
147, 148, 149, 151, 165, 166, 168,	السالبة لصبغة كرام, 117, 204, 635, 636, 637,
172, 184, 203, 211, 213, 215, 216,	638, 641
217, 218, 225, 228, 229, 262, 395,	السايتوبلازم, 70, 121, 130, 147, 148, 153,
428, 463, 469, 470, 471, 472, 480,	166, 203, 205, 209, 211, 219, 237,
481, 483, 485, 487, 490, 492, 495,	635
497, 519, 521, 525, 562, 573, 581,	السرطانات, 36, 101, 127, 128, 225, 227,
636, 643, 652	228, 232, 485, 489, 526, 527, 529,
الخلية العصبية 346, 347, 348, 349.....	541, 665, 702, 713

السقط..... 93, 430, 511
 السلاسل الكارهة للماء..... 174
 السمكة الكروية..... 123
 السمكة المخططة..... 38, 698
 السمنة..... 697, 706
 السوائل الجسمية..... 687
 الشبكات الايضية..... 703
 الشبكات الثانوية..... 678
 الشبكات العصبية الاصطناعية..... 345
 الشبكات العصبية البايولوجية..... 345
 الشبكات العصبية الصناعية..... 25, 633
 الشبكة الاندوبلازمية..... 177, 184, 203
 الشبكة العصبية الصناعية..... 509
 الشبكة العنكبوتية..... 3, 265, 266
 الشجرة المجذرة..... 419
 الشجرة غير المجذرة..... 420
 الشجرة متعددة الأفرع..... 419
 الشد السطحي..... 665
 الشرى..... 717
 الشريط الحساس..... 149, 484, 534
 الشريط النقيض..... 149, 242, 534
 الشعيرات الجنسية..... 115
 الشكل التطوري..... 422
 الشيوخة..... 526, 541
 الصبغات..... 138, 250, 251, 260, 475, 476, 543, 547
 الصويا..... 715, 716, 721, 722, 723, 724
 الصيغ..... 14, 15, 17, 19, 20, 23, 26, 41, 42, 44, 47, 49, 384, 423, 487, 536, 549, 600, 646
 الضغط الانتخابي..... 170, 415, 471, 621
 الضغط التنافذي..... 582, 583
 الطاقة الحرة..... 535, 539
 الطبقات المخفية..... 350
 الطبقة المخفية..... 349, 350, 354

الطحالب..... 50, 463
 الطرح الجينومي..... 7
 الطرف الأميني..... 177, 178, 195, 202, 203, 204, 636, 719
 الطرف الكربوكسيلي..... 177, 178, 186, 195, 202, 204, 237
 الطرق الخارجية..... 479, 480, 481, 502
 الطرق الداخلية..... 479, 481
 الطرق السريعة..... 118, 261, 465, 466, 477, 492, 511, 591, 604, 614, 617, 618, 677
 الطرق المتحدة..... 658
 الطرق الموجهة..... 390, 402, 404, 410, 458, 618
 الطرق الهجينة..... 479, 649
 الطفرات..... 7, 36, 79, 92, 93, 101, 102, 103, 104, 107, 113, 121, 122, 137, 141, 151, 161, 190, 211, 217, 225, 233, 238, 240, 377, 384, 385, 402, 406, 412, 414, 416, 421, 424, 425, 426, 432, 433, 436, 453, 461, 489, 490, 506, 512, 518, 529, 562, 583, 605, 652, 653, 658, 660, 672, 678, 680, 691, 695, 696
 الطفرات التعويضية..... 653, 672
 الطفرات المؤذية..... 7
 الطفرات المفيدة..... 113
 الطمغ..... 238, 239, 240, 260
 الطمغ الجزيئي..... 260
 الطمغات..... 55, 238, 239, 466, 477
 طور الاستوائي..... 137
 طور البيني..... 70
 الطوي..... 172, 184, 185, 186, 187, 189, 190, 200, 581, 604, 606, 612, 617, 618, 633, 654, 656, 660, 661, 662
 الطوي التلقائي..... 186

العائيات, 50, 96, 111, 118, 154, 225, 233,	العوامل المساعدة.....602
234, 415, 520, 556, 578, 593	العوامل المطفرة.....104
العامل سيكما..148, 213, 214, 521, 522, 561	الغشاء الخارجي.....204
العرق.....139, 710	الغصون.....416
العروات, 14, 125, 157, 158, 159, 161, 197,	الفترة المغلقة.....290
201, 207, 580, 605, 615, 616, 620,	الفحوص التأكسدية.....729
653, 659	الفرتين.....206
العروات الداخلية.....157	الفردانية.....72, 88, 135
العضيات.....38, 44, 121	الفرضية البديلة.....315, 319, 327
العقدة الداخلية.....424, 436	الفسفرة التأكسدية.....60, 61, 679
العلاج اللاجيني.....228, 232	الفطريات, 50, 100, 131, 220, 233, 238, 573
العلاج باستبدال البروتينات.....190	الفقايع.....539
العلاقات التطورية, 2, 5, 11, 40, 156, 162,	الفهرسة.....34, 49
205, 412, 413, 415, 416, 417, 418,	الفيران, 46, 100, 237, 239, 241, 242, 418,
420, 422, 424, 426, 430, 432, 437,	505, 707, 718, 723
438, 445, 451, 452, 453, 455, 457,	القابلة للزرع.....542, 557, 690, 705, 709
462, 523, 543, 544, 596, 602, 618,	القاعدة الدانمركية.....586, 587, 614
621, 630, 652, 665, 681, 705, 709,	القافزات, 129, 130, 131, 133, 137, 163,
713	166, 225, 556
العلاقات التعايشية.....117	القافرات الارتدادية..130, 133, 137, 166, 225
العلاقات التوافقية.....465	القالب, 107, 144, 260, 529, 535, 536, 537,
العملية الهرمية.....186	538, 539, 540, 545, 547, 548, 614,
العمل.....25	615, 617, 618
العناصر الارتدادية.....129, 133	القضية المحورية.....148, 164, 369
العنقدة, 356, 362, 363, 364, 365, 366, 367,	القطع واللصق.....131
437, 438, 447, 449, 468, 477, 481,	القطيفات, 10, 12, 65, 95, 201, 481, 492,
492, 555, 556, 593, 602, 659, 681	506, 562, 602, 627, 628, 629, 630,
العنقدة الثنائية.....555, 556	634, 636, 653, 657, 659, 709, 727,
العنقدة اللاهرمية.....362, 366, 367	728, 729
العنقدة الهرمية.....362, 366, 367, 438, 477	القفز.....131
العنقود, 356, 363, 364, 365, 366, 384, 439,	القفز الارتدادي.....134, 512
681	القوارض.....50
العوامل الاشتراكية.....209	القوى الكهربائية المستقرة.....182
العوامل الرائدة.....211	القوى المساهمة.....181
العوامل السايكوبلازمية الكامنة.....209	القيم الاحتمالية.....294, 339
العوامل المرافقة.....35, 61, 63, 212	

القيم الإحصائية, 5, 15, 17, 18, 21, 48, 67, 394, 400, 462, 497, 631, 649, 662, 664, 684	الماتيكونندريا, 38, 69, 92, 96, 97, 121, 136, 165, 202, 203, 262, 263, 428, 525, 546, 688, 702
الكالسيوم..... 203.....	المتأريضات 6, 677
الكاميرات 203.....	المتجهات..... 274
الكتلة الإضافية 635.....	المتوسط, 301, 302, 304, 305, 306, 314, 315, 320, 552
الكتب 398.....	المنيل, 36, 106, 184, 230, 231, 232, 233, 234, 239, 240, 490, 529, 540, 541, 551
الكثافة التشفيرية..... 463, 519, 520, 707.....	المنثلية العالية الموضوعية 541
الكرامية للماء 184, 197, 640.....	المجاعة 582
الكربوهيدرات 39, 75, 184, 206, 706, 707....	المجال الجغرافي 266
الكروماتيدات 36, 129.....	المجال المغناطيسي 556
الكروماتين, 93, 97, 138, 139, 211, 213, 216, 219, 220, 223, 224, 226, 227, 228, 229, 230, 233, 234, 237, 491, 492, 526, 562, 578, 591	المجاميع العنقودية 601
الكروماتين الحقيقي 138, 230.....	المجاميع المتباينة..... 602
الكروموسومات, 46, 53, 69, 70, 72, 89, 107, 109, 113, 120, 121, 124, 125, 126, 127, 129, 134, 136, 138, 139, 140, 141, 142, 145, 146, 211, 232, 233, 238, 239, 240, 241, 242, 256, 378, 416, 463, 472, 489, 491, 493, 512, 525, 574	المجاميع المتوافقة..... 114
الكلابكان..... 707, 719, 728.....	المجاميع غير المتوافقة..... 114
الكلمات المفتاحية, 44, 46, 49, 51, 52, 64, 65, 600, 635, 644	المجتمعات الميكروبية..... 703, 707
الكمية العددية..... 280.....	المجسات, 243, 258, 259, 475, 478, 530, 546
الكواذب 253, 602.....	المجمعات الوراثية..... 710
الكوزميدات 258, 690.....	المجموعة الخارجية 420, 421, 422
الكيموحيوية..... 574, 630, 687.....	المجموعة الداخلية 420
اللاصقات 234, 637.....	المحاصيل... 7, 38, 686, 722, 723, 724, 725
اللاقطبية..... 187, 581, 623.....	المحاصيل المهندسة..... 722, 723
اللبائن, 88, 97, 130, 134, 169, 219, 223, 227, 231, 232, 237, 238, 239, 240, 420, 488, 515, 526, 540, 705	المحاكاة الحاسوبية..... 1
	المحاكاة الحيوية 1
	المحس, 714, 716, 717, 720, 722, 723, 729
	المحسسات المخفية..... 723
	المحور العصبي 347
	المخرجات, 14, 15, 26, 67, 349, 351, 353, 410, 462, 464, 602, 630, 662, 667, 674, 727

المدخلات, 13, 27, 30, 53, 347, 348, 349, 350, 351, 370, 658
 المدى..... 189, 305, 348, 412, 502, 722
 المراجع, 12, 26, 34, 35, 37, 45, 46, 47, 51, 52, 53, 55, 56, 59, 62, 63, 412, 494,
 552, 566, 602, 605, 617, 630, 682,
 686, 694, 695, 697, 698, 701, 709,
 727, 728
 المراسي..... 178
 المراوغة..... 435
 المرحلات الابتدائية..... 350
 المرحلة بعد الجينوم..... 688
 المركز, 28, 30, 35, 40, 41, 42, 47, 49, 55, 56, 66, 138, 139, 228, 302, 303, 305,
 407, 468, 563, 572, 586, 594, 659,
 713
 المركز الأوروبي, 28, 35, 41, 47, 55, 66, 572, 586, 713
 المراكز..... 69, 93
 المزامرة..... 186
 المساحة السطحية المتاحة.. 183, 611, 633, 647
 المسار الأيضي..... 5, 62, 170, 681
 المسارات الأيضية الميكروبية..... 40
 المسافات الوراثية..... 432, 441
 المسافة الأقلية..... 357, 358
 المستضدات..... 78, 203, 643, 713, 714
 المسكتات..... 486, 506
 المشجعات, 93, 208, 209, 210, 215, 231, 486, 487, 491
 المصفوفات, 40, 221, 223, 260, 261, 276, 277, 280, 281, 332, 333, 370, 383,
 384, 386, 398, 432, 475, 477, 480,
 488, 504, 542, 586, 591, 604, 629,
 649, 655, 663, 693, 695, 719
 المصفوفات الدقيقة, 40, 221, 223, 260, 475, 477, 504, 542, 586, 591, 663

المصفوفات النقطية..... 398
 المصفوفة الانتقالية..... 330
 المصفوفة الصفيرية..... 276
 المصفوفة القطرية..... 276
 المصفوفة المتماثلة..... 276
 المصفوفة المحايدة..... 276, 279, 280
 المطفرات..... 104, 129
 المعالجة الحيوية..... 117, 693
 المعايير..... 247, 367, 389, 412, 430, 550
 المعدل الرياضي غير المرجح..... 440
 المعلوماتية الحيوية السريرية..... 696
 المعلوماتية الكيماوية..... 699
 المعلوماتية المناعية..... 711
 المعنوية الإحصائية..... 372, 388, 394, 400
 المفتاح الرئيس..... 31, 49
 المقاومة, 117, 129, 130, 162, 244, 415, 696, 724, 725
 المقدمات الجينومية..... 545
 المقروءة..... 252
 المقياس الداينميكي..... 578
 المكررات, 11, 15, 21, 29, 48, 107, 119, 123, 124, 129, 137, 223, 233, 253,
 377, 378, 390, 395, 396, 463, 467,
 470, 471, 472, 520, 521, 529, 568,
 569, 570, 571, 573, 577, 585, 586,
 620, 726
 المكررات الكثيرة..... 124, 395, 463
 المكررات المعتدلة..... 123
 المكررات المقلوبة..... 472, 569
 المكررات المنتشرة..... 123
 المكررات الموضوعية..... 521
 المكون البروتيني, 172, 216, 220, 260, 574, 600, 643, 719
 المكون الميكروبي المركزي..... 703, 709
 المكوثر السداسي, 480, 497, 505, 507, 509, 524, 598

المكوثرات السداسية المشفرة.....	508	النسخ الارتدادية.....	134
المكوثرات العشرية.....	505	النسخ العكسي, 130, 133, 134, 244, 258, 259,	466, 504, 545, 546
الملف البسيط.....	35	النسخ ثم اللصق.....	131
المملكة الحيوانية.....	69, 231, 430	النشوء المتجانس.....	431, 452
المملكة النباتية.....	69	النظام البيئي الميكروبي.....	708
الممهد, 93, 161, 208, 213, 214, 215, 219,		النقل الأفقي.....	118, 424, 463, 471, 523
226, 227, 484, 485, 486, 488, 489,		النماذج البيئية.....	44, 595, 691, 692, 705
490, 492, 493, 494, 495, 498, 521,		النمط البيئي.....	681
522, 566, 579, 663		النمط المظهري, 93, 101, 140, 227, 238,	478, 518, 605, 678, 680, 681, 695
الممهد الأقرب.....	487	النهايات المتدلية.....	656
الممهد البعيد.....	487	النهاية الكربوكسيلية.....	164, 177, 355, 727
الممهدات الدنيا.....	487	النهج التوجيهي.....	406
الممهدات القوية.....	149	النواة, 68, 69, 70, 71, 91, 94, 95, 97, 100,	101, 119, 121, 122, 124, 130, 135,
المناطق البينية.....	493, 517, 561, 579, 583	146, 148, 150, 153, 166, 169, 197,	202, 209, 210, 211, 213, 215, 219,
المناطق الجسمية الكاذبة.....	242	220, 226, 227, 236, 237, 253, 263,	415, 463, 479, 485, 486, 487, 488,
المناطق الساخنة للتنوع.....	689	489, 490, 491, 494, 495, 496, 497,	498, 502, 503, 506, 508, 509, 510,
المناطق الغنية بـCpG.....	37	512, 515, 517, 519, 520, 528, 543,	546, 553, 569, 573, 583, 621, 634,
المناطق المشوشة.....	475	666, 703	
المنظر, 63, 92, 186, 285, 339, 352, 430,		النوافذ المنزلقة.....	470
450		النواقل.....	244, 258, 467, 528
المناعة الفطرية.....	710, 713	النوع, 27, 30, 32, 41, 61, 62, 85, 86, 95,	
المنظومات الحيوية.....	694	96, 110, 111, 112, 113, 121, 132,	
المنوال.....	304	137, 139, 140, 141, 162, 163, 181,	
المهامة, 10, 11, 13, 19, 23, 65, 207, 350,		182, 197, 202, 204, 207, 216, 219,	
369, 401, 498, 503, 507, 510, 649,		234, 247, 257, 267, 271, 285, 286,	
720		295, 298, 312, 314, 315, 316, 318,	
الموائمة.....	48, 370, 606, 630	319, 320, 321, 328, 332, 333, 334,	
المواد العصبية.....	63, 115, 117	367, 370, 381, 382, 384, 389, 390,	
المواصفات الداخلية.....	647, 648, 649	402, 422, 424, 475, 481, 485, 487,	
المواقع الجغرافية.....	139	497, 517, 520, 528, 557, 561, 574,	
الموجبية.....	445		
الموجبات.....	623		
الميكانيك الكمي.....	581		
النافذة المتحركة.....	377		
النحفاء.....	707		
النسائل.....	129, 243, 244, 695		

579, 593, 594, 635, 685, 689, 698,
704, 705, 716, 721
121, 223.....النوية
248.....النيوكلوتيد الشاذ
714, 716, 720, 727.....الهستامين
37, 209, 211, 213, 215, 219, 226, 228, 229, 230, 234, 235, 236,
237, 238, 239, 242, 490, 491, 493,
582, 584, 591
349.....الهندسة المعمارية للشبكة
116, 117, 118, 214, 541, 690, 721, 724
الواسمات, 93, 232, 238, 243, 428, 465, 546, 573, 578, 686, 687
428.....الواسمات الجزئية
686, 687.....الواسمات الحيوية
417, 418, 420, 422, 423, 424, 426, 427, 428, 442, 445, 447,
452, 453, 458, 459, 461, 705, 709
110, 417.....الوحدة التصنيفية
652.....الوراثة الجديدة
36, 89, 227, 228, 229, 230, 236, 238, 492, 493, 526, 527, 528,
529, 551
527.....الوراثة اللاجينية الحاسوبية
238.....الوراثة المنديلية
303, 304, 469.....الوسيط
186, 189, 190, 596.....الوصيفات
190.....الوصيفات الصيدلانية
573.....الوضوح الواطئ
172, 206.....الوظائف الميكانيكية
468.....الوهمية
أ
127, 228, 720.....أمراض المناعة الذاتية

أ
416.....انتخاب المجتمعات الثانوية
218.....انتخاب ايجابي
113.....انتخاب دارون
577.....انحراف محتوى القواعد
153.....انحياز الشفرات
232, 233, 541.....انخفاض المثيلة
541.....انخفاض المثيلة العامة
إ
144, 216, 259, 530, 533, 551.....إنزيم الكوثرية
128, 133, 259, 545.....إنزيم النسخ العكسي
217.....إنزيمات إزالة الأدينين
234.....إنزيمات المثيلة اليتيمة
113, 125, 137, 142, 163, 243, 257, 432, 545
113, 125, 142.....إنزيمات القطع الخارجي
137, 142, 163, 243, 545.....إنزيمات القطع الداخلي
أ
264, 269.....أنظمة التشغيل
234.....أنظمة القطع والتحويل
أ
697.....انقسام الشخصية
70, 71, 72.....انقسام السايكوبلازم
70.....انقسام النواة
140.....انقسام خيطي داخلي
101, 108, 115, 148, 160, 162, 163, 465, 523, 577
160, 163, 215, 520, 577, 592, 652.....اوبرونات

إ
إيجاد الجين, 479, 495, 498, 508, 515, 519,
524, 525
إيجاد الممهدات, 484, 490, 492, 493, 494,
524

ب

ببتيد الإفراز 635
ببتيدات الإشارة 203
بدء التضاعف .. 92, 121, 144, 560, 578, 585
بدائية النواة, 16, 21, 27, 37, 60, 68, 69, 89,
92, 93, 94, 95, 98, 101, 109, 110,
118, 122, 123, 124, 148, 149, 150,
151, 153, 166, 190, 203, 213, 214,
215, 217, 225, 226, 263, 463, 465,
470, 479, 482, 483, 484, 485, 487,
508, 515, 519, 521, 526, 553, 573,
579, 581, 582, 621, 637, 638, 652,
678

برامج التنبؤ, 4, 14, 159, 173, 177, 489, 506,
524

برامج المحاكاة 63
برامج وأنظمة الاسترداد 45
برنامج الجوار الأقرب 705
برنامج الشعار, 20, 22, 522, 630, 631, 668,
671

برنامج الشعار الصوري 20
بروتوكول نقل الملفات 265
بروتينات 164, 263
بروتين رابط 487
بروتينات الصدمة الحرارية 189, 210
بروميد الاثديوم 534
بصمة الببتيد 262, 635
بكتريا اللبن 151
بلازميد الخصوبة 115, 116
بلازميد المقاومة 115

بلازميدات اقترانية 113
بلازميدات الخمائر 118
بلازميدات الضراوة 115
بلازميدات مفككة 115
بنك الأدوية 687, 699, 700
بنك الجينات, 15, 37, 41, 45, 49, 50, 51, 52,
53, 64, 65, 66, 466, 468, 469, 542,
557, 562, 572, 574, 586, 587, 666,
682, 686

بواى الاكسونات 541
بواى التضخيم 552
بيئة المستخدم التصويري 264
بيئة المستخدم التصويرية 264
بيانات إضافية 26, 33, 34, 43
بيروكسيد الهيدروجين 583
بيورينية 82, 105, 383

ت

تبادل كروماتيدي شقيقى 129
تباين المعدل الموضعي 436
تباينية الاستبدال 435
تجاوز الاكسون 167, 169, 504
تجاوز الاكسونات 169, 504
تحديد النسق 367
تحسس الزحام 690
تحليل الاصطفافات 66
تحليل البيانات 369
تحليل المتغيرات المتعددة 703
تحليلات الحاسوب 43, 53, 59, 65
تحوير الهستونات 234, 236, 237
تحويرات الكروماتين 228
تحويرات بعد الترجمة 184, 206, 212
تخليق البروتينات, 69, 92, 156, 172, 202,
217, 635
تداخل الأدوية 698, 700

تداخل البروتينات الأصلي 170
تداخل الجينات 515
تدريب الشبكة 352
تذبذب 107, 351, 352, 585
تراكيب البروتين الثانوية 182
تراكيب ثانوية فائقة 159
ترتيب الجينات 464, 465, 524, 621, 691
ترجمة شاذة 164
تركيب أساسي مشترك 650
تركيب الحافلة 271, 272
تركيب العيدان والكرات 644, 645
تركيب بوقي 157
تصميم الأدوية الشخصية 702
تصميم البودائ 11, 12, 91, 402, 528, 535, 538, 540, 542, 548, 549, 550, 551
تصنيف البروتينات 40, 205, 602, 606, 621, 627, 649
تضخيم الانتروبي 628
تضخيم الطفرات 540
تضخيم العقدة 546, 547
تضخيم المسافات الطويلة 548
تطابق زائف 395
تطور الجينومات 224, 227
تطور ايجابي 429
تعطيل للكروموسوم الأنثوي 241
تغاير الأطوار 521
تغايرات النيوكليوتيد المفرد 12, 36, 46, 140, 228, 256, 400, 517, 541, 605
تفاعل التضخيم 539
تفضيل الشفرات 542
تقفي الأثر التطوري 480
تقنيات المصفوفات الدقيقة 259
تقنيات النقل 268
تقوس DNA 578
تقوس أشربة 21

تقييم أداء البرامج 671
تقييم الشك 284
تماثل التوالي 371, 372, 385, 602, 614
تمثيل افتراضي 678
تنظيم الترجمة 215
تنظيم الجينات 147, 165, 166, 210, 236, 513, 526
توافق 41, 220, 437, 457, 464, 507
توالي الاستعلام 11, 48, 374, 382, 388, 389, 390, 394, 395, 396, 397, 400, 479, 480, 508, 509, 513, 605, 606, 611, 613, 617, 629, 636, 642, 655
توالي الإشارة 203, 204, 636
توالي كوزاك 507, 521
تواليات الإشارة 263, 726
تواليات الإقحام 130, 416
تواليات المرأة 472
تواليات صناعية 50
تواليات مكررة 119, 230, 495
توجيه البروتينات 184, 203, 262, 263
توحد 17, 459
توحيد التسميات 513
توحيد تسمية الجينات 24
توزيع بواسون 296
توزيع كمبيل 395
توليد الحساسية 718, 719, 720, 722
توهين 165

ث

ثملات 17, 52, 53, 106, 158, 177, 183, 184, 197, 207, 229, 235, 236, 237, 334, 355, 370, 371, 381, 382, 383, 385, 390, 430, 548, 562, 620, 622, 624, 629, 634, 636, 637, 648, 663, 665
ثملات العوامل المرافقة 430

178, 182..... ثنائية القطب

ج

جبر المصفوفات 274, 277
جرع الجين 117, 140, 241, 242
جزاء فتح فجوة 381
جزاءات الفجوات المرتبطة 381
جزاءات الفجوة 380
جزر الامراضية 556, 593
جزر التعايش 556
جزر ايض الكبريت 556
جزئيئات التداخل الصغيرة 166
جسيم الخياطة 167, 501, 503
جنون البقر 190, 604
جهة اليمين 85, 148, 485
جهد الكهربائية المستقرة 722
جين الإعلان 261
جينات الإدامة 213, 488, 547
جينات الظل 525
جينات صامتة 695
جينات غريبة 424
جينات كاذبة 512, 513
جينات متباينة الأصول 489
جينوم مرجعي 594, 595, 693
جينوم مركزي 110
جينومات 7, 37, 65, 96, 97, 104, 107, 108, 109, 111, 112, 119, 121, 122, 123, 129, 130, 131, 135, 141, 147, 154, 163, 166, 217, 231, 234, 239, 382, 395, 462, 463, 464, 469, 470, 471, 497, 510, 512, 517, 519, 520, 521, 526, 528, 540, 543, 553, 554, 555, 557, 558, 562, 565, 567, 571, 572, 573, 574, 582, 587, 621, 665, 689, 690, 693
جينومات المضيف 7

جينومات بدائية النواة 107, 119, 122

ح

حب الشباب 708
حجم الانتروونات 65
حجم الجينومات 463, 469, 470
حد الإشباع 434, 436
حد العتبة 347, 348, 349, 483, 684
حدثة قواعد البيانات 65
حدس 4, 471, 479, 486, 489, 490, 491, 493, 494, 495, 501, 502, 505, 506, 509, 511, 512, 513, 515, 522, 569, 604, 605, 606, 609, 610, 611, 612, 613, 615, 617, 618, 619, 620, 621, 622, 623, 624, 628, 633, 634, 635, 636, 639, 641, 642, 643, 647, 651, 653, 654, 657, 658, 659, 660, 661, 662, 663, 664, 665, 667, 695, 700, 722, 728
حدس التركيب الثانوي 610, 611
حدس التركيب المتماثل 617, 618
حدس الحلزونات 641
حرارة الالتحام 535, 537, 542, 550
حرارة انصهار 536
حركات التحلق 579
حساسية 261, 389, 398, 429, 445, 486, 490, 504, 545, 546, 617, 627, 690, 716, 722, 729
حشرات 151, 724
حقيقية النواة 16, 27, 37, 60, 66, 68, 69, 91, 94, 98, 107, 113, 122, 123, 127, 130, 131, 138, 142, 144, 145, 165, 168, 215, 219, 226, 230, 236, 263, 413, 425, 430, 463, 472, 479, 482, 483, 484, 486, 490, 495, 496, 507, 551, 553, 556, 573, 597, 621, 634, 665, 666

حلزونات ألفا, 186, 192, 193, 194, 196, 197, 263, 261, 207, 206, 199, 198, 624, 613, 611, 610, 606, 562, 642, 641, 639, 638, 627, 643
حلزونات سالبة الالتفاف.....582

خ

خارج الأنظمة الحية, 186, 217, 245, 248, 491, 604, 512, 523, 528, 543
خرائط التقييد.....253
خلية عصبية اصطناعية.....347
خميرة الخبز, 100, 118, 127, 130, 153, 517, 573
خوارزميات العنقدة.....362
خوارزمية الجوار الأقرب.....621

د

داء السكري.....137, 697
داء الصدفية الجلدي.....697
دالة التجزئة.....655, 660
دالة التفعيل.....348, 349, 350, 353
دالة التنشيط.....353
دالة الدرجة.....403
دالة السكمويد.....353
دالة السكواش.....353
دالة الكثافة الاحتمالية.....301
دالة الكشف.....397
داينميكية.....5, 68, 238, 239, 424, 585, 653
دبوس الشعر.....197
دخيلة التشابه.....419
دراسات الجينوم المقارنة.....463, 465
دراسة الجينومات المقارنة.....205
دراسة الغصون.....418
دراسة مكنون الأيضي.....681
درجات التشابه, 388, 399, 401, 405, 599, 626

درجات التطابق.....379, 390, 396, 399, 551
درجات التوافق.....664
درجات الكراهية للماء.....641
درجات اللوغاريتم الشاذة.....384
درجة الشك.....395, 397
دوال التفعيل اللاخطية.....348
دورية القطيفات.....562

ذ

ذاكرة الخلية.....233
ذباب النوم.....520
ذباب الفاكهة.....38, 97, 122, 166, 168, 682
ذيل الأدنين, 95, 101, 149, 168, 170, 217, 221, 259, 466, 497, 503, 512
ذيول الهستونات.....230, 235, 236

ر

ربائط خارجية.....209
رسم الأثر.....584
رسم الجبل.....669
رسم الصندوق.....552
رسم نقطي.....464
رشاد الصخر, 38, 64, 65, 96, 98, 130, 227, 440, 441, 454, 455, 515, 516, 517
رقائق الجين.....259, 473
رقم التسجيل.....16, 39, 46, 49, 51, 52, 667
رموز المجموعة.....284

ز

زاوية الالتواء.....584
زامم الليوسين المنزلق.....627, 628

س

سلاسل ماركوف.....328, 332
سلسلة ماركوف.....328, 330, 331, 333, 341

724..... Cry سموم
489..... سيطرة رجعية
213, 214, 485, 487, 521, 522, 561... سيكما

ش

شبكات 15, 111, 213, 214, 215, 218, 219, 226, 266, 267, 345, 369, 489, 516, 652, 677
شبكات الايض..... 677
شبكات المنطقة الشخصية..... 266
شبكات المنطقة الواسعة..... 267
شبكات خلوية..... 215
شبكات منطقة العاصمة..... 267
شبكة الانترنت..... 2
شبكة ثانوية..... 683
شبيه بالهستونات..... 213
شجرة الجينات..... 416, 424
شجرة الدليل..... 405, 438
شخصنة الأدوية..... 695
شفرة المصادقية..... 59, 65
شفرة الهستون..... 235
شفرة الوراثة اللاجينية..... 232
شفرة الوقف 98, 103, 148, 153, 154, 483, 501, 565
شفرة مقابلة..... 155
شلالات الإشارة..... 489
شوكة التضاعف 120, 145

ص

صبغات متقلورة..... 248, 260
صبغة كمزا..... 138
صف التواليات 9, 11, 375, 401, 402, 405, 406, 611, 648
صفائح بيتا 194, 196, 197, 199, 261, 608, 610, 623, 638, 646
صهر الحزون 581

صينغ الملفات..... 23, 26, 27, 549
صيغة التوالي العادية..... 26

ض

ضم الجوار 441, 442, 445, 446
ضم الجوار المرجح 445

ط

طاقات الارتباط..... 494
طاقة الإدانة..... 68
طاقة الرص 21, 536, 557, 577, 581, 585, 586
طبوغرافية 145, 419, 422, 580, 618, 639, 682
طبوغرافية المسارات..... 682
طرق الانتقال الأفقي..... 118
طرق التنبؤ..... 479, 605
طرق الحاسوب..... 1, 502, 614, 678
طرق العنقدة..... 437
طرق الكرة الجديدة..... 481
طريقة إرساء البروتينات..... 621
طريقة الاصطفاف التكراري..... 406
طريقة التجذير الوسطي..... 420
طريقة التحزيم..... 138
طريقة التنشيطية..... 252, 253, 467, 472, 524
طريقة الجوار الأبعد..... 705
طريقة القطع والتسلسل..... 407, 408
طريقة الكلمة الموجهة..... 390
طريقة المركز 363, 365, 366
طريقة المعدل الرياضي 405, 406, 438, 440, 442
طريقة المعيار الأمثل..... 367
طريقة المعيار التسلسلي..... 366
طريقة المعيار المتوازي..... 366
طريقة رسم ملئ الفراغ..... 645

طريقة ورود 365
 طمغات 168, 466
 طمغات التواليات المعبر عنها 466
 طوي البروتينات, 185, 184, 183, 182, 179, 5
 604, 596, 190, 189, 188, 187, 186,
 663, 651, 618, 612

ع

عامل الصدمة الحرارية 210
 عامل سيكما 214, 521
 عدد الجينات, 218, 136, 121, 99, 88, 64
 569, 567, 553, 516, 511, 469, 463
 عدم البدء 551, 538
 عروات, 197, 161, 158, 157, 124, 113, 97
 641, 491, 215
 عروات متعددة الأذرع 157
 عروة الإزاحة 125
 عروة مضادة الشفرة 159
 علاقة تطويرية كاذبة 452
 علم التطور الجزيئي 416
 علم الحياة الصناعي 462
 علم المناعة الحاسوبي 711
 عمر الخلية 126
 عمليات إظهار 410
 عمليات الاصطفاف, 378, 370, 205, 40, 17
 393, 391, 388, 387, 382, 381, 380,
 464, 430, 429, 412, 409, 406, 403,
 550, 542, 505, 502, 498, 497, 492,
 628, 615, 612, 605, 602, 566, 551,
 682, 681, 680, 662, 661, 657, 649,
 727, 726, 719, 684, 683
 عمليات الانزلاق 521
 عمليات البدء الكاذبة 551
 عمليات التصحيح 172
 عمليات التصفية 550, 516, 65, 64
 عمليات الزرع 704, 689

عمليات طوي .. 604, 527, 186, 184, 177, 40
 عملية الحجب 395
 عناصر التنظيم, 524, 522, 492, 489, 486, 10
 عناقيد الجينات المتناظرة 38
 عوائل البروتينات 620, 602, 205, 40
 عوائل كبيرة 650, 602, 209, 205
 عوازل الدومينات 493
 عوامل الإطلاق 155, 154
 عوامل الانتساخ, 210, 209, 208, 207, 10
 227, 226, 219, 216, 215, 212, 211,
 485, 484, 481, 261, 236, 235, 231,
 522, 494, 492, 491, 489, 487,
 699, 580

عوامل الانتساخ العامة 227, 210
 عوامل البدء 153
 عوامل الرخصة 146
 عوامل الضراوة 115, 112

غ

غطاء 149

ف

فحص الانحياز 461
 فحص وخز الجلد 729
 فحوص إثارة الحساسية 725
 فرضيات الاختبار 314
 فرضية العدم, 320, 319, 317, 316, 315, 314
 327, 326, 325, 324, 323, 321
 فستق الحقل 724, 716
 فسحة محيطية 204
 فسفرة الهستونات 236
 فقر الدم المنجلي 597
 فلق, 497, 481, 247, 184, 166, 101, 96, 95
 725, 636, 516

ق

- قاعدة بيانات الحوالم 713
- قاعدة صنع القرار 316, 317, 318
- قاعدة كرامر 281
- قطع او كازاكي 120, 126, 144
- قطيقات الانفلاق 65
- قطيقات تركيبيية 197, 201, 653
- قطيقة متوافقة 263
- قناة المعلومات الرئيسة 265
- قواعد البيانات, 2, 3, 4, 5, 7, 11, 14, 15, 23, 24, 25, 26, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 52, 54, 55, 56, 57, 58, 59, 60, 62, 63, 66, 205, 262, 264, 265, 340, 341, 370, 374, 382, 387, 388, 389, 390, 393, 395, 396, 397, 398, 400, 402, 460, 466, 467, 468, 469, 470, 479, 480, 481, 482, 493, 494, 497, 506, 508, 513, 514, 515, 516, 517, 519, 542, 563, 568, 569, 574, 576, 586, 592, 597, 599, 600, 602, 604, 605, 613, 614, 615, 617, 618, 619, 620, 621, 622, 628, 629, 630, 636, 643, 644, 650, 651, 657, 664, 666, 677, 682, 684, 685, 686, 687, 688, 690, 691, 693, 698, 699, 700, 708, 712, 713, 719, 725, 726, 727, 728
- قواعد البيانات الثانوية 41, 46
- قواعد البيانات الحيوية 41, 58
- قواعد البيانات الشخصية 41
- قواعد البيانات العلاقية 23, 31, 32
- قواعد التصنيف 56
- قواعد بيانات الايض 58, 677, 682, 686
- قواعد بيانات الايض والانزيمات 58

- قواعد بيانات البروتينات, 54, 400, 468, 483, 600, 606
- قواعد بيانات الترحيل الكهربائي 41
- قواعد بيانات المسارات الايضية 59
- قواعد بيانات أولية 41
- قواعد بيانات فرعية 65
- قواعد للكلوبيولينات 713
- قوى فاندرفال 182
- قيم ترجيحية 406, 440
- قيم ثقة 477, 516
- قيم معيارية 359
- قيمة P 323, 324

ك

- كائن خارق 703, 707
- كاشف البقع 644
- كتلة الجسم 706
- كثافة العقدة 437
- كروماتين متباين 239
- كشافات الجينات 468
- كلونة DNA 243
- كماشة GC 538
- كيمورا 434, 435, 436

ل

- لذابة الفاكهة 3
- لعبة الأحجار المقطعة 472
- لغات البرمجة 3, 9, 14, 32
- لهجة وراثية 692

م

- ماركوف, 328, 329, 331, 332, 341, 342, 343, 344, 345, 495, 507, 508, 636, 642, 720
- ماشات الشعر المتقابلة 158

113, 157, 220, 223, 475, 521, 525, 538, 539, 659	641..... محبة الدهون
484, 486..... ماكينة الانتساخ	174, 722, 725, 727 محبة الماء
90, 163, 562..... مبدأ المراوغة	109, 559 محتوى الجينومات من AT
274, 275..... مبدلة المصفوفة	127 محدودية هايفليك
723..... مبيدات الأدغال	17, 410, 442 مخرجات البرامج
690..... متأبطات	179, 194, 646 مخطط راماجاندران
257, 518..... متباين الزيجة	727 مخططات محبة الماء
257, 518..... متجانس الزيجة	724 مدة الصلاحية
10, 467, 468, 691..... متجاورات	396 مدرج لوني
468..... متجاورات تخمينية	269, 270, 272, 273 مدى التردد
274, 275, 280, 282, 283, 348, 349, 351	534 مرحلة الاستقرار
متجهات المتوسط	534 مرحلة التباطؤ
416..... متحجرات جزئية	151 مرض النوم
68, 107..... متضاعفات	716 مزدوج متجانس
178..... متعادل الحموضة	106, 156, 219, 220, 230, 236, 488, 534, 535, 538, 539, 540, 546, 547, 548, 550, 581, 582, 583, 584
140, 225, 485, 495, 716, 722..... متلازمة	534, 547 مزدوجات البادئ
140..... متلازمة داون	106, 236 مزدوجات التايمين
419, 421..... متناظرة التشابه	535 مزدوجات القالب
63..... مثالب	642 مزدوجة الصفات
36, 37, 86, 226, 230, 231, 232, 233, 235, 236, 238, 239, 242, 490, 492, 527, 528, 529, 541	25, 518, 528, 613, 652 مزودات خدمة
36, 37, 230, 231, 232, 233, DNA مثيلة	518 مزودات خدمة إضافية
236, 239, 242, 490, 492, 527, 528, 529, 541	39, 681, 687 مسارات نقل الإشارات
231, 233..... مثيلة الأدينين	723 مساعدات المحسسات
226, 235..... مثيلة الهستونات	208 مساعدات تنظيم الانتساخ
282..... مجتمع متعدد المتغيرات	180, 181 مستوى الأصرة
268..... مجمع الرسائل	213, 216 مستويات تنظيم الانتساخ
403, 406..... مجموع الأزواج	189 مسخ البروتينات
173..... مجموعة الأمين	576 مشاريع الجينومات البيئية
418..... مجموعة التطور النظرية	665 مشجعات الاكسونات
173, 178, 206..... مجموعة الكربوكسيل	687 مشروع الايض البشري
265, 266, 267, 270..... مجهز الخدمة	228, 687 مشروع الجينوم البشري
	228 مشروع الجينوم البشري اللاجيني
	703 مشروع المكنون الميكروبي

مشروع المكنون الوظيفي.....	702
مصادقية.....	63, 65, 66, 388, 461, 516, 551
مصفوفات استبدال.....	384, 405, 407, 519
مصفوفات الترتيب الموضوعية.....	506
مصفوفات اللون.....	19
مصفوفات المسافات.....	431
مصفوفات النسق.....	628, 629
مصفوفات درجات المواقع المحددة.....	333, 342
مصفوفة.....	274, 276, 277, 279, 280, 283, 340, 329, 331, 333, 334, 338, 339, 340, 341, 358, 375, 376, 378, 385, 386, 387, 390, 391, 392, 399, 402, 403, 405, 437, 438, 439, 445, 446, 448, 449, 450, 452, 655, 656, 658
مصفوفة الارتباط.....	283
مصفوفة التثقيط.....	375, 376, 378, 390, 655, 656
مضاهايات قواعد النيوكليوتيدات.....	106
مطياف الرنين النووي المغناطيسي.....	191
معادلة والاس.....	535, 536, 551
معالجة الصور.....	476
معامل ارتباط بيرسون.....	308, 320
معامل ارتباط سبرمان.....	312
معامل الارتباط.....	282, 308, 309, 310, 312, 313, 357, 509, 622, 672
معامل الارتباط التقريبي.....	509
معامل الارتباط اللامعلمي.....	312
معامل الارتباط بيرسون.....	313
معدل الترابط.....	363, 365
معدل ترسيب.....	167
معضلة الفرع الطويل.....	452, 453
معطي الانفلاق.....	166, 496
معقد النقب النووي.....	219
معكوس المصفوفة.....	279
معنوية الاختبارات الإحصائية.....	314
مفاهيم إحصائية.....	284
مفكك الرسائل.....	268
مقارنة الجينومات.....	505, 552, 553, 556, 557, 572, 588, 593
مقارنة النيوكليوتيدات الرباعية.....	557
مقاييس الاختلاف.....	305
مقاييس النزعة المركزية.....	301, 305
مقاييس كره الماء.....	623
مقبرة الخلية.....	184
مقياس التداخل.....	610
مقياس التشابه.....	357
مقياس سبرمان راو.....	312
مكتبات DNA.....	257, 258
مكررات التواليات المترادفة.....	544
مكررات الكروموسوم.....	123
مكررات النسخة الواحدة.....	123
مكنون RNA.....	216
مكنون الإفراز.....	635
مكنون الإنسان الميكروبي.....	693, 703, 708, 710
مكنون الايض.....	6, 574, 681, 686
مكنون البروتينات المفسفرة.....	237
مكنون الجلد الميكروبي.....	708
مكنون العناصر المتحركة.....	113
مكنون الفم الميكروبي.....	708
مكنون المجترات الميكروبي.....	710
مكنون المهبل الميكروبي.....	709
مكنون النسخ.....	6, 148, 156, 164, 172, 219, 220, 224, 259, 369, 465, 478, 551, 696, 701
مكنون ايض الإنسان.....	686
مكوثر سداسي.....	124
ملفات النصوص.....	45
مناطق الإقحام والحذف.....	430
مناطق بدء الانتساخ.....	121
مناظرات المحسسات.....	728
منزوعة البروتون.....	178

منصات الارتباط.....491
 منصات البروتين 260
 منطقة الرفض 321
 منطقة الظلام 372
 منطقة العاصمة 266
 منطقة الغيش.....605
 منطقة الغروب 372
 منظمة الصحة العالمية.....56
 مواقع الجينات 2, 253, 463, 556
 مواقع الزعزعة 582, 583
 مواقع خفية 505
 موجبة لصبغة كرام 16, 27, 118
 موديلات ماركوف.....628, 629, 636
 موقع الجين 11, 213, 556
 موقع الفرع 167
 موقع بدء الانتساخ...485, 487, 488, 508, 522

ن

نافذة استعلام.....47
 نبات البتولا 722
 نتائج سالبة كاذبة.....48, 630
 نتائج موجبة كاذبة 389, 534
 نزعة التداخل المشتركة 664
 نسبة التعلم.....349, 351, 352
 نسبة العزم 349, 352
 نسق الاصطفاف 402
 نسق التعبير, 541, 527, 502, 492, 221, 6
 547, 596, 695
 نسق التعبير الجيني.....6, 492, 527, 596, 695
 نسق الكراهية للماء.....640
 نسق الممهدات.....566, 567
 نظام احتساب الدرجات 647
 نظام حيوي.....689
 نظام ماكنوش 26
 نظرية النزعة المركزية.....317, 318, 325

نقاط تعادلها الكهربائية 260
 نقطة الاشتباك العصبي 348
 نقل الإشارات الهرمونية.....63
 نماذج الاتصالات.....267
 نماذج التطور 431
 نماذج كوبنهايكن 614
 نماذج ماركوف, 501, 490, 409, 341, 328
 506, 508, 510, 525, 620
 نماذج ماركوف الخفية 490, 508, 525, 620
 نماذج ماركوف المخفية.....341, 409
 نمذجة مونتي كارلو.....628
 نموذج Gibbs 628
 نموذج الجوار الأقرب.....659
 نهايات الكروموسومات 138, 165
 نهاية الكروموسومات.....124, 126, 491
 نواقل.....42, 117, 243, 257, 258, 466, 690
 نيوجرسي 35

ه

هجرة الالكترونات 68
 هلام, 534, 262, 260, 253, 248, 247, 243
 538, 551, 578
 هندسة الأغذية وراثيا.....721
 هيئة الممهد 489

و

ورق البرسيم.....661
 وسائل المحاكاة 682
 وسائل المعلوماتية الحيوية, 12, 11, 10, 8, 7, 2
 16, 170, 604, 634, 637, 651, 702,
 709, 720, 726, 729
 وسائل ريوتية 474
 وسط غذائي أدنى 678
 وقت التنفيذ 506
 وقت الجيل 413, 415, 426

فهرس انكليزي

1	Allergome 719
16S rRNA161, 162, 428, 525, 690, 691, 703, 705, 706, 707, 708, 709	Allergy 714
2	Allermatch TM 714
2µm 118	Allotropic expression..... 202
A	Alternative hypothesis 315
AAS.....518, 605	Alternative splicing..... 134, 168
Ab initio 481, 516, 599	Alu exonization..... 169
ABC transport system 204	Alu sequence..... 97, 137, 220, 232, 529
Accessible surface area 183	aMAZE 63
Accession number 15, 49	Amino acid substitution..... 518
ACT.....554, 555, 572, 594	Amphipathic 207, 642
Actinobacteria ..521, 559, 691, 705, 707	Amplification refractory mutation systems 540
activation domain 208	Analogy..... 650
Activation function.....348	Anaphylaxis 716
Adaptive evolution 429	Aneuploidy 140
Adhesins 234	Angelman syndrome..... 227, 238
Aeroallergens 721	Anisotropic bendability 584
Affine gap penalties 381	Anisotropic flexibility..... 584
<i>Ag. rhizogenes</i> 589	Annotation 2, 11, 587
Agent 25	ANSI..... 269
Agglomerative.....437	Anticodon 90
Aggregation.....186	Antigens..... 75
Ago proteins 164, 227	Antisense..... 94, 118, 165, 534, 666
<i>Agrobacterium</i> 115, 588	AOL 265
<i>Agrobacterium tumefaciens</i> 588	Apoptosis 127
Algorithms 4	Apple Macintosh..... 264
Alignment algorithms 375	<i>Arabidopsis thaliana</i> 38, 60, 100, 125, 441
Alignments 6	AraCyc..... 60, 62, 686
Aliphatic.....175	Archaea 68
Alkylation.....105	Archaeobacteria 68
Allergen avoidance database tools...721	Argonaut 221
Allergen database 719, 726	ARPA..... 265
Allergenic adjuvant 723	Artemis comparison tool 554
Allergenicity.....719	<i>Arthrobacter luteus</i> 137
	Artificial life form..... 470
	Artificial neural network..... 25, 345

Artificial neural networks	345	Biological databases	41
asm	56	Biological diversity.....	689
<i>Aspergillus nidulans</i>	131	Biological neural networks.....	345
Asymmetric PCR	546	Biome.....	689
Authentic sites.....	505	biomining.....	1
Autism.....	225	BioMOBY	25
Autocrine.....	209	Biomolecules	72
Autoradiogram	247	Bionetworks.....	369
Average linkage	365	Bioperl	14
B		Biophyton	14
<i>Bacillus subtilis</i>	99, 118, 524	Biopreservation.....	115
<i>Bacillus thuringiensis</i>	724	Bioremediation	693
Background pixels.....	476	Biosimulation.....	1
BacMap	574, 575	Biotinidase	236
Bacteriocin signal peptide	636	Biotinylation	230
<i>Bacteroides fragilis</i>	136	Bisulfite.....	37
BALiBASE.....	412	Bitmap.....	21
Ball and stick structure.....	644	bits.....	21, 270, 673
Bandwidth	269	bl2seq.....	394
Basophiles	716	BLAST4, 15, 19, 20, 49, 340, 341, 370,	
Bayesian	460	389, 390, 391, 393, 394, 395, 396,	
<i>Bdellovibrio bacteriovorus</i>	637	397, 398, 400, 464, 468, 475, 497,	
Bendability	562	502, 538, 542, 550, 551, 554, 555,	
<i>Betula verrucosa</i>	722	572, 574, 575, 594, 595, 597, 601,	
Beyond the genome.....	688	602, 614, 628, 654, 657, 664, 667,	
Biased scoring.....	48	691, 693, 705, 709, 713, 719, 727	
Bifidobacteria.....	707	BLAST atlas	594, 693
Bile salt hydrolase	708	<u>B</u> locks <u>S</u> ubstitution <u>M</u> atrices.....	386
BioAgent	25	BLOSUM384, 386, 387, 388, 392, 430,	
BioBase	39	727	
BioCyc	25, 40, 61, 62, 63, 687	Blueprint	127
Biodiversity hotspots.....	689	BMI.....	706
Bioenergetics.....	74	Body mass index.....	706
Biofluids.....	687	Boolean operators	44
Bioimaging.....	1	Bootstrap value	462
Bioinformatic tools	2	<i>Borrelia</i>	99, 113, 124
Bioinformatician	9	Box-and-Whisker.....	469, 470
		<i>Bradyrhizobium japonicum</i>	553

Branch site.....	167	Chaperonin.....	189
BRENDA	25, 58, 59, 682	Chaperonin 60.....	709
Brown Norway rat.....	726	Character-based methods.....	452
BSH.....	708	Charge-dipole	182
Bt toxins	724	Check points	70
BTG.....	688	ChemFinder	682
<i>Buchnera aphidicola</i>	23, 111	Chemical mapping fingerprints	653
Built-in	551, 700	Chemoinformatics.....	1
Bulge loop	157	Chi sites	520
Bulges.....	539	Chimeric	466, 489
<i>Burkholderia xenovorans</i>	470	Chimerism.....	468
Butyrates	707	<i>Chlamydia trachomatis</i>	111
C		Chloroplast targeting peptide.....	203
<i>C. elegans</i>	64, 97	Chou-Fasman.....	612
<i>Caenorhabditis elegans</i> ..3, 38, 100, 231		Chromatin atlas.....	591
Calmodulin.....	203	Chromatin silencing.....	213
Cancer epigenetics	527	Circular dichroism	610
Candidatus.....	557	Clades	416
Canonical sequence.....	494	Cladistics.....	413, 418
Capping	149	Cladogenesis.....	418
CarBank.....	35	Cladogram.....	422
<i>Carsonella ruddii</i>	470	Clastogenic agent.....	104
CATH.....	40, 650	Client- server model	267
CAZy database	39	Clinical bioinformatics	696
CAzyme.....	706	Cloning vectors.....	243
CBS	28, 586, 587, 614	<i>Clostridium tetani</i>	471
CDD	40	Clover leaf	661
CDS.....	21, 51, 96, 498, 569, 575	Clustal378, 402, 405, 410, 445, 605, 658, 727	
Celiac disease	715	ClustalW	727
Centimorgan.....	257	Cluster analysis.....	356
Central dogma	148	Clustering.....	2, 362, 437
Central limit theorem	317	Clustering algorithms.....	362, 437
Centrioles	69	Clustering methods	437
Centroid.....	365, 659, 669	COBRA	42
Centroid method.....	365	Coding potential.....	568
Centroid structure.....	669	Coding sequences	2
Chaperones.....	186	Coding statistics.....	507

Codon degeneracy.....	429, 497	CRISPR	165, 225
Codon usage	38, 90, 91, 562	CURVATURE.....	581
Codon usage bias	91	Cut & paste	131
coefficients Correlation.....	307	C-value.....	88, 135, 469
Cofactor.....	78	Cy5/Cy3.....	477
Cognate tRNA.....	90	<i>Cyanidioschyzon merolae</i>	463
COGs.....	602, 603	Cyanobacteria	96
Coiled coil proteins	624	Cyclins	72, 146
Coleoptera	724	Cyclization kinetics	579
Communications models.....	267	Cystic fibrosis	190
Compactosome.....	107	Cytochrome.....	425, 514
Comparative atlases	593	Cytogenetic maps.....	257
Comparative genomics.....	6, 205, 463	Cytokines	206
Compatibility	114	Cytokinesis	70
Compensatory mutants.....	653	Cytological maps	574
Complementary patches.....	619	Cytoskeletons.....	596
Complete linkage	364		
Computational epigenetics	527	D	
Computational immunology	711	<i>Danio rerio</i>	38, 698
Conditionally essential genes.....	679	Dark matter	123
Confidence value.....	477, 516	Data signals.....	461
Conjugation.....	113	Database curators.....	35
Consensus motif.....	177, 263	Databases	2, 29
Contact allergen avoidance program	721	dbGSS	36
Content sensors	481	DBMS	29
Contigs	10, 53	DBPCFC.....	720
Copenhagen models	614	dbSNP	36
Copy & paste.....	131	dbSTS	36
Copy number.....	117	DDBJ	28, 41, 42, 49, 52, 682
Core genome	110, 463, 557, 594	DDI	698
Core microbiome.....	703	<i>De novo</i> methylation.....	232
Cosmids.....	258	Deadenylases	217
Covariation.....	657	Decamers	505
CpG37, 123, 230, 231, 232, 233, 490,		Degeneracy	91
493, 494, 497, 508, 511, 527, 540,		Degenerated primers.....	537, 550
541		Delayed hypersensitivity	715
CPK.....	646	Deletion mutants.....	197
Crayfish.....	17	Denaturation	189

Dendrograms 437
 Derived metadata 34
 Desensitization 721
 Dichotomy.....419
 Digital human.....702
 Dinucleotide repeats.....123
 Diploidy 139
 Dipolarized.....178
 Dipole-dipole interactions.....181
 Disabled genes 92
 Disabled pseudogenes 512
 Disease database.....698
 Dispensable genome.....111
 Disposable buffers.....127
 Distal promoter.....487
 Distance matrices 431
 Disulphide bridges 75
 Divided and conquer alignment 407
 Divisive 437
 Dizygotic twins 37
 DNA breathing.....582
 DNA curvature 569, 577, 578
 DNA fingerprinting.....123
 DNA methyltransferases 36
 DNA polymorphism.....93
 DNA sequencing 247, 250
 DNAP120, 121, 144, 145, 216, 248,
 259, 529, 530, 531, 533, 534
 DNase 1.....584, 586
 Dnmt3b.....233
 DnmtI 232
 Domain.....68
 Dot matrix 375, 656
 Dot matrix method 375
 Downstream analysis 8
 Draft sequence.....247
 Drosha 220
 DrugBank 687, 699, 700

Dual polarization 191, 610
 Dualprimer.....543
 Duchenne muscular dystrophy locus
573
 Duplexes 82
 Dynamic programming.....48, 375, 378
 Dynamic scale.....578
 Dystrophin gene.....515

E

E value 48, 727, 729
*E. coli*17, 62, 63, 91, 108, 109, 111,
 115, 116, 118, 120, 121, 123, 153,
 160, 161, 163, 164, 213, 214, 216,
 226, 234, 243, 254, 522, 524, 528,
 530, 553, 567, 580, 583, 678, 680,
 705, 714
 EBI-MSD.....35
 EC numbers 61
 ECF.....561
 EcoCyc.....682, 683
 Ecogenomics.....689
 Eczema.....708
 Edman degradation 261
 E-Ejn 541, 542
 EID.....53, 64, 65
 Eigen values.....280
 Eigen vectors 280
 electrophoresis 257, 260
 Electrostatic forces.....182
 Electrostatic potential 722
 e-mail 265
 EMBL27, 28, 35, 41, 42, 46, 47, 49, 52,
 590, 682, 694, 713
 EMBL format.....27, 47
 e-Medicine 702
 Emission probability.....342
 Endocrine.....209

Endomitosis.....	140	Exact matching	629
Endoplasmic reticulum.....	177	ExInt	53, 64
Endoplasmic reticulum retention signals	177	Exon shuffling	96
Endo-RNase	162, 163	Exon skipping	167, 504, 505
Enhancers	93	Exonic enhancers	665
enologo	672, 673	Exonization	134
Ensemble	37	Exons	97
Enthalpy	187, 536, 539	Exonucleases.....	142, 220
Entrez	45, 46	Exosome complex.....	217
Entropy	21, 185, 536, 539	ExPASy	27, 39, 576, 600, 682
Environmental genomics.....	689	Expectation value.....	394
ENZYME25, 39, 40, 58, 62, 576, 694, 709		Exportins.....	219
Epigenetic code	232	Expressed sequence tag	36, 168
Epigenetic memory	226	Expression atlas	147
Epigenetic remodeling events	237	Extracellular factor	561
Epigenetic therapy.....	228	Extrinsic methods	479
Epigenetics	36, 227, 526		
Epigenome.....	227	F	
Epigenomics	227	False priming	551
Epinephrine	720	Farnesylation.....	177
Episome.....	115	Fast ethernet.....	270
Epistatis.....	701	FASTA4, 14, 15, 26, 27, 28, 45, 52, 53, 389, 390, 397, 398, 399, 400, 401, 468, 497, 549, 572, 614, 628, 700, 719, 726, 727, 729	
Epitopes.....	183	Fatty acids.....	78
EPS.....	21	FBA	678
<i>Escherichia coli</i>	3, 38, 99	FDA	699, 719
Essential gene database	37	Features.....	49, 51
ESTs36, 49, 466, 467, 468, 478, 479, 481, 482, 497, 502, 504, 505, 506, 513, 515, 516		Ferritin	206
Euchromatin	93	File Transfer Protocol	265
Euclidean distance.....	357	Final exons.....	98
Eukaryotes.....	68	Firmicutes471, 521, 559, 561, 691, 705, 707	
Everted repeats	472, 569	Fitch-Margoliash.....	450
Evidence code	59	Flat file.....	26
Evolutionary divergence	384	FlyBase	38, 682
Evolutionary epigenetic	527	Fold recognition.....	617

Folding	172, 604	Genetic circuits	679
FooDB	687	Genetic dialect	692
Footprint	584	Genetic distances	432
Footprinting	248	Genetic drift	512
Forward primer	536	GeneWiz	587
Fosmids	690	Geninfo identifier	51
Frameshift reading mutants	103	Genome6, 21, 25, 34, 36, 41, 51, 135, 228, 520, 575, 576, 688	
FTP	27, 265	Genome agents	25
<i>Fugu rubripes</i>	123	Genome atlases	34
Functional genomics	6, 465	Genome imprinting	228
Fusion protein	204	Genome projector	575, 576
Fuzzy matches	630	Genome reduction	520
G		Genomic islands	556
G1	71, 72, 146	Genomic inserts	545
G2	71, 72, 146, 237	Genomic SELEX	217
Gap penalties	380	Genomic stress	129
Gaussian distribution	300	Genomics	6
G-banding	138	Genotyping	162
GC clamp	538	GEO	41
GC skew	21, 121, 523, 575, 585	Geometric distribution	508
Gel Bank	41	GFF	24, 515
Gemini negative control	146	gi 51	
GenBank	15, 36, 38, 49, 50, 52, 549	Gibbs free energy	535
GenBank qualifiers	52	GIF	19, 21
Gene chips	259	GIS	57
Gene economy	168	Gliadin	715
Gene expression	6, 7, 284, 590	Global alignment	374
Gene expression clustering	7	Global methylation	231
Gene expression profile	6	Global networks	369
Gene index	468	Glycan	707
Gene ontology	24, 43, 513, 702	Glycolipids	78
Gene signals	481	Glyphosate	723
Gene silencing	218	GOLD	586
Gene synteny	556	Golgi apparatus	184
GeneCards	688, 696	Gonnet	388, 407
GeneFinder	64	Google	55, 574, 575
Genetic algorithm method	459	GPCRDB	39

G-proteins.....	178	Heuristic word method	390
Graphical user interface	264	Hexamer frequency.....	480
Greedy heuristics.....	661	Hexamer statistics.....	568
GSS	51	Hf 9	226
GSview	672	Hidden allergens	723
GUI.....	264	Hidden Markov chain	341
<i>Guillardia theta</i>	553	Hierarchical classification	44
Gumbel distribution constant	395	High noise level	377
Gut microbiome	706	High throughput.....	465
Gx.....	583	Highly repetitive sequence	124
H		High-scoring segment pair.....	391
H- DNA.....	569	HINARI	56
<i>Haemophilus influenzae</i> 3, 57, 99, 253, 254, 524		Histone.....	209, 237
Hairpin loops.....	113	Histone acetyltransferase	209
Half-life.....	93	HIV	38, 133, 169, 711
<i>Haloarcula marismortui</i>	107	HMDB	686, 700
Haploid.....	72, 88	HMM	509, 525
Haplotype	527	HMMs.....	341
Haptens.....	713	HMP.....	687, 703, 705, 706
Hard masking	396	hn RNA.....	101
Hardware	10, 424	<i>Homo sapiens</i>	3, 38, 100
Hash function	397	Homologous.....	117, 371, 651
Hayflick limit	127	Homologous recombination.....	117
Header	49	Homology	30, 118, 340, 606, 614
Heat map	705	Homoplasmy.....	431, 452
Heat shock factor.....	210	Homozygous	257
Heat shock proteins.....	189	HOT	708, 709
Helical junctions.....	158	Housekeeping genes	213
Helicases	120	HSP	391, 392, 394
<i>Helicobacter pylori</i> 99, 111, 201, 205, 642		HTML	15, 26, 56
helix-turn-helix.....	197	Human Drug Metabolism Database.....	699
Herbal pesticides	724	Human epigenome project.....	228, 528
HETAM	602	Human gene mutation database.....	36
Heterochromatin.....	93	Human genome project.....	3, 228
heterogeneous nuclear RNA	148	Human gutome database.....	708
		Human metabolome database	686
		Human metabolome library	687
		Human metabolome project.....	687

Human metagenome.....	703	<i>In silico</i> gene deletion.....	678
Human microbiome.....	693, 703	<i>In silico</i> media.....	678
Human microbiome project	703	<i>In silico</i> simulation	1, 528
Human microRNA disease Database	697	<i>In silico</i> studies	12
Human oral taxon.....	708	<i>In situ</i> PCR.....	546
HumanCyc database.....	687	Indels.....	402, 403, 430
Huntington	190	Inflammatory bowel disease	707
Hydrazine	247	Informatics.....	1
Hydropathy index.....	623	Initial exons	98
Hydrophilic	78, 175	Initiation.....	153
Hydrophilicity plot.....	727	Initiator	488
Hydrophilicity value.....	727	initn	399
Hydrophobic.....	78, 175, 208, 609, 610	Innate immunity.....	710
Hypergeometric distribution	298	Inputs	13
Hypomethylation.....	232	Insertion mutants.	197
Hypothetical	55	In-stop codons.....	691
Hypothetical proteins.	55	insulators.....	486
I		Integrases	450, 451
IBD.....	707	Integrins	713
Identity	276, 371, 373, 374, 599	Inteins	202
Identity matrix.....	276	Interaction propensity	663
Identity scores	599	Interchromosomal repetitive sequences	123
IEEE	269	Interferometry	610
IgE715, 716, 717, 718, 719, 720, 721, 723, 725, 726, 727, 728, 729		Interior loops.....	157
IHF	213	Internal exons.....	98
IMGT	40, 711, 712, 714	Internal ribosome entry sites.....	95
Immediate hypersensitivity	715	Internal transcribed spacer.....	162
Immunoblotting.....	260	internet	264, 265
Immunological bioinformatics	643	Internet.....	45, 266
Imprinting control regions	239	Internet Explorer	45
In group	420	Interpreted genes.....	92
In homologous similarity	452	Interspersed repetitive sequences	123
<i>In silico</i>	1, 12, 53, 528, 678, 680	Intrachromosomal repetitive sequence	123
<i>In silico</i> biology	680	Intraperitoneal.....	718
<i>In silico</i> biotechnology.....	1	Intrinsic curvature.....	21
		Intrinsic methods	479

Intronic RNA.....218
Inverse PCR 545
Inverted repeats 472, 569
Isoacceptor tRNA.....90
Isoallergens 728
Isomerization.....186
Iterative alignment methods.....406
Iterative optimization 648
IUPCA.....543, 550
IUSI.....714

J

Jones-Taylor-Thornton.....388
Jukes-Cantor.....432, 433, 436, 457
Jukes-Cantor model.....432
Junk DNA 93, 511

K

Karyology.....138
Karyopherins.....219
KEGG.....40, 60, 61, 682, 696
Kernel function.....663
Kimura.....434
Kissing hairpins.....158
Klebsiella pneumoniae 115
Klenow subunit 530
K-means clustering.....366
K-mer 464
Kozak sequence.....507
K-tuple.....389, 509

L

Lactobacillus delbrueckii 151
LANs.....266, 267
Latent cytoplasmic factors 209
lateral gene transfer.....118
Latex.....715, 717
Lb. bulgaricus 153

Lepidoptera 724
Leptospira interrogans 519
Leucine zipper 197, 627, 628
Licensing factors.....146
Licensing positive control 146
Ligands 7, 211
Lineage 417, 418
Linear epitopes.....729
LINEs.....130, 137, 512
Linguistic methods.....508
Linkage maps.....257
Linkout.....56
Linux.....14, 549
Lipidation.....635
Local alignment 374
Local direct repeats.....569
Local hypermethylation 541
Locus.....49
Logo20, 22, 522, 563, 630, 668, 671, 673, 674
Log-odds ratios 384
London dispersion forces.....581
Long distance PCR 548
Long-distance evolutionary relationships 162
Lookup.....47
Lossy heuristic.....405
Low complexity regions 395, 475
Lyonization 241

M

M phase.....71
Mac 26, 264
Machine learning tools 719
Mainframe computers.....264
Maintenance energy.....68
Malassezia furfur 714
MANs 266

MAP kinases	237	Metadata	24, 33
Margaret O. Dayhoff.....	3	Metagenomes.....	595
Markov chains.....	328	Metagenomic atlas	693
Marsupials.....	239	Metagenomic libraries	690
Masking.....	231	metaORFA.....	690
Masking proteins.....	231	MetaRouter	693
Mast cells	714	MetaServers	518
Matching	370, 680	method Ward	365
Mathematica.....	19	Methylases	226
Matlab	19	Methylated genes.....	36
Matrices algebra.....	274	Metropolitan area networks	266
Matrix attachment regions.....	491	Mfold	540
Matrix inversion.....	279	MHC	40, 643, 713, 714
Maxam & Gilbert.....	247	Microarray technology.....	259
Maximum likelihood.....	457	Microarrays.....	40
Maximum parsimony	452	Microbial bioinformatics	634
Mean.....	301	Microtubules.....	69
Measures of central tendency.....	301	Midnight zone.....	372
Measures of variation.....	305	Midpoint rooting approach	420
Median.....	301, 303, 552	MIGS/MIMS	44
Medline	51	Minimal gene set.....	110
Megaplasmid.....	113	Minimal genome.....	464
MEME.....	628, 729	Minimal medium	678
Meningococci.....	111	Minimal promoter.....	487
MEROPS.....	39	Minimum evolution	451
MeSH	55	Minor networks.....	369
Message assembler.....	268	MIPS	601
Message disassembler	268	miRNA39, 165, 217, 218, 219, 220,	
Meta HIT.....	706	221, 222, 223, 224, 225, 226, 526,	
MetaboCard.....	687	654, 662, 666, 697	
Metabolic islands	556	Mirror repeats	472, 569
Metabolic pathway	5, 59, 680, 685	Mismatch repair	107
Metabolic pathway databases.....	59	Mismatches	535
Metabolic pathway tinkering.....	685	Mispriming	538
Metabolome.....	6	Missense mutants.....	102
Metabonomics	6	Mitochondria targeting peptide.....	203
MetaCrop	686	MITOP	38
MetaCyc	62, 682	Mitosis	70, 71

MLVA.....	119	Nanoparticles	172
Mobilome	113	Natural antisense transcripts	219
Mode	301	NCBI27, 28, 36, 37, 38, 40, 41, 42, 45,	
Moderately repetitive sequence.....	123	49, 52, 55, 56, 57, 58, 66, 340, 468,	
Molecular biology	1	489, 518, 550, 574, 601, 602, 604,	
Molecular clock.....	412	696, 709, 713, 714	
Molecular fossils	416	ncRNA164, 165, 214, 220, 230, 515,	
Molecular phylogenetics	416	517, 652, 662, 665	
Molecular printing.....	260	NDB.....	35
Molten globule state	186	Nearest – neighbor approach	364
Monte Carlo methods.....	578	Needlman-Wunsch	381, 464
moses.....	662	Neighbor joining.....	441
Motifs	10, 65, 201, 562	Neighbor-joining method.....	405
Motifs periodicity.....	562	<i>Neisseria</i>	521
Mountain plot.....	669	Nested PCR.....	546, 550
Multi layer perceptron.....	353	Network architecture	349
Multi-FASTA.....	26	Neurodegenerative diseases.....	190
Multiple Sequence Alignment.....	401	<i>Neurospora</i>	125, 131, 231, 236
Multiplex PCR	548	<i>Neurospora crassa</i>	125, 131
Multivariate population.....	282	Neutral drift	653
<i>Mus musculus</i>	38	Newick.....	423
MutH	234	Nitrous acid.....	104
Mutual information function.....	664	NMR	35, 156, 653
MVPs	232	Node density effect	437
My SQL.....	45	Non parametric correlation coefficient	
MYcDB.....	38		312
<i>Mycobacterium tuberculosis</i> 38,	99,	Nonhierarchical.....	362, 366
111, 163		Nonhierarchical clustering.....	366
<i>Mycoplasma</i> .90, 99, 120, 526, 565, 567		Non-homologous end joining	127
<i>Mycoplasma pulmonis</i>	565	non-messenger RNA.....	165
<i>Mycoplasma capricolum</i>	120	non-redundant	15, 43
<i>Mycoplasma hyopneumoniae</i>	526	Nonsynonymous SNPs	36
MyGrid.....	25	NORFs	690
Myristoyl anchor	177	Normal distribution.....	300
N		Northern blot.....	221
<i>Nanoarchaeum equitans</i>	553	nsSNP	140
Nanobodies.....	260	Nuclear pore complexes	219
		Nucleolar organization region	138

Nucleosides	80
Nucleosome dynamics	228
Nucleosomes	226
Nucleotide signature.....	581
Nucleotides.....	81
Null hypothesis	314
Nutraceuticals.....	699

O

Object-oriented database	32, 41
Offline programs	16
Okazaki fragments	145
Oligo - skew	577
–ome	6
–omics	5, 574
OMIM	696
Oncogenes.....	210
Ontology.....	44
Open reading frame.....	3
Open system interconnection	269
Opera.....	45
Operational taxonomic units	705
Optimality algorithms	437
Optimty-based methods	449
Optimizing threshold procedure.....	367
Oral allergy syndrome.....	716
Oral microbiome	708
Oral sensitization.....	726
ORFome	6, 482, 690
Orphan entries	728
Orphan methylase.....	234
Orphelia.....	690
Orthologous groups.....	38, 622
Orthologs.....	419
<i>Oryza sativa</i>	38
OSI	269
Outgroup	420
Outputs	14

OUTs.....	705
OxySRNA.....	225

P

P value	477
Pairwise sequence alignment.....	370
Palmitoyl group	177
PAM384, 385, 386, 387, 388, 407, 430, 480, 604	
PANs.....	266
Parallel threshold procedure	366
Paralogs.....	419, 421
Paralogy	512
Paraphyletic	418
Parkinson	190
Partition function	655
Passive demethylation	231
Pathogenicity islands	556
PathPred.....	683
PathRanker.....	678
pBR ₃₂₂	116, 244
PcGrx	233
PCR11, 29, 124, 245, 246, 478, 504, 505, 512, 528, 529, 530, 532, 533, 534, 537, 540, 541, 542, 543, 545, 546, 547, 548, 550, 551, 690	
PDB35, 41, 600, 601, 602, 605, 614, 625, 644, 650, 682, 687	
PDF	19, 21
PDI.....	186
Peanut	716
Peer to peer model	267
Peptidases	39, 177
Peptide fingerprint	262
Peptide plane.....	180
Peptide vaccines.....	711
Perl.....	14, 15
Permutation test	461

Peroxisomes	69	Pleiotropic.....	101
Pfam	40	PMCID.....	55
PFGE	257	PMID	55
PGDB	688	PMN.....	59
Pharmaceutical chaperones	190	Point Accepted Mutation	384
Pharmacogenomics	489	Polycomb proteins	229
PharmGKB.....	699	Polymerase chain reactions.....	528
phase 0.....	96	Polymerase slippage	107
Phase 0	65, 95	Polynomial time.....	403
Phase 1	95	Polynucleotide kinase	248
Phase 2	96	Polyploidy.....	139
Phase variation	234, 521	Polysaccharides.....	75
Phospholipids	78	Polytomy	419
Phosphoproteomics	237	Population bottleneck	512
<i>Photobacterium profundum</i>	160, 162	Population epigenetics	527
Phylogenetic relations	412	Position effect variegation	491
Phylogenetic trees	2, 412	Position preference	585
Phylogeny.....	11	Position specific scoring matrices ...	333
Phylogram	422, 431	Posterior probability	460
Phylome atlas	595	Post-translational modifications	184
Physical maps.....	574	PPI	186, 686
Physiome.....	6, 702	Prader-Willi syndrome.....	227, 238
Physiome project.....	702	Prediction.....	4, 479
Phyton	14	pre-mRNA .97, 166, 169, 483, 500, 508	
pI260, 599, 600, 609, 721		Pre-protein	204
PIGS	222	Primary key.....	31, 49
Pioneering factors.....	211	Primase	120
Piperidine	247, 248	Primer design	11
PIR	35, 52, 54, 63, 482, 600	Primer set	548
PIR- PSD.....	600	Primers.....	29, 528, 550
PiRaNhA	665	PRINTS	602
piRNAs.....	223	Prion.....	190
Pixels.....	19	Probability density function(pdf).....	301
PKR.....	39	Probe hybridization.....	530
Plain sequence.....	26	Processed genes	92
Plain sequence format	26	<i>Prochlorococcus marinus</i>	155, 569
PlantCyc	59	Profile drift.....	341
Plasmids	68	Profiling	367

Programs 13
 Progressive alignment method 404
 Prokaryotes.....68
 PROMISE 39
 Promoters 7, 148
 Propeller twist angles 584
 PROSITE 630
 Prosthetic groups.....39, 183, 184
 Proteasome 184
 Protein – induced deformability.....583
 Protein expression 147
 Protein folding.....5
 Protein kinases 72
 Protein replacement therapy.....190
 Protein scaffolds.....260
 Proteobacteria96, 151, 162, 205, 233,
 234, 521, 637, 691, 707
 Proteome 6, 39
 Proteomics.....6, 172
 Proteopathy diseases 190
 protocols 269
 Providing.....265
 Proximal promoter 487
 Pseudo amino acids.....621
 Pseudoautosomal region.....242
Pseudococcus nipae 239
 Pseudogenes 92, 511
 Pseudoknots 158
Pseudomonas 25
 PSI-BLAST 339, 340, 341, 613
 PSSM 333, 339, 342
 PTGs.....223
 PubMed 45, 51, 53, 55, 697
 PUFs.....664
 Purifying selection 429
 Purines.....80
 PWMs.....506
 Pyrimidines 79

Pyrococcus abyssi..... 120
 Pyrosequencing..... 541, 705
 Pyrrollysine..... 565

Q

Q-banding 138
 QNet..... 686
 QPath 686
 Quadratic forms 280
 Quantitative PCR..... 534, 546
 Quantum mechanical calculations ..581
 Quartet puzzling.....458
 Query box 47
 Query sequence..... 11
 Quorum sensing.....690

R

R/M systems 234
 Ramachandran plot 179
 Random priming 550
 range 305
 rasiRNA 166, 223
 RasMol.....601
 Raster graphics..... 19
 R-banding 138
 RCSB 601
 Reactome 63
 Reads..... 10, 252
 Readseq..... 27
 REBASE 40
 Recombinases 141
 Recombination..... 113, 141
 Reference allergen 726
 Refseq 43, 53, 54, 396, 513, 691
 Refseq database 396
 Regular Markov chain 332
 Regulation.....213
 Relational databases.....31, 41

ReLiBase.....	40	RNAP121, 148, 149, 169, 178, 207,	
Remote protein homology.....	647	209, 210, 211, 213, 215, 220, 225,	
Repeats.....	48, 470, 585	226, 227, 484, 485, 487, 521, 522,	
Replica imprint.....	258	561, 580, 582	
replication.....	117	RNase.....	165, 215, 220, 545
Replication slippage.....	107	Roche.....	576
Replicons.....	68	Rolling circle replication.....	117
Replisome.....	144	Rooted tree.....	419
Reporter gene.....	244	Rose plot.....	23, 566
Resolvase.....	132	Rosetta stone.....	620
Response elements.....	208	RT PCR.....	547
Restriction enzymes.....	234	Rubinstein-Taybi.....	485
Restriction maps.....	11, 253	Ruminant microbiome.....	710
Retention signal.....	177	Runs.....	298
Retrieval of data.....	11		
Retrieval systems.....	43	S	
Retrocopies.....	134	S phase.....	71, 138, 142
Retroelements.....	129	Safari.....	45
Retropseudogenes.....	133	SAGE.....	477
Retrotransposons.....	129	<i>Salmonella</i>	111
Reverse primer.....	536	Saturation transposon mutagenesis..	118
Reverse transcriptase.....	504	Scaffold attachment regions.....	491
Rfam.....	39, 66, 666, 667	Schizophrenia.....	225
Rfam database.....	66	SCOP.....	40, 649, 650, 651
RFLP.....	141, 257	scores.....	384, 664
Rhodamine.....	250	Scoring function.....	403
Ribbon Structure.....	646	Scoring matrices.....	383
Ribonucleoprotein.....	126, 219	Sec-dependent.....	204
Riboregulators.....	218	SECIS.....	95
Ribosome decoding sequence.....	164	Secretome.....	635
Riboswitch.....	218, 225, 659	Selenocysteine.....	91, 95
RiceCyc.....	686	Selfish genetic elements.....	202
Rickettsia.....	553	Semantics.....	24
RISC.....	221, 222, 223	Semiglobal alignment.....	382
RNA – Decoder.....	665	Sensitivity.....	48, 389
RNA editing.....	149, 172	Sequence alignment.....	9, 11
RNA interference.....	220	Sequence identity.....	371
RNAi.....	164, 220, 223, 230, 528	Sequence similarity.....	284, 371

Sequential threshold procedure	366	SNPs	12, 25, 93, 140, 228, 232, 256, 505, 517, 518, 541, 551, 586, 695, 701, 702
Shadow coding sequences.....	501	snRNA	97, 165, 167, 224, 666
Shadow genes.....	525	<i>Sodalis glossinidius</i>	520
Shelterin complex.....	125	Soft masking	396
Shine-Dalgarno	483	Software.....	10, 13, 424
Short tandem repeats	123, 544	Solvent accessibility	518
Shotgun	10, 53	<i>Sorangium cellulosum</i>	470
Shotgun sequencing	10	Southern blot.....	141, 243
shRNA.....	223	Space – filling structure	645
Shuttle plasmids	116	Speciation	162, 419
SIDD	582, 592, 593	Species diversity	689
SIDD atlases.....	592	Specificity	48, 389
Sigmoid activation function	353	Sphingolipids	78
Signal peptide.....	203	Spirochaetes.....	521
Signal- sensing domain	207	<i>Spiroplasma</i>	90
Silent mutants.....	101	Spliceosomal RNA	97
Silver-Russell syndrome	238	Spliceosome.....	167, 501
Similarity.....	357, 370, 373, 374, 599	Splicing.....	134, 166, 503, 506
Similarity measures.....	357	Sporulation.....	214
Similarity scores.....	599	Spot identifier	644
SINEs	130, 137, 512	SQL.....	14, 23, 31, 41, 45, 694
Single exons	98	Squashing function	353
Single linkage.....	364	sRNAs.....	225
siRNA	36, 93, 166, 217, 219, 222, 223, 224, 225, 513	SRS	46
Sister chromatid fusion	129	Stacking energy	536
Skewness test	461	Stacking effect	382
Skin microbiome	708	Stalled ribosome	154
Skin prick test.....	720	standard deviation.....	305
Small molecule pathway database ...	701	Standard ethernet	270
SMART	631	Standard normal distribution	301
SMILES	686, 700	<i>Staphylococcus aureus</i>	572
Smith-Waterman	382, 389, 399, 480, 481, 482	Stationary phase.....	161
SMPDB	687, 699, 701	Statistical concepts.....	284
snoRNA genes.....	515	Statistical evidence	315
		Stem-loop.....	95
		Step- down PCR	542

Stochastic sampling.....	659	Tat secretion system	204
Storage area network.....	270	TATA binding protein	487
<i>Strep. clavuligerus</i>	113	Taxid.....	57
<i>Streptomyces</i> 62, 112, 124, 151, 559, 561		Taxonomic diversity	689
<i>Streptomyces coelicolor</i>	559	Taxonomy browser	57
Stress-response factors	190	Taxonomy databases.....	56
STRING	622	TCA	59, 677, 678
Structural alignment.....	409	Telomerase.....	126, 127, 128
Structural genomics.....	6, 463	Telomeres	69, 124, 125, 165
Subject sequence	374	Terminal inverted repeat.....	130
Sulfation	635	Thalassemia	104
Sum of pairs	403	Therapeutic target database	699
Sumoylation	230, 235, 236, 237	Thermal cyclers	245
Supercomputers.....	708	Thermodynamic.....	86, 549
Superfamilies.....	650	Thermodynamic theory.....	549
Superfamily.....	205, 602	<i>Thermoplasma acidophilum</i>	202
Superorganism	703	<i>Thermus aquaticus</i>	530
Support vector machine.....	634	Threading prediction.....	614
Surrogate proteins	723	Ti115	
SVM	620, 634, 663	TIGR database	37
Symbiosis islands.....	556	TMHMM	13, 642
Synapses.....	346, 509	tmRNA39, 66, 148, 151, 154, 155, 159, 163, 164, 466	
Synonymous codons	89	Tn.....	130, 131
Syntenic relationships	465	Topoisomerases	145, 582
Synteny.....	465	Topology.....	271, 461, 651
Synthetic biology	462	Torsion angles.....	179
Systems biology	694	Touchdown PCR.....	542
SYSTOMONAS.....	25	Touchstones	412
T		ToxCard	701
T cells receptors	713	Trace of matrix	276
T- COFFEE	615, 658	Transcription.....	6, 121, 147, 207, 208
T3DB.....	687, 700	Transcription coregulators.....	208
Tandem mass spectrometry.....	261	Transcription factors.....	207
Taq polymerase	245, 533	Transcription start site	121
Targeting signal.....	177	Transcriptome	6, 148
ta-siRNA	223	Transcriptomics	6, 164
		Transduction	118

TRANSFAC.....	39
Transformation.....	118
Transformers	17
Transit peptide.....	263
Transition	330, 434
Transition matrix.....	330
Translation-trans-translation	164
Translocation.....	139, 553
Translocon.....	188
Transmission technology	268
Transposable elements	129
Transpose of a matrix.....	274
Transposition.....	131
Trans-translation	154
Transversion.....	434
TrEMBL	35, 54, 600, 682
TRF1	125
Triglycerides	78
Trinucleotide repeats.....	123
Triplex DNA	569
Trithiorax proteins.....	229
Tropomyosin	722, 725
True matching	48
TrxG	236
Tsix gene	242
TSS.....	121, 485, 487
Tuple	377
Turner syndromes.....	140
Twig	417
Twin arginine	204, 636
Twin arginine signal peptide.....	636
Two component systems.....	203
U	
Ubiquitin	230, 236
Ubiquitylation	230, 235, 236, 237
UM-BBD.....	40, 63, 685
UniGene	468
UniProt.....	35, 601, 688
Unitary pseudogenes.....	512
Universal primers.....	543
Unrooted tree	419
Untranslated regions	93
UPGMA.....	405, 406, 438, 445, 447
UTRs.....	93, 94, 95, 136, 218, 220, 221, 225, 467, 480, 482, 498, 500, 515, 516, 697
V	
Vaginal microbiome	709
van der Waals force	182
variance.....	305
Vector graphics.....	19
Vectorette PCR.....	546
<i>Vibrio parahaemolyticus</i>	565
Vienna RNA webserver.....	667
Viewing tools.....	35
Virtual human project	702
Visualization.....	410
VNTR	119, 123
W	
Wallace formula.....	535
WANs	266, 267
Web sites.....	2
Weighbor	445
Weight matrices	496
Weighted parsimony.....	455
Weil.....	519
Western blot.....	260
Wet-lab-data	1
WHO.....	56, 725, 726, 727
Wide area networks	266
<i>Wigglesworthia glossinidia</i>	151
Window.....	377
Windows.....	14, 26, 264, 549

Windows vista.....	264
Wireframe	644
Wobbling hypothesis.....	90
Word method.....	375
Word weight algorithm	55
World Wide Web	265
Worm Base.....	38
WWW	265, 266

X

X chromosome	240
X-chromosome inactivation.....	228
Xenobiotics	63
Xenologs	419
XML.....	15, 42
X-ray crystallography.....	156

Y

Yeast artificial chromosome	257
YeastCyc.....	62
YIP	118
YRP	118

Z

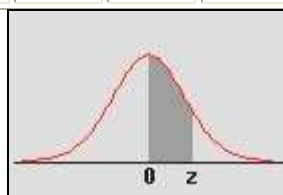
Z-DNA.....	86
Zero matrix	276
ZFN.....	117
Zinc finger	117
Zip code	263
Z-score	358, 359
Zwitterion	178

A

α -defensins.....	708

ملحق 1
(قيم Z)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990



رقم الأيداع في دار الكتب والوثائق ببغداد 1445 لسنة 2012



المعلوماتية الحيوية
Bioinformatics
2012

