

**Mohammed Baqur S. Al-Shuhaib
Ali Hmood Al-Saadi
Haider Kamil Zaidan**

Principles of Gene Biology

An Illustrated Textbook of Molecular Biology

Dedication

***This book is dedicated to the man who will fill the earth
with justice and fair after it has been filled with
injustice and unfair***

Preface

More than six years of efforts since 2006 have yielded this textbook of molecular genetics. It's with great pleasure for us to write this textbook of *gene biology*. This writing has been accompanied not only by rapid progress in many areas of molecular biology, but also by revolutionary changes in some of the ways of that molecular biologists are approaching their science. The success of genomics has also spawned global experimental approaches in other areas and given rise to the new field of systems biology, which combines large-scale experimentation with quantitative analysis and modeling.

Having been engaged in the teaching of genetics, molecular genetics, genetic engineering, and biotechnology for the last few years, we have written this book to cater to the needs of undergraduate and postgraduate students. Our objective is to introduce the undergraduate, as well as postgraduate students of science, agriculture, medicine and environment to some essentials of molecular biology of the genes.

Although this is the first edition of gene biology, it has been incorporated these advances and highlighted the exciting opportunities that mark the future of gene biology. The text begins with the structure of genetic material, moves – through ten chapters – to the function of that material, and concludes – in chapter eleven –with the application of the knowledge gained in this study. In the entirety, then, the text is concerned with the structure, expression, manipulation of genetic material with special emphasis on the role of genes in this scene.

The idea is to give the student a frame of reference to either continue the study of molecular biology or genetics or to use the information acquired to further an understanding of other courses. When, however, the experimental evidence is crucial to an understanding of how advances in the field occurred, the experiments are described in some detail.

Eventually, we sincerely hope that this text is capable to help students to get to grips with this interesting area of gene biology.

Authors



Main Contents

Number	Details	page
Chapter one	Structure and Packaging of Nucleic Acids	1 – 24
Chapter two	Gene Concept	25 – 44
Chapter three	DNA Replication	45 – 72
Chapter four	Gene Transcription	73 – 96
Chapter five	Gene Translation	97 – 127
Chapter six	Regulation of Gene Expression	128 – 162
Chapter seven	DNA Mutation and Repair	163 – 188
Chapter eight	DNA recombination and Transposition	189 – 206
Chapter nine	Systems of DNA Transfer in Bacteria	207 – 224
Chapter ten	Principles of Gene Cloning	225 – 257
Chapter eleven	How to Deal with the Gene	258 – 290
	Glossary	291 – 301

Contents

Number	Details	page
Chapter one	Structure and Packaging of Nucleic Acids	
	Introduction	1
	The Chemistry of Nucleic Acids	2
	DNA is the Genetic Material	5
	Double Helix Discovery	8
	DNA X-ray Crystallography Technique	8
	Chargaff's Ratios	9
	Watson – Crick Model	10
	DNA Denaturation	12
	Forms of DNA	13
	DNA Packaging	15
	DNA Packaging in Prokaryotes	16
	DNA Packaging in Eukaryotes	17
	Nucleosome Discovery	18
	Nucleosome Structure	19
	Chromosome Condensation	20
	Chapter One Appendix	22
	The Variable Applications for Exotic DNA Structure	
	Chapter One Revision Questions	22
	Further Readings	24
Chapter two	Gene Concept	25
	Introduction	25
	Gene Size	26
	The genetic information	27
	Gene Histology	29
	Prokaryotic Genes	29
	Eukaryotic Genes	30
	Noncoding DNA Sequences	31
	First; Intragenic (Gene Related)Noncoding Sequences	32
	A. Introns	32
	B. Pseudogenes	33
	Second; Extragenic Noncoding Sequences	33
	A. Repetitive Sequences	35
	B. Unique sequences	36
	The Flow of Genetic Information	39
	Reverse Transcription	40
	RNA Self-Replication	41
	DNA Involvement in Translation	41
	Chapter Two Appendix; Application of Prion and Its Interesting	42
	Role in understanding The Central	
	Dogma	
	Chapter Two Revision Questions	43
	Further Readings	44
Chapter three	DNA Replication	45
	Introduction	45
	Requirements of DNA polymerase	50
	The Story of DNA polymerase I and DNA polymerase III	51
	How DNA polymerase I Work in <i>E. coli</i>	51

	How the Main Replicating Enzyme Work in <i>E. coli</i>	52
	Problems Facing DNA polymerase III in Replication	55
	Solving Unwinding Problem	55
	Solving Initiation Problem	57
	Solving Directionality Problem	58
	Solving Topological Problem	61
	Solving End Replication Problem	63
	Summary of Replication Steps	66
	Other Replication Structures	67
	Rolling Circle Model	68
	D – Loop Model	69
	Chapter Three Appendix	70
	Replication in Eukaryotes	
	Chapter Three Revision Questions	71
	Further Readings	72
Chapter four	Gene Transcription	73
	Introduction	73
	From DNA to RNA	74
	Promoter	76
	1. In Prokaryotes	76
	2. In Eukaryotes	77
	RNA polymerase	78
	1. In prokaryotes	
	2. In Eukaryotes	80
	Description of Transcription	80
	Initiation Stage	81
	Elongation Stage	82
	Termination Stage	83
	RNA Processing	85
	Processing of Messenger RNA	86
	Splicing of Hetero-nuclear RNA (hnRNA)	87
	RNA Types	89
	Messenger RNA	89
	Transfer RNA (tRNA)	91
	Ribosomal RNA (rRNA)	92
	Small stable RNA (ssRNA)	93
	Chapter Four Appendix; The Ribozyme	94
	Chapter Four Revision Questions	95
	Further Readings	96
Chapter five	Gene Translation	97
	Introduction	97
	Codons and Protein Synthesis	98
	Cracking the Genetic Code	99
	Each Codon Consists of Three Nucleotides	101
	Characters of Genetic Codes	101
	Reading Frame	103
	Start and Stop Codons	104
	Protein Synthesis	105
	Initiation Stage	106
	The differences between Initiation of Translation in Eukaryotes and Prokaryotes	108

	Elongation Step	111
	Termination Stage	113
	Ribosomes Roles in Protein	114
	Synthesis	
	Protein Synthesis has Proofreading	116
	Protein Structure	117
	Post-Translational Modifications	119
	First; Three-Dimensional Structure Modifications	119
	Second; Amino Acids Side Chains Modifications	122
	a) Phosphorylation	123
	b) Glycosylation	123
	Third; Polypeptide Backbone modifications	124
	Proteins Breakdown	124
	Chapter Five Appendix	125
	The Inhibitors of Protein Synthesis.... A Tool for Treatment!	
	Chapter Five Revision Questions	126
	Further Readings	127
Chapter six	Regulation of Gene Expression	128
	Introduction	128
	Regulation of Gene Expression in Prokaryotes	129
	First; Transcriptional Control on Gene Expression	130
	The Discovery of Operon	130
	<i>Lac</i> Operon	131
	<i>Lac</i> Operon; The Complete Picture	136
	Tryptophan Repressor	138
	Attenuation: the premature termination of transcription	140
	<i>Trp</i> Operon; The Complete Picture	143
	Transcription When Tryptophan Levels are High	
	Transcription When Tryptophan Levels are Low	144
	Other Means to Regulate Gene Expression in Bacteria at	146
	Transcriptional Level	
	Second; Translational Control on Gene Expression	147
	Third; Post-translational Control (Feedback or End Product	148
	Inhibition)	
	Regulation of Gene Expression in Bacteriophages	149
	Regulation of Gene Expression in Eukaryotes	150
	The Levels of Gene Regulation in Eukaryotes	151
	First; regulation of Gene Expression at Genomic Level	153
	Second; Regulation of Gene Expression at Transcriptional Level	156
	Third; Regulation of Gene Expression at Post-Transcriptional	159
	Level	
	Fourth; Regulation of Gene Expression at Translational Level	159
	Fifth; Regulation of Gene Expression at Post-Translational Level	159
	Chapter Six Appendix	160
	Improving Tomato Quality by Preventing the Expression of	
	polygalacturonase	
	Chapter Six Revision Questions	161
	Further Readings	162
Chapter seven	DNA Mutation and Repair	163
	Introduction	163
	Types of Mutations	164
	First; Classification of Mutation According to Type of Cells	164

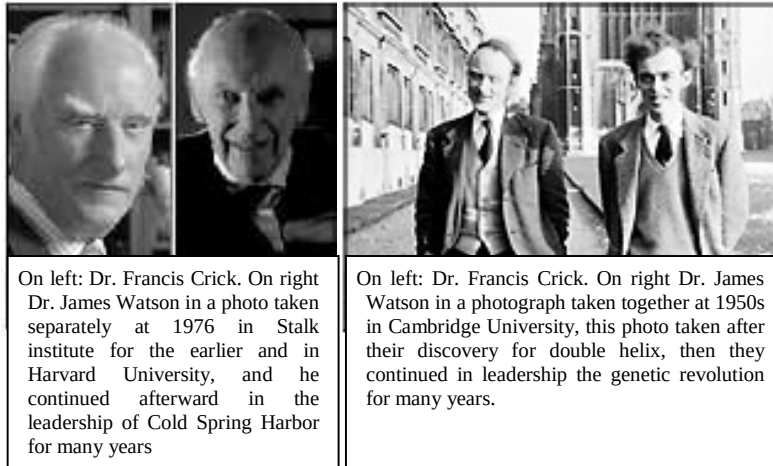
	2. Classification of Mutations According to the Size and Quality	164
	Third; classification of mutations according to their effect on the amino acid sequence of the resulting protein	105
	Fourth; Classification of Mutation According to Magnitude of Phenotypic Effect	167
	Fifth; Classification of Mutation According to the Direction	167
	Sixth; Classification of Mutation According to the Origin	167
	Mutagenes	170
	Effect of Chemical Mutagens on Nucleotide Sequence	171
	Problems of Dealing with Chemical Mutagenes	174
	DNA Repair	174
	The chemistry of DNA Bases Facilitates the Detection if its Damage	174
	DNA Repair Mechanisms	175
	1. Excision repair mechanism	175
	2. Mismatch excision repair	176
	3. Recombination (post-replication) repair	178
	4. SOS Response Repair	178
	5. Photoreactivation Repair	181
	Chapter Seven Appendix	184
	Cancer An Accumulation of Mutations	
	Chapter Seven Revision Questions	187
	Further Readings	188
Chapter eight	DNA recombination and Transposition	189
	Introduction	189
	Homologous Recombination	190
	The Role of <i>Rec A</i> Protein in Homologous Recombination	191
	Holliday Junction	192
	The Role of Homologous Recombination	193
	Site Specific Recombination	194
	Transposons	195
	General Features of Transposons	196
	Transposons Classification in Prokaryotes	197
	Insertion Sequences (IS)	197
	Composite Transposons	198
	Transposons Classification in Eukaryotes	199
	Class I	199
	Class II	199
	Mechanisms of Transposition	200
	Direct Transposition Mechanism	201
	1. Replicative Transposition	
	2. Non-replicative Transposition	202
	Indirect Transposition Mechanism	203
	Transposition via RNA intermediates	
	Chapter Eight Appendix	204
	The Advantages of Transposons	
	Chapter Eight Revision Questions	205
	Further reading	206
Chapter nine	Systems of DNA Transfer in Bacteria	207
	Introduction	207

	Plasmids	209
	First: Conjugation	211
	Mechanisms of Conjugation	214
	The Importance of Conjugation	216
	Second: Transformation	217
	The Mechanism of Transformation	218
	The Importance of Transformation	219
	Third: Transduction	219
	The Mechanism of Transduction	220
	The Importance of Transduction	221
	Chapter Nine Appendix	222
	The Destiny of Foreign DNA in The Bacterial System	
	Chapter Nine Revision Questions	223
	Further Readings	224
Chapter ten	Principles of Gene Cloning	225
	Introduction	225
	The Successive Tools for Gene Cloning	226
	First; Tools of Separation of DNA containing the Desired Gene	226
	Second; Tools of cutting and Sealing	227
	DNA Fragments	
	Third; Tools of DNA Transfer (Cloning Vehicles)	235
	a. Genes entry into prokaryotic cells	244
	b. Genes entry into eukaryotic cells	244
	1. DNA uptake (co-precipitation technique)	245
	2. Electroporation technique	245
	3. Protoplast generation technique	246
	4. <i>Agrobacterium</i> injection technique	246
	5. Biolistic technique	248
	6. Pronuclear microinjection technique	249
	7. Viral infection (transfection) technique;	251
	8. Liposome mediated DNA transfer technique	252
	9. Somatic cell nuclear transfer (SCNT) technique	252
	Chapter Ten Appendix	254
	Cloning Dolly the Sheep	
	Chapter Ten Revision Questions	256
	Further Readings	257
Chapter eleven	How to Deal with the Gene	258
	Introduction	258
	How to Obtain the Gene (or the Desired DNA Fragment)	259
	First method (Isolation of Genetic Material and Digestion of	259
	Restriction Enzymes	
	The second method (gene library construction)	262
	Third Method (direct oligonucleotides synthesis)	266
	How to Identify the Gene Sequence	271
	Labeling of Probes to Detect the Target	274
	Gene	
	How to Detect the Gene	276
	How to Amplify the Gene	280
	Polymerase Chain Reaction (PCR);	281
	Hot start PCR;	285
	Reverse Transcription PCR (RT-PCR);	285

Quantitative Real Time PCR	286
Chapter Eleven Appendix	288
How to Identify the Genetic Identity	
Chapter Eleven Revision Questions	289
Further Readings	290
Glossary	291

Chapter One

Structure and Packaging of Nucleic Acids



Introduction

Genes consist of DNA, which is the common, or RNA since the later is the genetic constituent in some viruses, otherwise, DNA is the provider of genetic information, which transcribed into RNA, which, in turn – in most cases – used later to direct protein synthesis to precisely complete gene expression. The different functions for the different nucleic acids are reflected according to their different positions in the cell, where the mature forms of RNA are localized generally in cytoplasm where the process of protein synthesis is taken place.

The total amount of DNA in a single cell differs widely among different organisms. In mammals, the content of DNA in these cells is thousands folds (approximately 8000 folds) higher than its content in bacterial cells. As for bacteriophages, such as in T-phage which infects *E. coli*, they contain only tenth or 1/12 of the amount of DNA that present in the host bacterial chromosome. The DNA content of the smallest viruses in turn is only tenth of its contents in the smallest T-phage, containing merely of genetic material distributed on ten genes. These data are entirely compatible with the truth which says that viruses doesn't have enough genetic material for independent growth, but they can grow only in parasitic mode on the host cells they infect. However, the content of DNA in the cell is not only a direct measurement for the amount of genetic material in the organism because complex Eukaryotic organisms contain of very large amount of non-informational DNA (which do not code for any genetic information) in their chromosomes, that its function still unknown.

In addition to the main DNA that exists in the Prokaryotes or in viruses, there is an extra-chromosomal informational DNA exists in specific organelles such as chloroplast (or chDNA) and mitochondria (or mtDNA) which is estimated from one thousand to three thousand copies in each Eukaryotic cell. Genes of mtDNA are encoding 13 proteins in

respiratory chain. These genes are so important since any mutation in these genes can cause some genetic diseases in human. It has to be noted that mtDNA is linked with maternal genetics, this means that cells of fertilized oocyte producing individuals consist of mtDNA resulting only from their mothers. Add to that, the studying of mtDNA holds a great importance with respect to the evolutionary chain for organisms.

The Chemistry of Nucleic Acids

When we know that DNA and RNA are the genetic material we have to investigate the chemical structure of these molecules, since their structure tells us a great deal about how these molecules are work. Nucleic acids are composed of the binding of nucleotides in a repeated way into polymers resemble the chain in their shape. Each nucleotide consists of three components; phosphate, sugar, and nitrogen base (figure 1.1). When nucleotides involved in the structure of nucleic acid, each nucleotide consists of one phosphate group, but when this nucleotide is floating freely in the cell's pool, it consists of three groups of phosphate. The nucleoside is the sugar – nitrogen base component, *i.e.*, it's a compound consists only of sugar and nitrogen base, therefore the nucleotide is nucleoside having phosphate (figure 1.1).

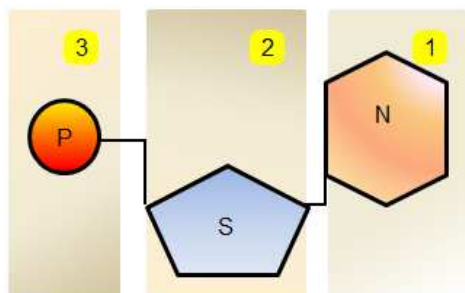


Figure (1.1). The three components of nucleotide which is the basic building block for DNA. N refers to nitrogen base (A, G, C, T, or U) of nucleotide. S refers to pentose sugar (ribose or deoxyribose). P refers to phosphate group.

It is necessary to notice that ATP (adenosine triphosphate), which is the energy currency for the cell, is nucleoside triphosphate. Sugar differs only in the presence of ribose in RNA, that's why it is called ribonucleic acid, where the oxygen is absent the molecule number two (the 2' position), note that carbon atoms in the sugar are numbered from 1' to 5'. The "" or "prime" is used to avoid confusion with the numeration system of nitrogen bases (figure 1.2).

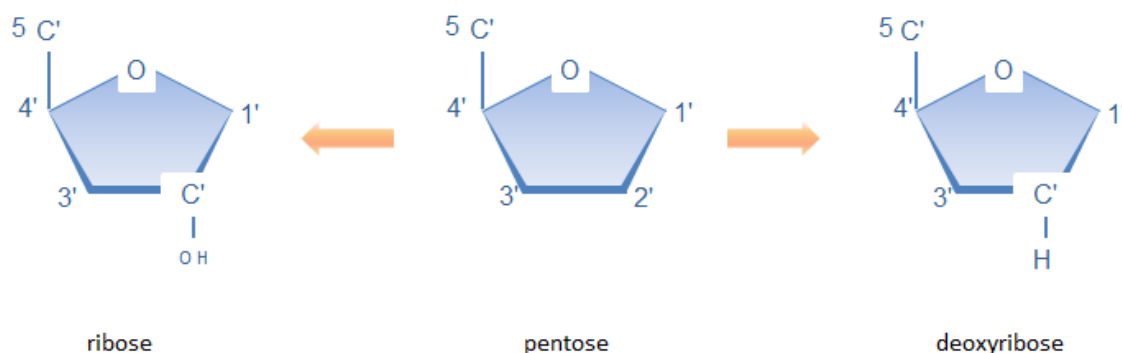


Figure (1.2). The chemical structure of ribose and deoxyribose

Both DNA and RNA have four bases (two of purine and two of pyrimidine) in their nucleotide chains (figure 1.3). Both molecules have the purines adenine and guanine and the pyrimidines cytosine. DNA has the pyrimidine thymine while RNA has the pyrimidine uracil. Thus, three of nitrogen bases are present in both DNA and RNA, while thymine is unique in DNA and uracil is unique in RNA (figure 1.3).

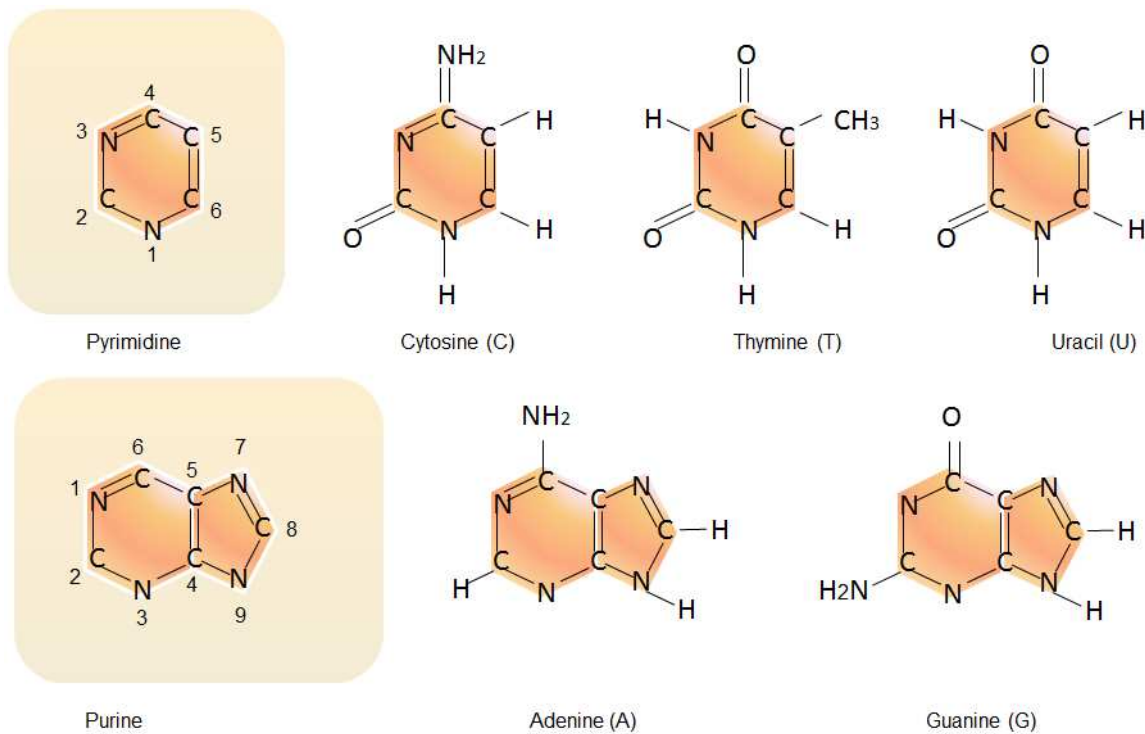
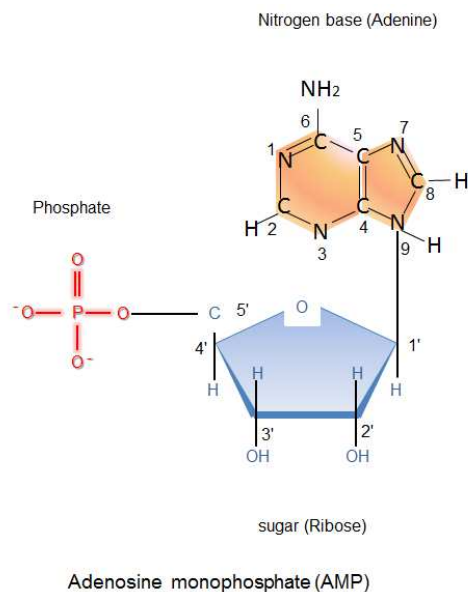
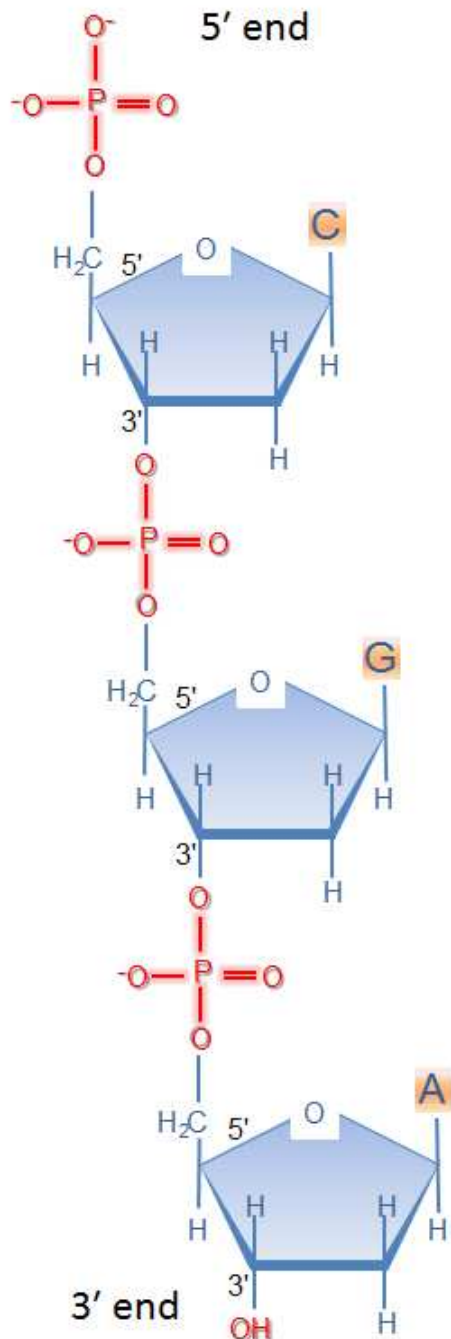


Figure (1.3). The chemical structure of the five nitrogen bases exist in nucleic acid molecules



Each nucleotide is constructed in the cell by connecting the nitrogen base (of atom number 9 in purine, or of atom number 1 in pyrimidine) with carbon atom number 1' in sugar molecule by glycosidic bond, and by connecting phosphate with the same sugar molecule by ester bond from carbon atom number 5' (figure 1.4). It does deserve to note that nucleotides take their names from the type of the nitrogen base (figure 1.4).

Figure (1.4). The chemical structure of nucleotide. Note that nitrogen base (A here) is connecting with deoxyribose through glycosidic bond, whereas deoxyribose is connecting with phosphate group through ester bond.



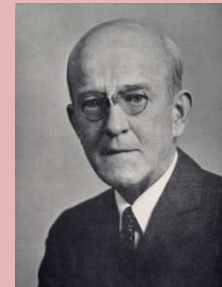
Nucleotides are connected with each other (polymerized) by forming a bond from carbon atom number 5' of a nucleotide and hydroxyl group (OH group) from carbon atom number 3' of the adjacent nucleotide, this bond called phosphodiester bond (figure 1.5). Now, a question appears; why the willing of Allah Almighty has chosen phosphate and nothing except phosphate is a link through which nucleotide are bound the each other? The answer to this question is held in phosphate group itself which have many features made it ideal for linking DNA subunits (nucleotides) into long polymers. The first one, phosphate group can form linking bonds – thus, it can connect two compounds with each other (figure 1.5).

Linking bonds that originated from phosphate (phosphodiester bonds) have additional features give them enough stability in order not to be broken easily by enzymatic hydrolysis. In another meaning, bonds are stable but nucleotide residues can be removed in order to be kept during many processes such as replication, repair (see chapter eight) and others, in such away when the nucleotide removed it can't be broken in the process. After formation of phosphodiester bond, oxygen atom that belonging to phosphate works spontaneously (because the resistance of negative charge to nucleophilic attack) that DNA enjoys in retaining it within membranes, especially Eukaryotic nuclear envelope and Prokaryotic cell membrane. Since negatively charged compounds are lipids non-soluble compounds, phosphates are retained within membranes by such ionization. Consequently, all of these conditions are corresponding to phosphoric acid and nothing except phosphoric acid!

Figure (1.5).The chemical structure of polynucleotide. This illustration shows how 5' end is phosphoric and how 3' end is hydroxylic. It must note that the other strand (not shown here) if its drawn it will extend from in opposite direction to this strand. Accordingly, its phosphate group localizes beside its hydroxyl end of this strand, whereas its hydroxyl end localizes beside phosphate end of this strand.

DNA is the Genetic Material

There are several important historical beginnings to discover the fact that DNA is the genetic material. Fred Griffith discovered during his studying of *Streptococcus pneumonia* (1928) that nonpathogenic strains could be converted to pathogenic strains by exposing to the extract of killed pathogenic strains. After sixteen years, Oswald T. Avery, Maclead and McCarty (after several years of continuous handwork) that DNA is “the transforming principle”. Within these years, those scientists isolated and purified the factor responsible on transformation process. They realized that the nature of this factor was similar to the nature of DNA molecules and differs entirely from the nature of proteins. To make sure of this, they added proteolytic enzymes such as trypsin and chemotrypsin and they found that these enzymes don’t had any mentioned effect on the transforming principle. Therefore, the search on the nature of transforming principle was excluded on the nucleic acids; DNA and RNA only. Basing on this, they added RNA destroying enzyme (RNase), they found it didn’t have any effect as well, but when they added DNA destroying enzyme (DNase) they noticed a breaking down in the biological activity of transforming principle. Therefore, the factor responsible on transformation is DNA and not other than DNA.



Oswald Avery



Maclyn McCarty

These results were published at 1944 but they were not accepted by scientific society at that time where many professionals in this field remain insisting on the fact that the protein is the genetic material. This discovery established the role of DNA as a genetic material for bacterial cell. Those researchers found that DNA extracted from the pathogenic strain of *Streptococcus pneumonia* (known as *pneumococcus*) transforms nonpathogenic strain for this microorganism into the pathogenic form (figure 1.6).

Avery and his colleges concluded that DNA extracted from pathogenic strain carried a genetic message can to be inherited among generations. Anybody was not ready to accept these conclusions, because the existence of the DNA contaminating proteins which could be a carrier for genetic information. After several experiments, once researchers excluded this phenomenon by treating DNA with proteolytic enzymes that didn’t destroy the transforming activity, but the treatment with DNase (the enzyme that destroy DNA) accomplished this task.

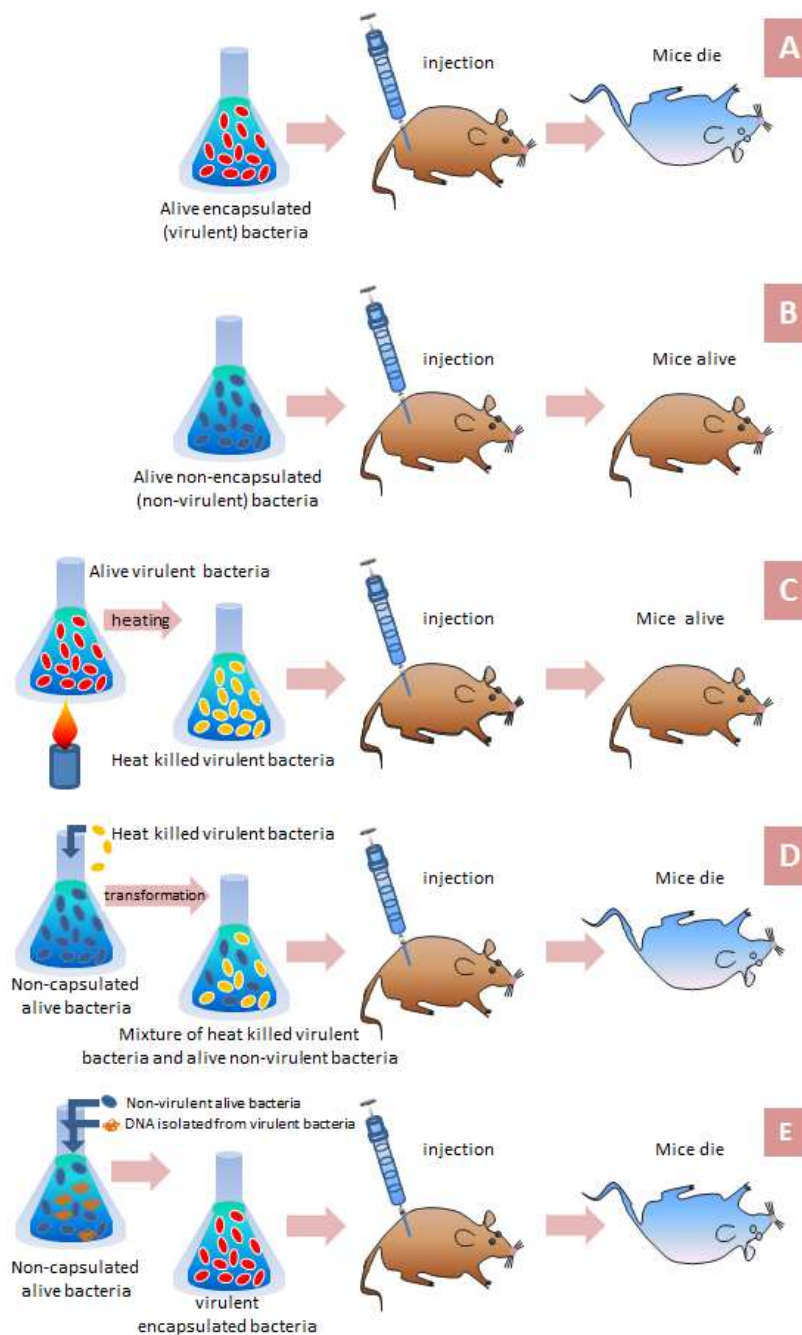
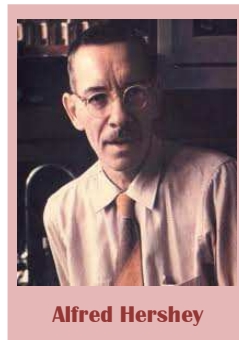


Figure (1.6). Avery MacLeod-McCarty experiment. (a) When inject mice with capsulated strain, mice die as a result to this injection, this train described as lethal, (b) whereas non-capsulated strain as long as (c) heat killed capsulated strain non-virulent. (d) A previous paper done by Frederick Griffith that the addition of heat killed virulent strain (non-harmful to mice) to the non-virulent alive strain the later is transformed temporally into encapsulated virulent strain. (c) Avery and his college extracted DNA from heat killed Pneumonia, removed proteins as much as possible, and added this DNA to non-virulent bacteria. DNA entering to non-virulent bacteria, was transform this strain temporally into virulent strain.

The another evidence emphasized that DNA is the genetic material had came from results obtained by bith Hershey and Chase form their studying to to the mechanism of T2 phage infection to *E. coli*. T2 phage is propagated through its attachment to the outer membrane to bacteiral cells and injecting its genetic material inside the cell where it replicates and directs the host cell to synthesize phage proteins in order to package its genetic material finally within phage proteins to be released outside the cell during its lysis. At that time when these two researchers were studying this phenomenon (where their paper was published at 1952), biology researchers were not realized how bacteriophage reproducing accurately. Everything they knew that T2 phage was composed of 50% proteins and 50% nucleic acids, moreover, they knew that phage binds to the surface of bacterial cell and releases from the same surface after the finishingof its life cycle. But the unknown thing for them was the mechanism of the transfer of phagal genetic material from one generation to another generation. Therefore, those two researchers was tracked the fate of protein and DNA of this phage using two radioisotopes; phosphore and sulfur since any binding molecule with these isotopes emits radioactivity could easily been detected. Consequently, the ability to know the location of radioactive molecule is been possible as well. Hence, radioactive phosphore was used (^{32}P) in the labelling of DNA because phopshore is one of the components of DNA backbone, and radioactive sulfur (^{35}S) was used in the labelling of protein because sulfur is one of the components of some amino acids of proteins. Add to that, DNA doesn't contain any sulfur and protein doesn't contain any phosphore. This mode of labeling ensures not to get any overlapped results.



Alfred Hershey

First of all, Hershey and Chase grew *E. coli* in a medium containing ^{32}P and infected the bacteria with T2 so that all the new phages would have DNA labeled with ^{32}P . They grew a second batch of *E. coli* in a medium containing ^{35}S and infected these bacteria with T2 so that all these new phages would have protein labeled with ^{35}S . Hershey and Chase then infected separate batches of unlabeled *E. coli* with the ^{35}S - and ^{32}P -labeled phages. After allowing time for the phages to infect the cells, they placed the *E. coli* cells in a blender and sheared off the now-empty protein coats (ghosts) from the cell walls. They separated out the protein coats and cultured the infected bacterial cells. Eventually, the cells burst and new phage particles emerged (figure 1.7).

When phages labeled with ^{35}S infected the bacteria, most of the radioactivity separated with the protein ghosts and little remained in the cells. Furthermore, when new phages emerged from the cell, they contained almost no radioactivity. This result indicated that, although the protein component of a phage was necessary for infection, it didn't enter the cell and was not transmitted to progeny phages. In contrast, when Hershey and Chase infected bacteria with ^{32}P -labeled phages and removed the protein ghosts, the bacteria were still radioactive. Most significantly, after the cells lysed and new progeny phages emerged, many of these phages emitted radioactivity from ^{32}P , demonstrating that DNA from the infecting phages had been passed on to the progeny. These results confirmed that DNA, not protein, is the genetic material of phages.

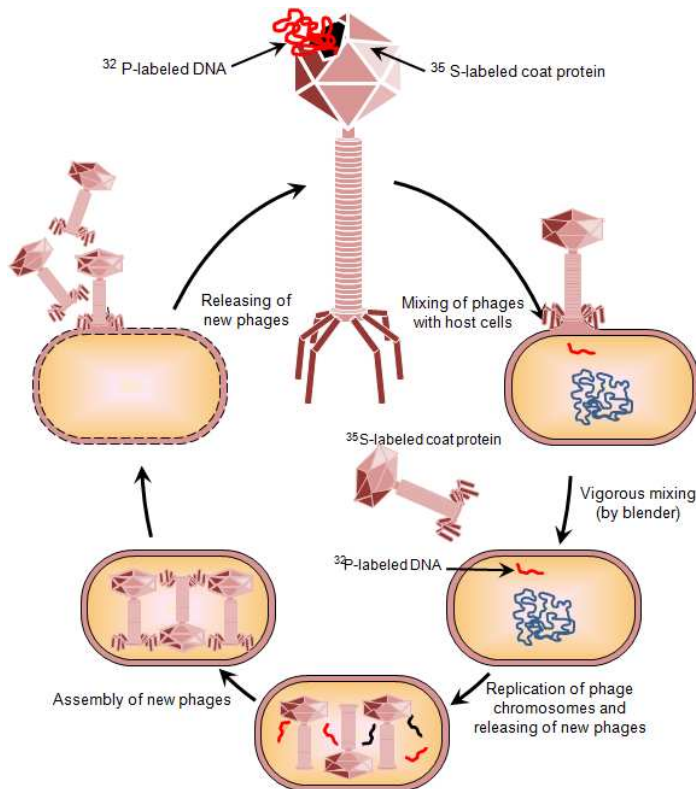


Figure (1.7). Illustration for Hershey and Chase experiment, in which it was demonstrated that DNA component of T2 phage is responsible about carrying of the genetic material, whereas protein coat only works as a protective shield.

Double Helix Discovery

There was three lines of evidences had led Watson and Crick in their known theories; the chemical nature of the previously mentioned DNA components, X-ray crystallography experiments, and what was called Chargaff's ratios.

DNA X-ray Crystallography Technique

At the beginnings of 1950s, Rosalind Franklin and her colleges utilized X-ray crystallography methodology (it's a technique used to know the three dimensional structure of molecules) in order to analyze DNA structure. Molecules of DNA in the crystal were arranged in a regular manner so that when X-ray bundle submitted on crystal it will be scattered in a regular manner. Scattering pattern can be detected on a photographic film or by computer controlled devices. The nature of this scattering depends on the structure of crystal (figure 1.8).



Maurice Wilkins



Rosalind Franklin

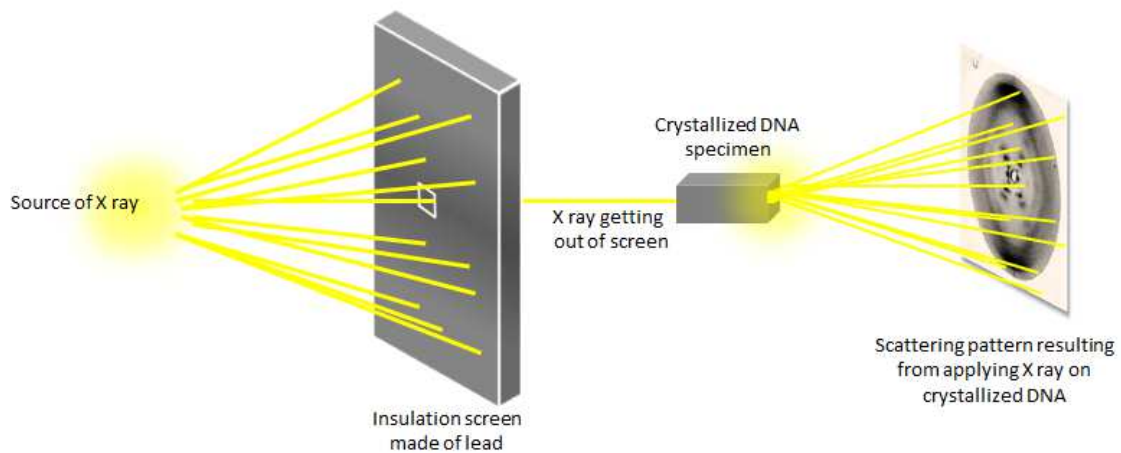
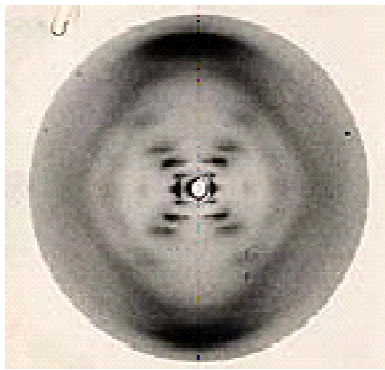


Figure (1.8).Steps of DNA X-ray crystallography.



11

Figure (1.9).The scattered mode of X-ray beam through crystallized DNA (Omoto and Larquin, 2004).

It was realized that the cross found in the center of photographic film (figure 1.9) indicated that the molecule was a helix, and the dark regions at top and bottom were came from the bases arranged vertically on the main axis of the molecule. From these results, Rosalind and Wilkins suggested that DNA was composed of two strands twisted around each other in a helical manner. The observation being DNA as a double helix was provided a major evidence that led Watson and Crick to absolutely identify the structure of DNA.

Chargaff's Ratios

Until the time at which Chargaff was working in the field of DNA, many scientists were suffered under erroneous hypothesis called tetra nucleotide hypothesis, in which, it was thought that DNA composed of equal amounts of four nucleotides, thus, the smallest DNA subunit composed of DNA consist of one copy of these four nucleotides. Chargaff carefully analyzed the nitrogen bases components in different organisms, he found, in spite of the fact that relative components of the given nucleotides were differed among organisms, the amount of adenine equal to the amount of thymine, and the amount of guanine equal to the amount of cytosine. For example, if human DNA contains 30% for each adenine and thymine, so the ratio of each of guanine and cytosine are 20%. That is, in the DNA in all studied organisms there is 1:1



correspondence between purine and pyrimidine. This is known as Chargaff's rule. But the observations made by Chargaff himself never demonstrate the tetranucleotide hypothesis, so that he noticed that the four nucleotides of DNA were not in 1:1:1:1 ratio. These results gave a flash of insight for each Watson and Crick to develop the model which they involved to build it since that time. Despite of what Erwin Chargaff was postulated in his hypothesis, he partially solved the base pairing puzzle because he showed that adenine ratio usually equal to thymine ratio, and the case is the same for guanine and cytosine.

Watson – Crick Model

After the above mentioned experiments, many information were accumulated in Watson's mind which made him capable to imagine how base pairing was taking place. In February at 1953, a flash of insight was happened to Watson when he saw that adenine and thymine bond was exactly the same length of guanine and cytosine bond, and when such bonds bind in this way the one stair of the helical ladder would be equal in length to the other stairs. By relying of both Watson and Crick on the above mentioned data, they managed to know the structure of DNA in University of Cambridge at 1953, then they both got Nobel Prize at 1962. According to the available data, Watson and Crick built their molecular models. They found that the reasonable structure of DNA was of two twisted helices around each other (double helix) with backbone consisting of sugar – phosphate located toward outside whereas nitrogen bases located inside (figure 1.10).

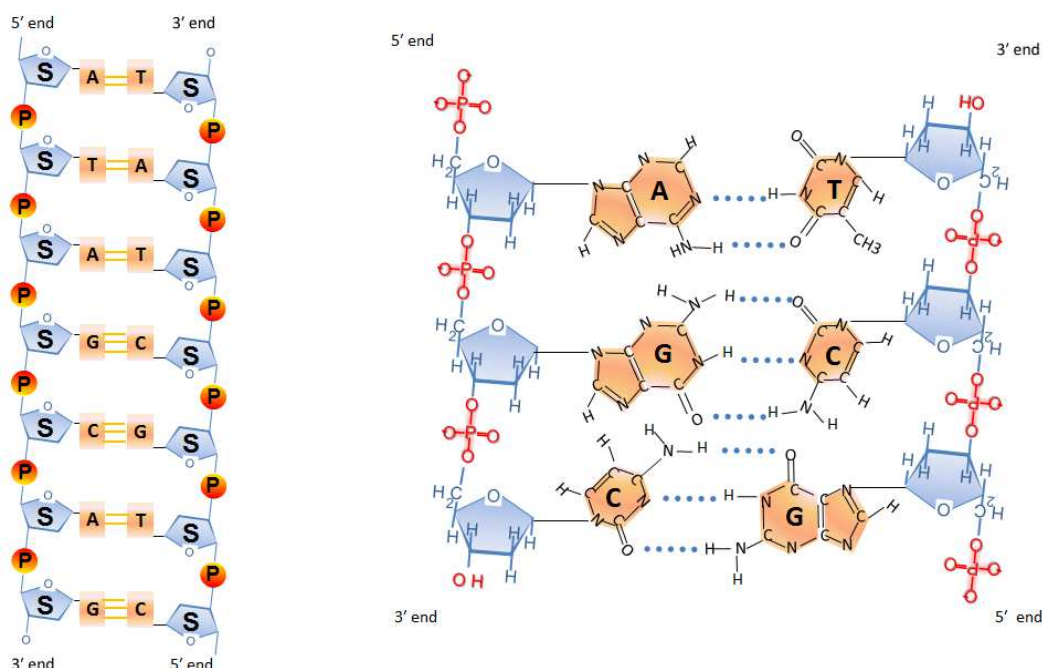


Figure (1.10). DNA double helix: chemical structure of DNA, in which nitrogen bases directed toward inside whereas sugar and phosphate directed toward outside constituting the backbone of the molecule. In the left portion of the figure the outside skeleton protrude to outside while nitrogen bases connect to each other by hydrogen bonds. The right portion of this figure explains the details of these connections.

This structure is concomitant with the suggested DNA dimensions in X-ray crystallography if the bases are localized oppositely in two strands and formed the stairs of the helical ladder. The dimensions of the double helix are 20 Å approximately (10 Å equals 1 nm) if there is one purine and one pyrimidine for the one transverse beam (figure 1.11). Two purine would be too big, as well as two pyrimidines would be too small.

After undergoing other experiments with such models, Watson and Crick found that hydrogen bonding was necessary to form the stairs of the helical ladder which would be taken place easily between specific base pairs (these pairs that Chargaff found in equal repetitions). These hydrogen bonds characterized by their very weak nature, they formed between two negatively charged atoms, as in oxygen atom and nitrogen atom. Hydrogen bonds have power estimated with 3 – 5% of the strength of covalent bonds. Hydrogen bonding thermodynamically stabilized between thymine and adenine and between cytosine and guanine. However, as long as $A = T$ and $C = G$ (according to Watson and Crick model), therefore, the total quantity of purine equals the total quantity of pyrimidine, as in the following equation; $(A + G = T + C)$.

Two types of grooves along the surface of DNA molecule are observed; one of them is wide and deep – major groove – and the other one is narrow and shallow – minor groove – both grooves are large enough in order to allow protein molecules to bind with bases (figure 1.11). In each case, the most huge nitrogen base, the base with two rings (purine), with the singular ring nitrogen base (pyrimidine), where – as mentioned above – adenine binds with thymine, and guanine with cytosine (figure 1.11). To increase the efficiency of base pairs packing, the two backbones composing sugar and phosphate twisted in a helical manner to form a double helix, in which each 10 base pairs create complete round (complete turn approximately 180°) (figure 1.11).

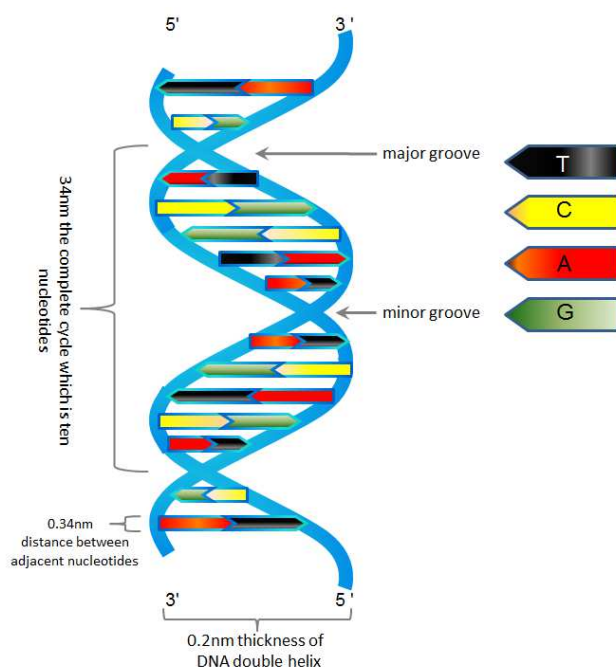


Figure (1.11). illustration for DNA double helix, in which the dimensions of most common form of double helices in cells is clarified (included major and minor grooves). Green colour refers to G, yellow colour refers to C, black colour refers to T, red colour refers to A.

Nucleotides covalently connect with each other in the chain through sugar phosphate groups, so that they make successive “backbone” of “.....sugar – phosphate – sugar- phosphate – sugar – phosphate.....”. But the long DNA chain consists of many nucleotides don’t differ in anything

except in their four nitrogen bases, therefore, it is reasonable to resemble DNA with necklace composed of four beads with different colours.

It is so important to notice that the DNA chain has two ends, one of them ends with 5' phosphate, and the other ends with 3' hydroxyl. It is so important to notice too that the direction of nucleotides synthesis is from 5' to 3'. So it become compulsory for us to read nucleotides sequences in DNA or in RNA from 5' to 3' (see chapter three). There is another point about DNA structure which correlates with the fact of existence of what is called the polarity in each strand. That is, the both strands are antiparallel (figure 1.12). If we suggestively refer to the absence of antiparallel feature for both strands, *i.e.*, we refer to the absence of hydrogen bonding between them. Therefore, in order to form hydrogen bonding between strands it is inevitable for the strands to be antiparallel.

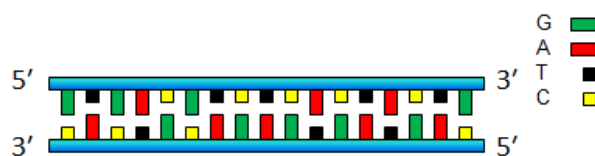


Figure (1.12). illustration shows the double directions for DNA double helix strands. Green colour refers to G, yellow colour refers to C, black colour refers to T, red colour refers to A.

The reason behind the hydrogen bonds existence between complementary base pairs hidden in the fact in which one of the strands able to work as a template to synthesize its second complementary strands. Thus, nucleic acids have unique ability to direct their own replication. Hence, nucleic acids possess the sovereignty on the cellular molecules (there is no other molecule in the cell possess's such feature) so that information kept in DNA are usually transferred into RNA then into protein, the later control most of the cellular activities.

It deserve to note that the double helix may be embodied by imagining the shape that appear when one of us take a ladder and twist it into a helical shape while its stairs stay in its vertical position. Sugar and phosphate molecules form the fence (the backbone) of the ladder in which sugar alternates with phosphate, while the stairs of the ladder represented by one purine and one pyrimidine. Within each chain (fence), it is possible for the four nitrogen bases to present in any ratios and in any system. We have to note that we only studied the primary and secondary structure of DNA, where DNA primary structure refers to how nucleotides binds with each other, whereas DNA secondary structure refers to the three dimensional configuration for the double helical structure discovered by Watson and Crick, as for DNA tertiary structure, it refers in turn to how DNA packages itself into chromosomes (see below).

DNA Denaturation

Denaturation studies referred that hydrogen pairing in DNA taken place in the way suggested by Watson and Crick. Although hydrogen bonds are too weak in the singular form, but they give structural stability for the molecule when their number increased. However, hydrogen bonds can be broken down and DNA strands separated by increasing the heating of DNA molecules in water. When increase heating into a certain point, thermal

agitation prevailing hydrogen bonding and the molecules become denatured, or melted (figure 1.13).

It is rational to say that the more hydrogen bonds included in the DNA, the more temperature is required to denature it. That is that the more G – C content ratio in DNA molecule, the more temperature is required for DNA denaturation. It is worth mentioning that temperature alone is not the only denaturing factor for DNA molecule. Alternatively, some chemical treatments – as in alkaline solution, formamide, urea....etc. – do the same role in denaturing DNA molecules. The resulting single stranded DNA molecules can bind with their complementaries in order to retain to their double strandedness state again. This is against denaturation as in getting the temperature into the normal degree in the lab, and this process is called renaturation (figure 1.13).

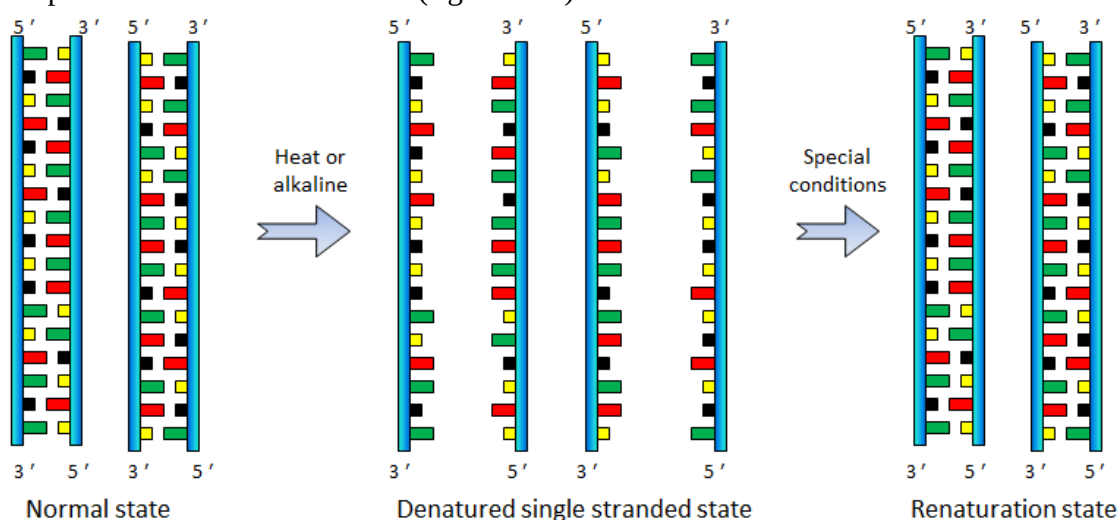


Figure (1.13). denaturation and renaturation of DNA molecule.



Forms of DNA

The shape of DNA that described till now is called B DNA. It is a right handed helix, which rotates in a clockwise manner when its axis viewed. Its bases arrange exactly vertical on the main axis, in which ten base pairs form complete turn in $34 \text{ \AA}$ (figure 1.14). The B form is the most popular form. However, DNA can be existing in other forms, if the content of water is more than 75%, at this point DNA double helices surge to each other and make new structures called A DNA, which characterized by the existence of 11 base pairs for each turn. In this form (A DNA) bases are slightly tilted with respect to the main axis. This form and other known forms are secondary variations of the common right handed B form which are out of the scope of this textbook.

Figure (1.14). The structure of common DNA which is known as B DNA.

But the thing that is really deserve to be noted in this aspect is the discovery made by Rich and his colleges in MIT at 1979. They discovered left-handed helix (figure 1.15), they called it Z DNA because it forms a zigzag structure.

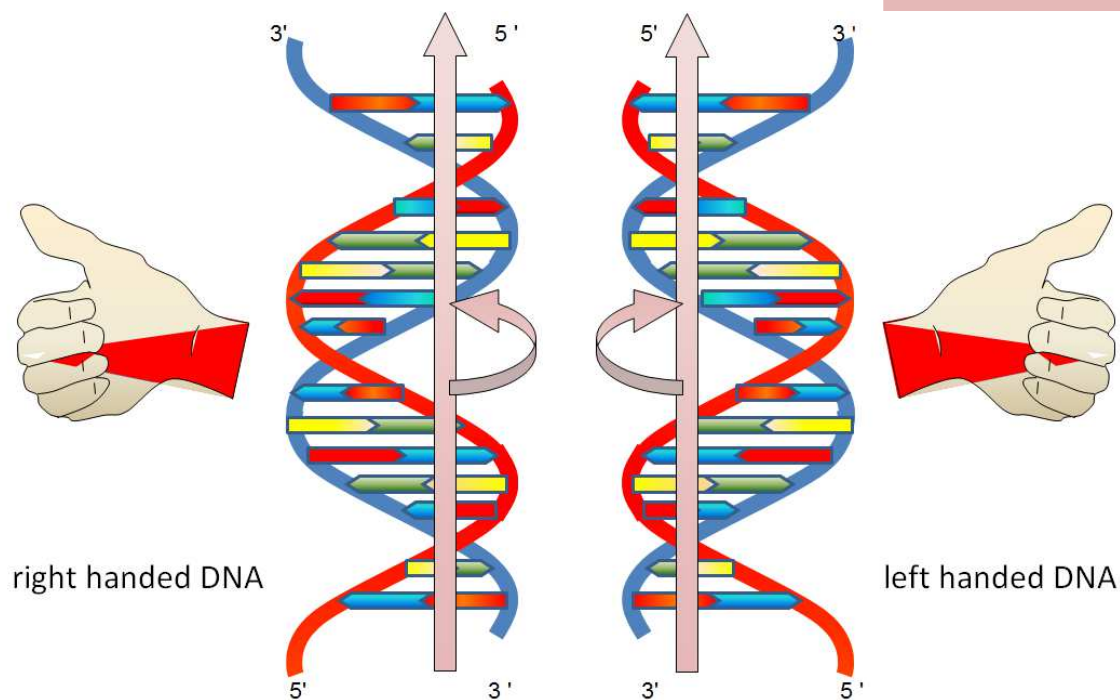
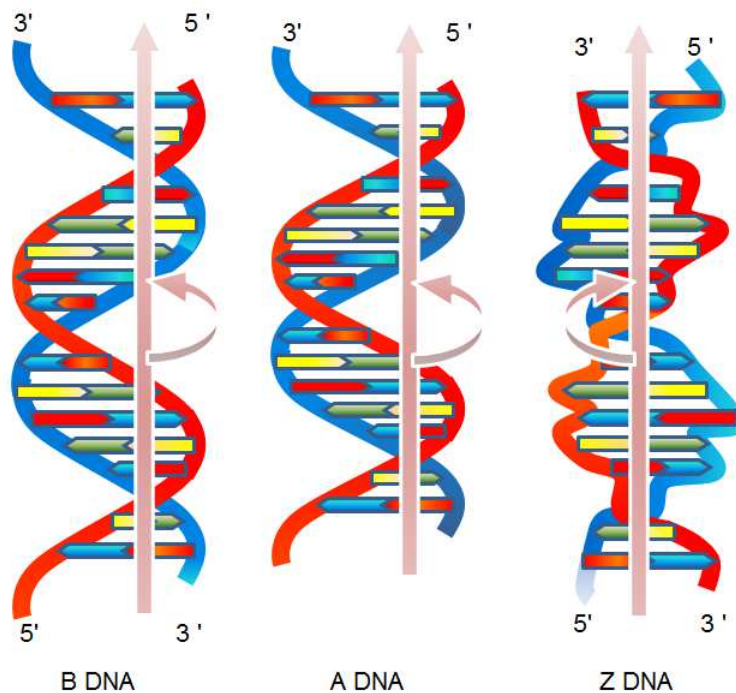


Figure (1.15). Illustration Shows the classic direction (right hand or clockwise) of DNA which is the same direction noted in B DNA and its derivatives, and the other rare DNA direction (left handed or anticlockwise) which is the same direction noted in Z DNA.

Z DNA was discovered by X-ray crystallography of very small DNA molecules which consist of repeated G- C sequences in one strand, and their complementary repeated C- G sequences in the opposite strand. It seems that Z DNA resembles B DNA in terms that its bases rotate 180° as well, but their turn here result in a zigzag form of left handed structure (figure 1.16).

The structure of Z DNA is noticeably thinner than B DNA and contains 12 base pairs instead of 10 for each one turn. it was thought initially that Z DNA was a structure don't deserve to get a lot attention of biologists (compared with B DNA) since it requires high concentration of salts in order to be stable, moreover, its biological function was not obvious at that time. But after performing several experiments it was demonstrated that Z DNA can be stabilized in physiologically normal conditions if methyl groups are added to the cytosines. Now it is clear that Z DNA involved in regulating gene expression in



A DNA structure.

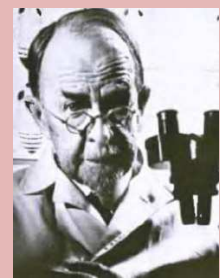
eukaryotes (see chapter six). However, although both Z DNA and B DNA having double helix, they both having opposite nucleotide strands, in addition to their commencing in the same pattern of binding of nitrogen bases of guanine with cytosine, but the amount of differences mentioned above are much more than the similarities.

Figure (1.16). three dimensional structure of zigzag DNA molecule which is known as Z DNA compared with B DNA and

DNA Packaging

At the beginning of 1902, some researchers noticed the separation of chromosomes during meiosis in the way postulated by Mendel. But the binding of specific genes in chromosomes was never been discovered until 1910 when Thomas Morgan noticed the white-eye mutation in *Drosophila melanogaster*. At the beginning of Morgan working, it was known at that time that the chromosomes were localized in the cells, but their importance are not known and not of interest. Some researchers were postulated that chromosomes carry the genetic material but they didn't provide enough clues to support this important hypothesis, therefore, Morgan provided the evidence which ensure that chromosomes were the particles responsible on inheritance. Morgan designed chromosome theory for inheritance, which dealt with genes as particles not capable on division that arranged in the form of beads on the string to form chromosome. These binding patterns provided clues to know the system of the longitudinal arrangements on chromosomes in order to understand recombination and chromosomal rearrangement (see chapter nine).

We will see how genes arranged in chromosomes (see chapter two), but in order to make a functional chromosome, DNA molecules must be capable on replication, and the



Thomas Morgan

replicated copies must be separated and distributed evenly into daughter cells in each replication carried out by the cell. This process is taken place through a series of stages called altogether "cell cycle".

DNA Packaging in Prokaryotes

Bacteria are unicellular organisms which lack the nuclear envelope. Bacterial chromosome usually takes circular shape which contain singular DNA molecule consisting of several millions of base pairs. For example, the length of *E.coli* genome estimated about 4.6 (4.6×10^6) million base pairs of DNA.

The genes of eukaryotic organisms doesn't much more complex than genes found in prokaryotes only, but the DNA of eukaryotic organisms organizes differently compared with the DNA found in prokaryotes. Prokaryotic genome consists of singular chromosome, which is a circular DNA molecule. In contrast, Eukaryotic genome consists of multiple chromosomes, each of which is a linear molecule of DNA. As it will be discussed later, DNA of eukaryotic cells binds tightly with small basic proteins called histones, which undertake DNA packaging in a regular way inside nucleus. This function is highly important in most eukaryotic organisms. Moreover, the process of DNA packaging in prokaryotic organisms is problematic too, but the way of packaging here is differ from that found in eukaryotes, as it will be discussed below.

Most bacteria contain genetic material consist usually of circular DNA molecule. In circular bacterial chromosomes, there is no relaxed circular or opened circle DNA because of inability of this circle to get inside the bacterial cell. Bacterial DNA never bind with histonic proteins (as in histonic proteins to be discussed in linear eukaryotic chromosomes). Consequently, bacterial chromosome called "naked DNA" for this reason, but this idiom is incorrect, since bacterial chromosome forms complexes with a number of proteins which bind with it in order to help in its condensation. When a bacterial cell is visualized with the electron microscope, its DNA frequently appears as a distinct clump, the nucleoid, which is confined to a definite region of the cytoplasm (see below). If the bacterial cell is broken up gently, its DNA spills out in a series of twisted loops (figure 1.17). The ends of the loops are most likely held in place by proteins. Many bacteria contain additional DNA in the form of small circular molecules called plasmids, which replicate independently of the chromosome (see Chapter nine).



Figure (1.17). The mode of bacterial chromosome folding into a series of coiled loops as it shown under electron microscope (Pierce, 2002).

Bacterial DNA binds with proteins that package and condense it into bacterial cells, that is they have function resembles the function of histones found in eukaryotes, but they differ from histones in the absence of the same structure. However, accumulated information are not available on the mode of bacterial DNA packaging, but there is a consensus on the fact that length of chromosomal DNA contained within *E.coli* cells is only about 2 μ m. The free DNA molecule in this size needs to condense for thousand folds until it reaches to the size of *E. coli* cells. However, many mechanisms are operated to package *E.coli* chromosomal DNA sufficiently to accommodate with the size of bacterial cell. However, in prokaryotes, the “beads on a string” structure which is common in eukaryotes is never seen here (see below). Since prokaryotes don’t have nuclear envelope, thus, condensed chromosomes with their surrounding proteins lie freely in the cytosol in a region called nucleoid (a region exists in bacterial cells that contains the bacterial chromosome) to show a sort of functional similarity with the nucleus that exist in eukaryotes.

It is impossible for DNA molecule to exist in a naked form in *E. coli*. For example, if the effect of digestive enzymes on DNA is abolished and suppose that DNA molecule in *E. coli* are not bound with other molecules, so, it is not possible entirely for these molecules to be compacted and entered into the cell, since condensation process needs folding of nucleotides on each other and this can’t be taken place because of charge repulsion among negatively charged phosphate groups. This can’t be happened because this effect is outstanced by the binding of DNA with the positively charged polyamines, such as spermine and spermidine, which mask the negative charges of DNA phosphate groups (because of their positive charges). Moreover, numerous small protein molecules associate with the chromosomal DNA, causing it to fold into a more compact structure. The most abundant of these proteins, H-NS (histone-like nucleoid structuring protein), binds tightly with DNA and compacts it considerably, as indicated by the increased sedimentation rate and decreased viscosity of DNA associated with H-NS compared with free DNA. Finally, *E. coli* chromosomal DNA is tightly supercoiled — that is, twisted upon itself. In *E. coli* there is an enzyme called DNA gyrase uses energy derived from ATP hydrolysis to wind supercoils in the DNA (see chapter three). Supercoiling contributes to the compaction that is necessary to fit chromosomal DNA into the bacterial cell.

DNA Packaging in Eukaryotes

DNA binding proteins in eukaryotes are divided into two main classes: histones, nonhistone chromosomal proteins. The complex of both classes of proteins with the nuclear DNA of eukaryotic cells is known as chromatin.

Typically, chromatin contains of equal amounts of proteins and DNA. Histones are the essential proteins in chromatin, which are small proteins containing high ratios of basic amino acids (lysine and arginine) which facilitate the binding with the negatively charged DNA molecules. The nature of this binding is an electrostatic attraction, in which positively charged histones are attracted (because the existence of positively charged amino acids) to the negatively charged DNA molecules (because of the existence of phosphate groups) that work efficiently to neutralize DNA charge. But the nature of this binding is unstable. Therefore, specialized proteins are operating to stabilize this binding between histones and

DNA, these proteins called nucleoplasmins, which with their help histones bind with DNA to form nucleosome. Though these proteins involved in this process, they never bind neither with DNA nor with chromatin. Although nonhistonic chromosomal proteins exist in low quantity compared with histone quantity, these proteins present in different types, which have acidic nature, these proteins participate in many activities such as DNA replication and gene expression (see chapter six). As for the DNA component, genomes of higher organisms contain DNA quantities much more of those required for genes since most of DNA doesn't contain coding functions (see chapter two).

Nucleosome Discovery

Nucleosomes are the basic units of eucaryotic chromosome structure, which discovered by Roger Kornberg at 1974. He carried out two types of experiments which led to build a model for nucleosome. The first one; the partial digestion experiments (the meaning of partial digestion is the temporal treatment of the specimen with an enzyme, which gives sometimes different results compared with complete digestion) of chromatin using micrococcal nuclease (DNA digesting enzyme), the addition of this enzyme result into separation of chromatin and the appearance of DNA pieces in about 200 base pairs in length. In contrast, Kornberg was exposed a naked DNA (not bound with proteins) to the same digestion conditions by the same enzyme, but the result was different, a continuous smear of randomly sized fragments were formed. These results suggested that the binding of protein with chromatin DNA protect regions of DNA from digestion with nuclease, consequently, it could be for the enzyme to attack DNA only at specific locations separated by 200 bps.

In accordance with this notion, electron microscope discovered that chromatin filaments had beaded appearance. Beads are separated from each other by 200 bp. When interphase nuclei were broken tightly and their contents were examined under electron microscope, it was noticed that most of the chromatin formed 30nm fibers (figure 1.18 A). when chromatin exposed to treatment that partially unfold it, it could be seen under electron microscope as a chain of "beads on a string" (figure 1.18 B).

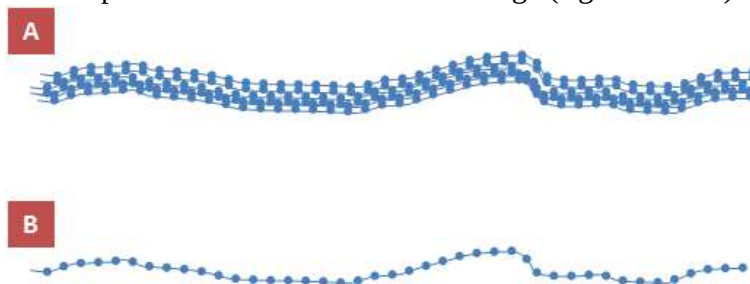
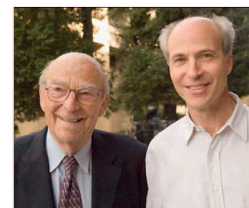


Figure (1.18). nucleosome as it illustrated under electron microscope. (A) directly isolated chromatin from interphase nucleus which appears like a 30nm thickness thread. (B) this electron microscope shows the length of chromatin which its condensation is experimentally removed after the separation of nucleosomes.

The string is the DNA, and each bead is the "nucleosome core particle", which consists of DNA twisted around histone containing core protein. Beads on a string is the



**The father Arthur Kornberg (left)
The son Roger Kornberg (right)**

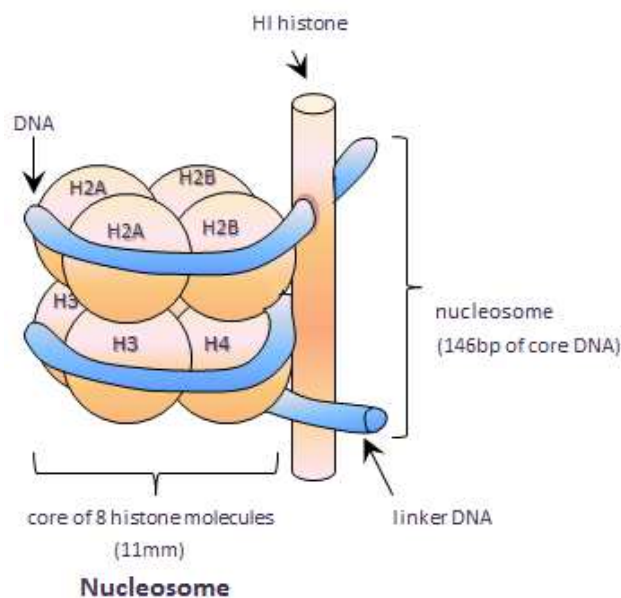
first level of DNA packaging levels. Thus, both studies (digestion with nuclease and electron microscopy examinations) suggested that chromatin consists of repeated subunits of 200bp which called nucleosomes.

Nucleosome Structure

The precise experiments showed that there were particles in the nucleosome core, it was shown that these particles contain 146 bp of DNA twisted 1.75 fold around histone core and contains two molecules of H2A, H2B, H3, and H4 (core histones). The fifth type of histones is only one molecule which is H1 which binds to the inlet and outlet of the core nucleosome. This in turn forms a chromatin subunit called chromatosome which composed of 166 bp of DNA twisted around core histones and held in place by histone H1 that called linker histone (figure 1.19).

DNA packaging in histones results in a 10nm chromatin fibers which consists of chromatosomes separated by linker DNA segments. The packaging of DNA into 10 nm chromatin fiber shortens its length six fold. Chromatin can condense much more this point, by coiling six nucleosomes into 30 nm fiber.

The interaction of histone H1 molecules with adjacent nucleosomes plays an important role in chromatin condensation stages. It is noted that condensation takes place in chromatin utilizing H1 histone being the less binding histones with chromatin, so, it is easy to be removed after its binding. Hence, the role of this histone involves in condensation. As for the mode of action of histone H1, it depends on adding and removing phosphate groups (hyper-hypo phosphorylation) into histone H1. For example, when a phosphate group is added into histone H1 it dissociates from its place to reduce chromatin condensation, and vice versa.



Each individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—and double-stranded DNA that is 146 nucleotide pairs long. The histone octamer forms a protein core around which the double-stranded DNA is wound (Figure 1.19). Each nucleosome core particle is separated from the next by a region of linker DNA, which can vary in length from a few nucleotide pairs up to about 80. (The term nucleosome technically refers to a nucleosome core particle plus one of its adjacent DNA linkers)(Figure 1.20).

Figure (1.19). The role of histone H1 in DNA packaging around nucleosome core particles.

At 1997, the molecular structure of nucleosome core particle was revealed in which it was found disk shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil (see figure 1.19). All four of the histones that make up the core of the nucleosome are relatively small proteins (more than 100 amino acids).

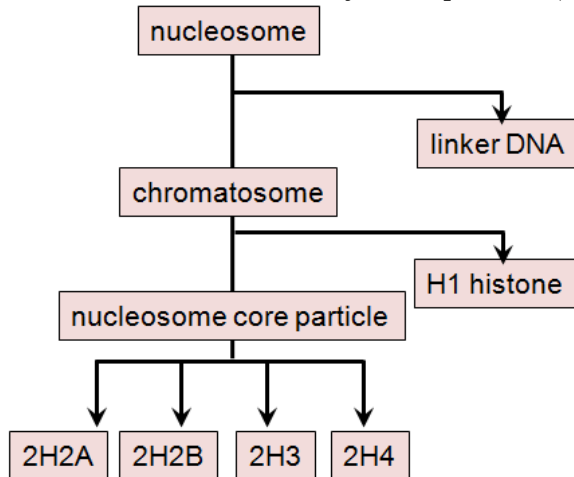


Figure (1.20).a diagram shows nucleosome structure

Chromosome Condensation

There is serious challenge of DNA packaging issues must be understood for the reader. Each human cell contains approximately 2 meters of DNA if stretched end-to-end; yet the nucleus of a human cell, which contains the DNA, is only about 6 μm in diameter. This is geometrically equivalent to packing 40 km of extremely fine thread into a tennis ball! The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization, preventing the DNA from becoming an unmanageable tangle. Amazingly, although the DNA is very tightly folded, it is compacted in a way that allows it to easily become available to many enzymes in the cell that replicate, repair, and use its genes in case of need to produce proteins (figure 1.21).

When cell enters mitosis its chromosomes become highly condensed and then are separated and distributed to the two daughter cells. It is thought that the 30 nm loops of chromatin fibers folds on itself much more than this to form compact metaphase chromosomes in dividing cells, in which DNA is condensed for ten thousands folds (figure 1.21). This condensed chromatin can't be used as template to synthesize RNA, by this, transcription is terminated through mitosis. Electron microscope studies indicated that DNA of metaphase chromosomes can be regulated into long loops through their binding with protein scaffold (figure 1.21), but now we neither exactly understand the associated structures nor the molecular mechanism involved in this process.

All eukaryotes have elaborate ways of packaging DNA into chromosomes. For example, human chromosome No. 22 contains about 48 million nucleotide pairs. If it stretched out end to end, its DNA would extend about 1.5 cm. Yet, when it exists as a mitotic chromosome, chromosome 22 measures only about 2 mm in length, giving an end-to-end compaction ratio of nearly 10,000-fold. This remarkable feat of compression is

performed by proteins that coil and fold the DNA into higher and higher levels of organization.

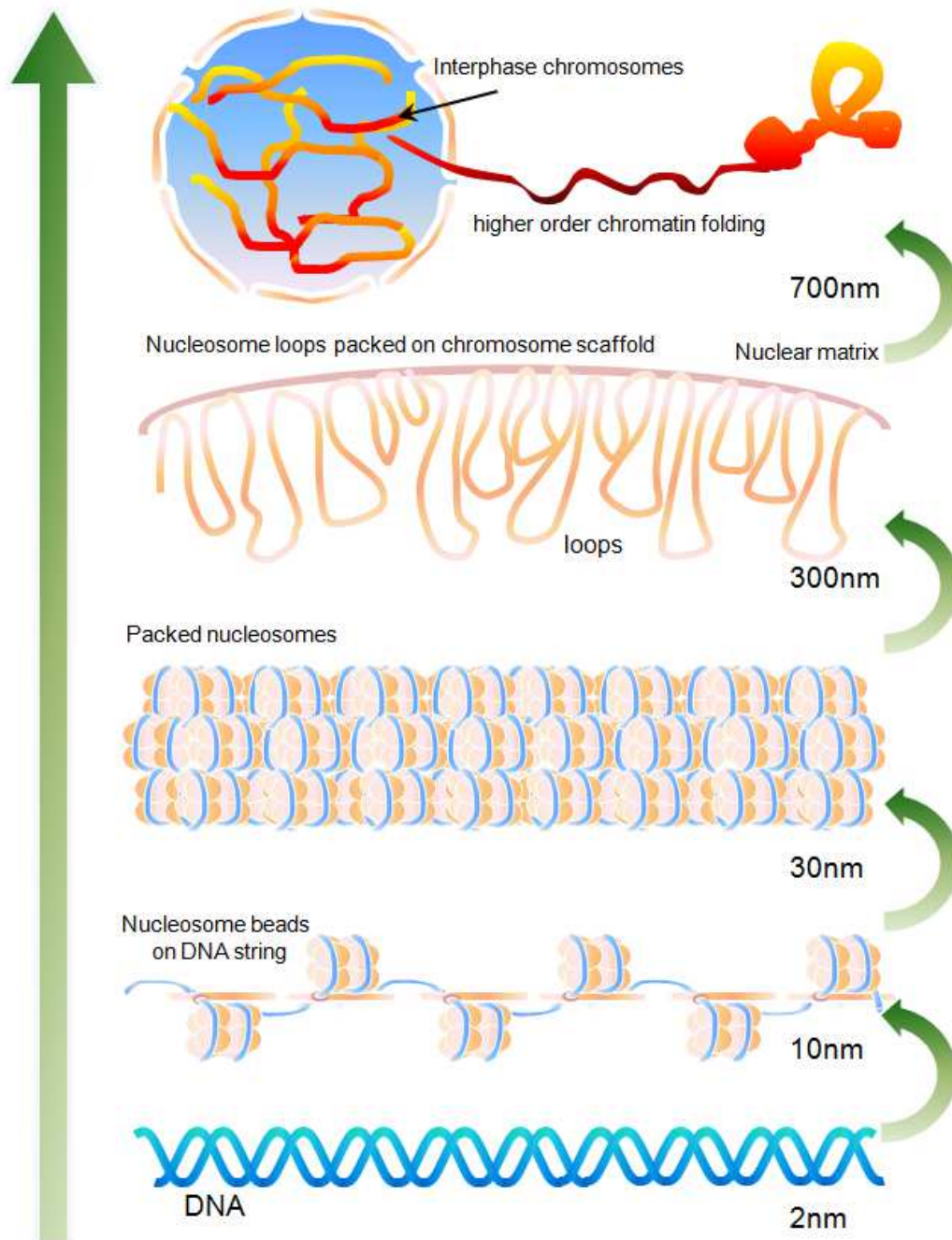


Figure (1.21).chromatin packaging; this figure shows some of the several levels of packaged chromatin to give a highly condensed mitotic chromosome.

When we read this chapter we have to put in our mind that chromosome structure is a dynamic structure, *i.e.*, it's in a diligent non-stopping movement. Not only do chromosomes condense in accordance with the cell cycle, but different regions of the interphase chromosomes condense and decondense as the cells gain access to specific DNA sequences for gene expression (see chapter six), DNA repair, and replication.

Chapter One Appendix

The Variable Applications for Exotic DNA Structure

It has to be noted that double stranded DNA under normal conditions is the rule. However, in certain laboratory conditions, it is possible to induce a third DNA strand to insert itself in the major groove of the double helix structure of normal DNA in a sequence specific manner. That is, the third DNA strand doesn't insert itself anywhere, but it will form stable triplex with sequence specificity (figure 1.22). These binding bases are slightly less accurate than nature, where all sequences can't be recognized, but recognition depends on the surrounding sequences instead. However, thymine in the third strand recognizes adenine in adenine – thymine (T – A•T) base pairing, while cytosine in the third strand recognizes guanine in guanine – cytosine (C – G•C) base pairing.

The first time at which triplex sequences discovered was at 1957 by three scientists; Alexander Rich, David Davis, and Gary Felsenfeld. It seems that such structure has multiple functions in laboratory and clinical fields. Rich was shined among them since he discovered Z DNA structure that it was thought it's just an exotic thing, but in future, it's got the attentions of scientists later since the conversion of B DNA to Z DNA hold great importance in changing nucleic acid behavior in gene expression. The nature of pairing of triplex doesn't follow Watson – Crick base pairing rules in binding with each other. Therefore, they alternatively called Hogsteen pairing, in accordance to Karst Hogsteen who recognized these exotic pairs for the first time at 1963. Thus, these exotic pairs allow triplex DNA formation.

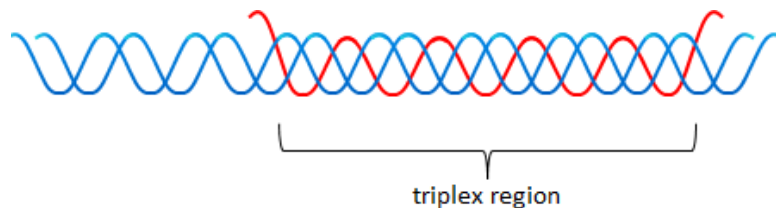


Figure (1.22). Structure of the triplex.

Nor Z DNA neither triplex DNA are the only two exotic structures of DNA, rather, four strands of DNA can pair with each other to form Tetraplex or quadruplex, but this occurs easily only for high Guanine ratio DNA sequences. There is another exotic structure for DNA known as H DNA, is found in polypyrimidine or polypurine tracts that also incorporate a mirror repeat. It produces acute bend in DNA. Such structures are found within regions involved in the regulation of expression of some eukaryotic genes. As it shown above, synthetic DNA strands designed to pair with these sequences to form triplex DNA could disrupt gene expression. This approach to controlling cellular metabolism is of growing commercial interest for its potential application in medicine and agriculture. Moreover, there are many applications for triplex DNA, since only the insertion of a third strand to form three-stranded mode in any gene is enough to disrupt the expression of this gene (see chapter six). This is not the only importance for triplex since it can be used in gene therapy in the same way previously mentioned in order to disrupt any pathogenic or undesired gene.

Chapter One Revision Questions

Q1- Justify the following:

1. It is possible for heat killed pathogenic bacterial strain to infect the normal bacterial strain
2. Avery and his colleagues discovered that DNA is the transforming principle not RNA or protein
3. Viruses in general can't grow outside their host they infect

4. The amount of intracellular DNA is not always a direct measurement for the quantity of genetic material in this organism
5. Prime (') used to numerate the carbon atoms of pentose sugar that exist in DNA
6. Phosphate involves in connecting nucleic acids nucleotides with each other
7. Grooves must exist in DNA molecule
8. We have to read nucleotide sequences in all nucleic acids from 5' to 3'
9. Both strands of DNA are antiparallel to each other
10. Base pairing must exist between nitrogen bases in nucleic acid molecule
11. Temperature required to melt G – C pairs is more than temperature required to melt T – A pairs
12. When the cell wants to divide, its chromosomes must be highly condensed
13. Histones have high ratios of arginine and lysine
14. Histone H1 is the less tightly bound histone with DNA

Q2- choose the correct answer:

1. Chargaff and his colleagues worked out for four years under erroneous hypothesis called
a. mono-nucleotide hypothesis b. di-nucleotide hypothesis c. tri-nucleotide hypothesis d. tetra-nucleotide hypothesis
2. The more ratio of component in DNA molecule, the less temperature needed for its denaturation.
a. A-T b. G-C c. "a" or "b" d. all
3. The common form of the double helix (B DNA) is directed toward direction.
a. right b. left c. anti-clockwise d. zigzag
4. If X-ray beam is shed on DNA molecule according to X-ray crystallography, the resulting picture of this molecule resembles the symbol.....
a. "θ" b. "-" c. "x" d. "φ"
5. In addition to the main DNA that exists in the cellular nuclei or in viruses, there is informational DNA exist in certain organelles such as.....
a- Golgi apparatus b- chloroplast c- cellular vesicles d- ribosomes e- endoplasmic reticulum
6. Nucleotides bind with each other by.....
a- covalent bond b- phosphodiester bond c- glycosidic bond d- hydrogen bonds e- ester bond
7. If human DNA contains 20% for both A and T, the ratio of each C and G are
a- 30% b- 40% c- 50% d- 60%
8. Thomas Morgan got Nobel Prize in regarding to his discoveries correlated with chromosomal roles of inheritance.
a. chemistry b. biology c. biochemistry d. physics and medicine
9. When chromosome resembles X letter, it will occur at phase.
a. interphase b. prophase c. metaphase d. anaphase
10. Histones have high ratios of amino acids.
a. histidine & asparagine b. glutamate & aspartate c. lysine & arginine d. glycine & proline
11. The nature of non-histonic chromosomal proteins is.....which against histonic chromosomal proteins which have charges against the charge of DNA.
a. basic b. acidic c. neutral d. non
12. The scientist discovered nucleosome.
a. Arthur Kornberg b. Roger Kornberg c. Evan Kornberg d. Schneider Kornberg
13. The represent the first level of chromosomal DNA packaging in eukaryotes.
a. beads on thread b. looped domains c. condensed chromosomes d. solenoid structure
14. The less tightly bound histones with DNA are
a. H2A b. H2B c. H3 d. H1
15. Each nucleosome core particle consists of
a. 2H2A & 2H2B b. 2H2A, 2H2B & 2H3 c. 2H2A, 2H2B, 2H3 & 2H4 d. 2H2A, 2H2B, 2H3, 2H4 & H1
16. In chromosomes couldn't be visualized easily and they resemble long fine tangled threads in the nucleus.

- a. interphase b. prometaphase c. metaphase d. anaphase
 17. One of the common amino acids in histones is which is responsible on their strong positive charge.
 a. methionine b. arginine c. alanine d. glycine
 18. The proteins that contributing in the binding between DNA and nucleosome are
 a. actins b. myocins c. tubulins d. nucleoplasmins
 19. the simplest level of chromatin packaging is about nm.
 a. 10 b. 30 c. 300 d. 700
 20. in Chromosomes could not be seen easily, they like long and fine tangled threads of DNA in nucleus.
 a. interphase b. prometaphase c. metaphase d. anaphase e. telophase

Q3- define the following:

Tetranucleotide hypothesis, renaturation, DNA secondary structure, DNA backbone, triplex, nucleoplasmin, linker DNA, chromatosome

Q4- suppose that you have three single stranded DNA belonging to three different DNA double stranded molecules, as follows;

A 5'- GCCGACTCGTCGATCAGCCAG - 3'

B 5'- ATCGAATCATCGATTAGACAT - 3'

C 5'- TCCAGACCAACGGATATCGAA - 3'

Which of the three strands the easiest to be denatured? Why? Which one turns to left? Why?

Q4- connect terms exist in right with its suitable counterparts in left

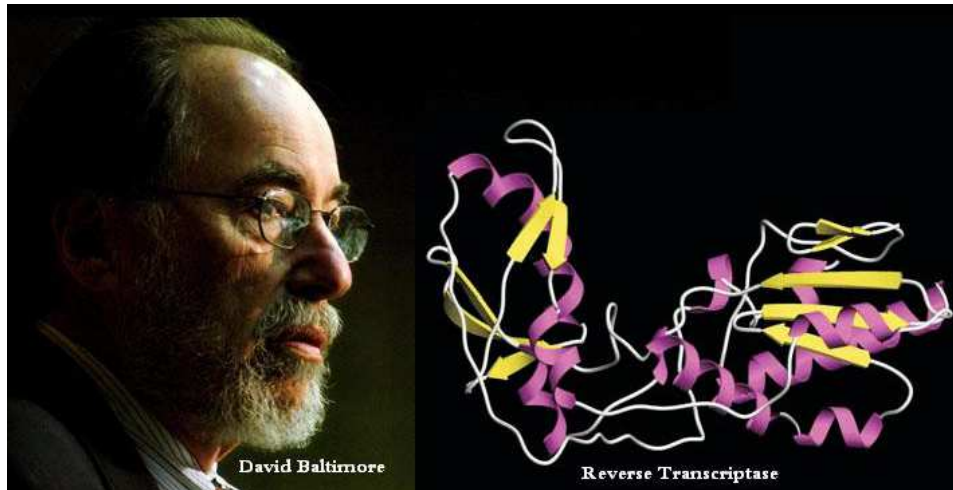
T ₂ DNA	zigzag
H DNA	common
B DNA	acute bend
medina	mitochondria
Z DNA	phage

Further Readings

- Alberts** B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fourth edition, Garland Science/USA. 2002.
Alberts B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
Bloomfield, V. Crothers, D., Tinoco, I. and Hearst, J., 2000. *Nucleic Acids: Structures, Properties, and Functions*. University Science Books.
Bolsover S., Hyams J., Shephard E., White H., Wiedemann C. Cell biology. Second edition, Jhon Wiley and Sons, 2004.
Cooper G., and Housman R. The Cell; a molecular Approach. Fourth edition, ASM Press, 2007.
Darnell J. 1985. RNA *Sci. Am.* 253: (4) 68-78.
Dickerson R. 1983. The DNA helix and how it is read *Sci. Am.* 249: (6) 94-111.
Dickerson R., Drew H., Conner B., Wing R., Fratini A., and Kopka M. 1982. The anatomy of A-, B-, and Z-DNA *Science* 216: 475-485.
Felsenfeld G. 1985. DNA *Sci. Am.* 253: (4) 58-67.
Koolman J., and Rohm K. Color Atlas of Biochemistry. Second edition, Thieme, 2005.
Lodish, H., Berk, A., Zipursky, L., and Matsudaira, P., Molecular Cell Biology (fifth edition). W. H. Freeman and Company, 2000.
Luger K., Mader A. Richmond R., Sargent D., and Richmond T. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolutions. *Nature* 389: 251-260.
Nelson DL, Menninger AL, Cox MM. Menninger Principles of biochemistry. Chapter 14. In: Principles of Bioenergetics. New York: Worth, 2000.
Omoto C., and Lurquin P., Genes and DNA; A Beginner's Guide to Genetics and its Applications. Columbia University Press, 2004.
Passarge E. Color atlas of Genetics. Third edition, Thieme, 2007.
Pierce B. Genetics: A conceptual approach. 2002.
Saenger, W., Principles of Nucleic Acid Structure. Springer Verlag, 1984.
Singer, M., Berg, P. *Genes and Genomes: A Changing Perspective*. University Science Books, 1991.
Tamarin. Principles of genetics. Seventh edition. The McGraw-Hill. 2001.
Verma P., and Agarwal V. molecular biology. First Edition Schandgoup, India (2004).

Chapter Two

Gene Concept



(Left); David Baltimore; when of two discoverers of reverse transcriptase. (Right); reverse transcriptase

Introduction

Genes are the genetic units that contain the information that arranged on the form of particular linear sequences of nucleotides (DNA) which are necessary to code for a particular polypeptide or particular RNA molecule. Polypeptide is a linear arrangements of amino acids connected to each other by the peptide bond. Therefore, “polypeptide” term refers to many peptide bonds in the molecules. Polypeptide chain may connect with other polypeptides to form biologically active protein molecule. Usually, polypeptides are biologically inactive. The two terms “polypeptide” and “protein” are interchangeably used many times, but the crucial point is to understand that proteins normally consist of more than one chain. Usually, genes belong to one of two wide functional classes; structural genes or regulatory genes. This classification depends on the biological function of the end product of genes.

1. Structural genes: code for polypeptide or RNA necessary for the cellular normal metabolism, such as enzyme, structural proteins and receptors. Thus, structural genes are amount of DNA or RNA in some viruses, which contain the necessary information that code for tRNA, rRNA or polypeptide.
2. Regulatory genes: code for polypeptides that form proteins employed in the control of structural genes expression (see chapter six). In their structure, structural genes resemble regulatory genes.

The important thing that should be kept in our mind is that the gene *per se* does not make, manufacture, or produce anything. It is a store of information. But this store is not an inert store away from exciting cellular events, instead, many molecules are acting upon it,

such as RNA polymerase or DNA binding proteins and other molecules to perform many cellular roles in gene expression (see chapter six). Therefore, this store does not act any action and remains in a continuous waiting to be acted upon. However, when a gene occupy a particular location within chromosome, it is possible to possess a particular effect within this location on the morphology or physiology of that organism. It is possible for this gene to be mutated (see chapter seven) or to recombine with other genes (see chapter eight).

The complete set of genes of an organism, what is called the genetic constitution of the organism, is its genotype. The physical or chemical manifestation or the expression of the genotype is the phenotype as in cell morphology and physiology. For example, genes that control eye color in animal included human are considered as part of the genetic constitution of animal genotype. The eye color whether it is blue, hazel or brown is just a chemical or physical manifestation or a result to the action of these genes, this result known as phenotype.

In a particular characteristic, such as a brown eye color, is a part of an organism's phenotype – that is, if it is expressed – it can be said that the individual has the gene for that characteristic. However, if a particular characteristic is not expressed, one cannot conclude that the gene is absent. Again, using eye color as the example, we cannot say that brown eyed individual does not have genes for blue eyes. Because gene expression can be repressed, as happens in bacteria and in high organisms such as animals, one gene may be recessive (not expressing; such as blue eyes) while another is dominant (expressing; such as brown eyes).

Genes can be located on either strands of DNA double helix. But, irrespective of which strand contains a particular gene, all genes are read in a 5' to 3' direction, and the strand containing the particular gene is referred to as the coding strand (see chapter four).

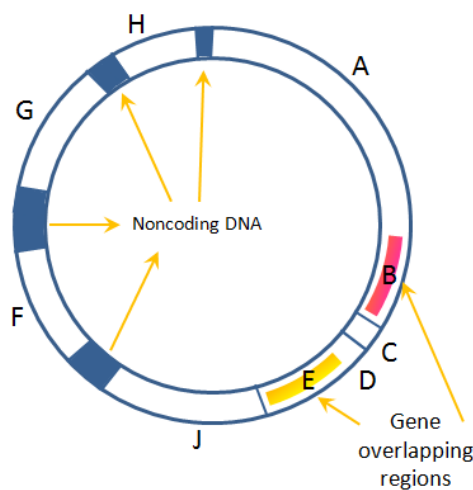
Gene Size

The sizes of the gene have been estimated using various techniques. Direct method to do so is simplified by division of the number of nucleotides in a chromosome by the number of genes known to exist for the organism. Another, but indirect, method is use of genetic mapping and cross-over frequencies. This method depends on the assumption that cross-over frequencies are about equal along an entire chromosome.

The problem is that a high frequency of cross-over can be interpreted to mean that the distance between the genes in question is greater than many in fact be the case since the farther apart genes are the more likely it is that crossing-over will occur. On the other hand, genes that show a low frequency may be seen as being closer together than they really are; these genes may cross over at a low frequency for some unknown reason. So, estimate of genome size and numbers of genes may be larger than is actually the case. The results, however, regardless of the method used, are comparable; the estimate is that an average prokaryotic gene contains somewhere between 900 and 1500 base pairs. The average eukaryotic gene is more difficult to determine, but genes on the order of several hundred thousand nucleotides have been discovered in the eukaryotic organisms. As it is expected, there is a relation between the complexity of an organism and the number of genes in its

genome. For example, numbers of genes are varied between prokaryotic and eukaryotic organisms from less than 500 genes for simple bacteria to 30000 genes in human.

Although the increased number of genes is a clear-cut evidence on the increasing of genetic information in it, but this doesn't mean for the gene when its being small in size it should have little genetic information. Many DNA viruses are very small, such as Φ X174 which needs 11 proteins and the number of their necessary amino acids are 2300. Since each amino acid is encoded by three nucleotides, the length of Φ X174 DNA must be at least 9600 nucleotides regardless of start and stop and other resembling sequences. Since the genome of Φ X174 consists of less than 4500 nucleotides, this microorganism solved this problem by developing what is called overlapping genes. Figure (2.1) refers to genetic mapping for Φ X174, in which overlapping characteristic is obviously clarified. Note, for example, gene B exists within gene A sequence, whereas gene E exists within gene D. This strategy allows containing a lot of genetic information within a limited length of DNA. In this case, it is permitted to each gene to exist within a distinguished length of the genome. That is, adjacent genes require more DNA amount than overlapping genes require.



Inside overlapping genes, overlapping information exists too, but each gene whether being overlapping or no doesn't have its own starting point. Despite of the importance of overlapping system in reducing the amount of necessary gene coding DNA but there are some problems confining this system, which are representing by mutation that takes place in the overlapping region, which doesn't only effect on only one gene but two genes are affected altogether. Consequently, two proteins will be damaged from one mutation in this system instead of one protein. Therefore, there is a duplicated danger of mutation on overlapping genes compared with their non-overlapping counterparts.

Figure (2.1). Nine genes map for Φ X174. Note that gene B is entirely included within gene A, while gene E is entirely included within gene D.

The genetic information

It is noted that information exist in DNA that acting as genetic material are on the form of linear parts called genes. The information resides in particular sequences of nucleotides. Two facts are fundamental here:

1. In the nucleotide of DNA, it is bases, not the sugar-phosphate, that are the information.
2. It is the linear sequence of the bases that is critical in the determination the identity of the gene (through the information it gives).

This can be clarified by resembling nitrogen bases arrangement system in DNA to form particular gene with letters arrangement system in the language to form a particular

word. The combination of different letters forms different words (different information), and the combination of different words forms different sentences (different genes).

Therefore, genes are composed of “words” made up of “letters”. As it shown in chapter one, different nucleotides represented by different letters: A, T, C, and G. in different combinations, those words give different information to the cell. Let us look at to the gene words and talk a little bit away from the words of the language. Hence, we notice that the word is always of three letters, no more, no less.

After we knew what genes composed of, let us focus much more on the behavior of genes in dealing with the cell. Since the gene consists of sequences represented by nitrogen bases, these sequences tell the cell either how to arrange its amino acids so that an almost infinite variety of proteins can be made or how to arrange other nucleotides to construct tRNA or rRNA.

The sequence of nucleotides in genes provides two types of information during protein synthesis. The first is the particular amino acid that the three-letter words in a gene specify. The second is the position each amino acid will occupy in the protein. These pieces of information – which amino acid to use and its position in the protein determine what is called the primary structure of the polypeptide (its linear arrangement of amino acids). The primary structure of the polypeptide ultimately determines the biological activity of the protein that contains the polypeptide. The primary structure determines the secondary, tertiary, and quaternary structures which are the physical shape of the molecule, and the polypeptide’s shape, in turn, determines how the molecule will associate with other polypeptides. And, determine whether the resulting protein having biological activity or haven’t (see chapter five).

Therefore, the final result of the arrangement and combination of informational sequences of gene is the biological activity, which is the mirror that reflects how much these sequences are successful in showing a right image for the resulting protein. Regulatory sequences usually located at the 5’ end of the gene. These nucleotide sequences respond to the chemical or physical signals that originate from inside and outside of the cell. As internal, external of both conditions change, the cell responds to these chemical signals via their interactions with regulatory sequences; these interactions activate or inactivate structural genes. In this way, the types, amounts, and the rate of synthesis of proteins are controlled. These proteins may be necessary inside the cell or on its surface, or they may be exported from the cell (e.g., exoenzymes of bacteria).

Regulatory sequences have no end products, but their function is to provide a recognition sequence for signal molecules and, in consequence of their interaction of the signal molecule and the regulatory sequence, to control the activity of structural genes. The regulatory sequences named according to their function: such as promoters, enhancers, operators, and attenuators (see chapter six).

The structural genes of various organisms, from viruses to prokaryotes to eukaryotes, differ with regard to their informational makeup and even with regard to their instability at a particular location within the DNA molecule. For example, there are genes whose information is contained in a continuous linear arrangement of nucleotides, as is commonly the case with prokaryotes. There are whose information is interrupted by

noncoding sequences of nucleotides, as generally happens in eukaryotes. Other genes have information that can be combined with information from still other genes to create a third gene, as in virus Φ X174 and in human beings. There are even genes that can relocate to different positions on a given chromosome or to another chromosome altogether, as happens in both prokaryotes and eukaryotes (see chapter eight).

Gene Histology

In general, gene consists anatomically of two parts, structural part; which is the part that represents the element to be genetically expressed, such as the various types of RNA molecules originated from template DNA strand. This part begins from transcription start site, and ends with transcription termination site (figure 2.2). The regulatory part consists these segments that specialized only in the regulation of structural part (as in transcription), these elements called non-genetic elements because they are not to be genetically expressed. These sequences are classified into two main classes in prokaryotes: operators and repressors which will be studied with some details in the next chapters (see chapter four, five, and six).

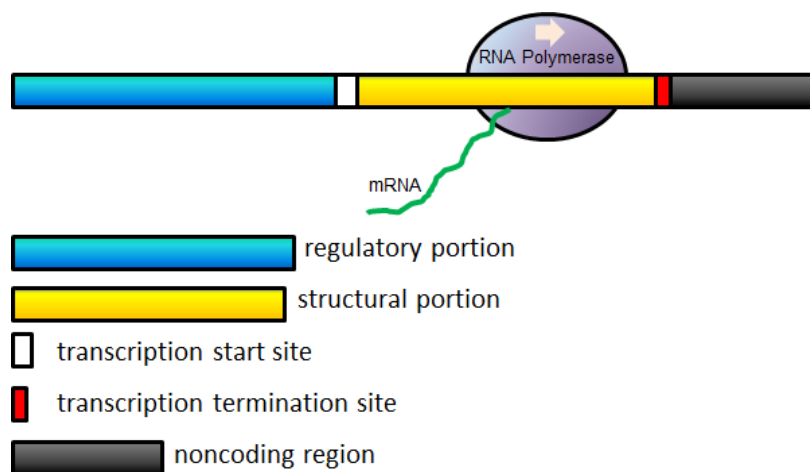


Figure (2.2). Gene histology. Note that gene consists of two portions; the regulatory portion which don't express on itself and structural portion which express's on itself.

Prokaryotic Genes

Many prokaryotic genes consist of non-interrupted nucleotide sequences that are not interrupted by noncoding sequences; though, in some *Eubacteria* and *Archaeobacteria*, noncoding sequences have been found. In general, however, the coding (expressing) nucleotide sequences found in prokaryotes are not interrupted with noncoding sequences. Prokaryotic structural genes are activated on a form of groups or clusters called polycistronic transcription units. In other word, the transcription unit contains information for construction of more than one polypeptide (figure 2.3).

The mRNA transcription units have the following nucleotide arrangements starting at the 5' end of the gene: a leader sequence, a start sequence, a nucleotide sequence for the polypeptide, a stop sequence, and a spacer sequence that consists of 5 to 20 nucleotides to separate one gene from the next one. All these sequences are repeated for each gene

decoded in a polycistronic transcription unit except for the leader sequence, which appears only once in each polycistronic transcription unit. In prokaryotes, mRNA carries more than one message; consequently, mRNA has information for more than one polypeptide.

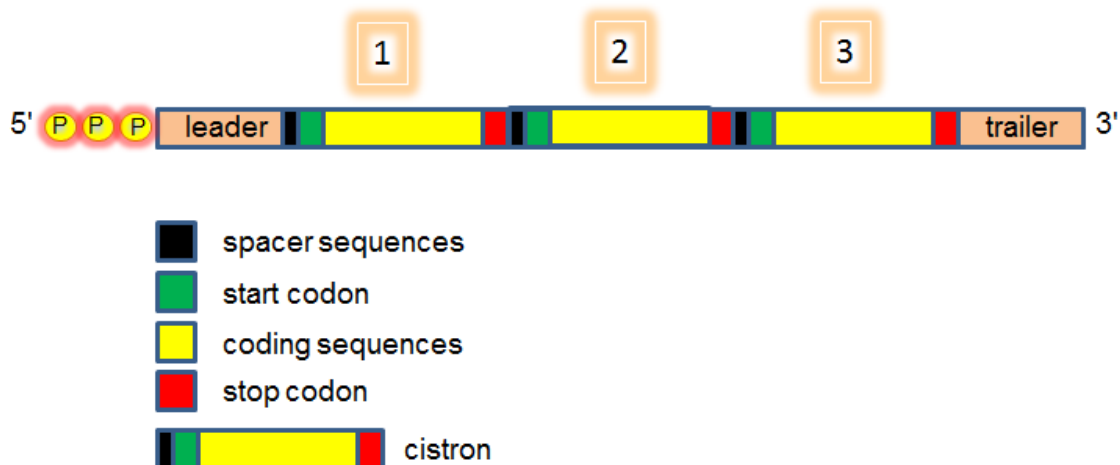


Figure (2.3). Polycistronic mRNA molecule that consists of three cistrons (1, 2, and 3 respectively). This molecule begins from leader sequence at 5' on the right portion of this figure into trailer at 3' end on the left portion of this figure.

Eukaryotic Genes

In contrast to prokaryotes, eukaryotes generally use monocistronic transcription units. The genes of these organisms are not activated in clusters but rather as single entities. Messenger RNA carries information for only single polypeptide.

The eukaryotic gene is most often as arrangement of sequences similar to that of the prokaryotic gene (leader, start, etc.), but these genes also have coding and noncoding sequences of nucleotides. Consequently, the eukaryotic gene is sometimes called a split gene; the information is dispersed along the DNA molecule. There are some rare exceptions to these arrangements. For example, the genes that encode the histone proteins needed to package eukaryotic DNA, in addition to the genes required for the synthesis of alpha and beta interferons, (these proteins that help in protection against viral infection) lack noncoding sequences.

Genes can be classified according to their specialty into housekeeping genes and luxury genes, the first genes code to the essential proteins that may express in every cell, whereas luxury genes (cell specific genes) which also called smart genes code for proteins that exist only in the specialized cells, such as globin, crystalline, fibroin, ovalbumin, casein, immunoglobulin which give an evidence for the complete inhibition in all the cells except the specialized tissues that characterized by their existence.

Noncoding DNA Sequences

It is not an unreasonable assumption to say that the more complex the organism, morphologically and physiologically, the more information it has. And, therefore, the more DNA (or RNA in case of viruses) is needed to store the information. The experimental evidence ruled this assumption out. For example, Φ X174 has a chromosome of 1.8 μ m long and contains approximately a dozen of genes. Lambda phage has almost 10 folds as much DNA and some 50 genes. *Escherichia coli* are estimated to have, 4700 genes in a DNA molecule of about 1.36 cm (that is thousand times as long as the cell). While the DNA of human diploid cells, if its placed end to end, would measure some 2 m, and contain estimated 50,000 – 100,000 genes. The DNA from human chromosomes individually is measured at from 1.5 to 8.5 cm in length.

There are, however, there are some exceptions to these data. The exceptions are referred to as the C-value paradox (C for “content”). The C-value is the amount of DNA in the haploid genome of a species. As noted above, there is a correlation between DNA content and complexity. But, paradoxically, for some organisms, there seems to be no correlation. Although the minimum C-value for eukaryotic species does increase as complexity increases, some organisms have much more DNA than appears to be required. For example, the range of DNA for mammals is between 2 and 3 pg (pictograms: 1 pg = 10^{-12} g, or 1 trillionth of a gram). But, for the amphibian, the range is between 1 and 100 pg; some flowering plants, the lily, for example, have 10 folds the DNA content of a human cell, as does the lungfish.

The current explanation of this phenomenon is that many organisms, including human beings, simply have more DNA they need. The extra DNA is thought to be noncoding DNA. That is, a good deal of the cell’s DNA does not code for protein, RNA, or regulatory information to control gene activity. This information make us ask ourselves the following questions: what is its function? What is the reason of all this noncoding DNA? Why have organisms accumulated so much more DNA than they apparently need for protein and RNA synthesis? Just the amount of DNA varies from species to species so too does the diploid number of chromosomes. For example, human beings have 46 chromosomes, dogs have 78, the fruit fly has 8, the hermit crab has 250, and the potato has 48. The most complex organisms do not necessarily have the most chromosomes (figure 2.4).

However, there are many examples show us the spreading of these noncoding sequences in higher organisms, perhaps the most important of such sequences are introns, tandemly repeated sequences, telomeres, centromeres, and pseudogenes.

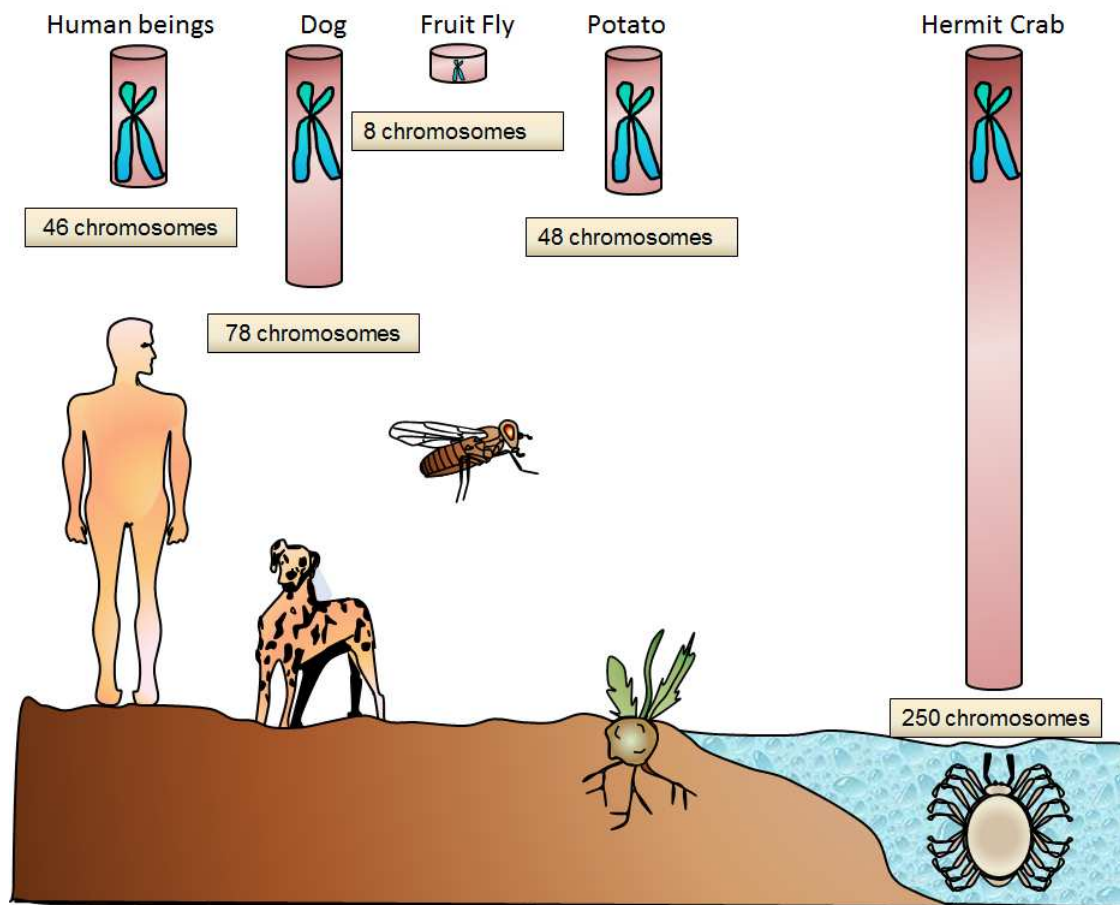


Figure (2.4). Illustration for the absence of direct relationship between number of chromosomes and the level of organism's complexity. It is clear in this figure that it is not necessary for the organism to be much more complex for the sake for his chromosomes to be in higher number. The length of pink columns refers to the number of chromosomes that exist in organism of interest.

Noncoding Sequences

There are several types for these sequences that doesn't code for genes and as follows;

First; Intragenic (Gene Related) Noncoding Sequences

Intragenic sequences are sequences exist within the gene or they are originally genes but not acting for somehow reason. Perhaps, the most common examples for these sequences are:

A. Introns

In bacteria, polypeptide are encoded by DNA sequences in accordance with the amino acids sequences, continuous along DNA template without stopping until all the coding required information are finished. But the notice that say that all genes are continuous has been refuted at 1977 when both Phillip Sharp and Richard Roberts are individually discovered that many polypeptide coding genes in eukaryotes and prokaryotes are interrupted by noncoding sequences called introns.

Amino acids encoding nucleotide sequences (coding sequences) in eukaryotes are called exons which came from the word “expressed”. While the noncoding nucleotide sequences which known introns were came from “*intervening sequences*”.

It is obvious now that there are noncoding long DNA sequences (don't participate in the genetic material that finally translated into amino acids sequences in the protein molecule) are intervening the amino acids coding sequences (exons) in many eukaryotic genes. As it shown above, intervening sequences or introns exist within most but not all higher eukaryotic genes. Eukaryotic precursor RNA copies contain intervening sequences (introns) in addition to exons. Therefore, they called pre-RNA copies. However, introns are cut from the primary transcript (pre-RNA), and exons bind with each other before the resulting RNA molecule is exist in the cytoplasm for translation purpose. This takes place by a process called splicing (see chapter four). Thus, why introns are removed after their existence in the structure of DNA? Some researchers want to verify this by removing introns from simian virus 40 or SV40 (animal infecting viruses having small, circular, and double stranded DNA, hence, it is easy to be studied and manipulated compared with big chromosomes), they found that SV40 intron removed mRNAs are unstable and didn't transport from nucleus to cytoplasm).

Introns function are not known till now but some theoretical biologists have resembled introns with some house stuffs which although nobody use them but nobody through them, and there is no house free of such stuff.

B. Pseudogenes

Pseudogenes are DNA sequences which have considerable similarity with their sequences in the “real” genes that derived from. It was thought that pseudogenes can encode for proteins, but they contain variations in their sequences made them nonfunctional. Pseudogenes don't have introns and contain many stop signals therefore RNA polymerase fails to more for long distances on pseudogenes. Even after the success of transcription of some of those genes and production of mRNA molecules, but these molecules are unable to be translated into active proteins.



Phillip Sharp



Richard Roberts

Compared with mRNA encoding normal genes, pseudogenes contain multiple mutations. It was thought they were accumulated since the first reversely transcribed mRNA molecule and inserted in the genomes of germ cells of old ancestors. These inactive gene copies referred to processed pseudogenes. Most of the pseudogenes are surrounded with short inverted repeats, this in turn support the hypothesis which suggests that they are generated by jumping events (see chapter eight) that include the cellular mRNA. Pseudogenes originated as a result of reverse transcriptase action (see below) which transcribes RNA into DNA double helix (that is known as cDNA or complementary DNA) but it works only in certain types of viruses (as in retroviruses) and after the reverse transcription of mRNA and its conversion to DNA, the resulting DNA inserts in the chromosome in low rate. These inserted copies don't express on themselves because they loss the right sequence that lead their expression (figure 2.5).

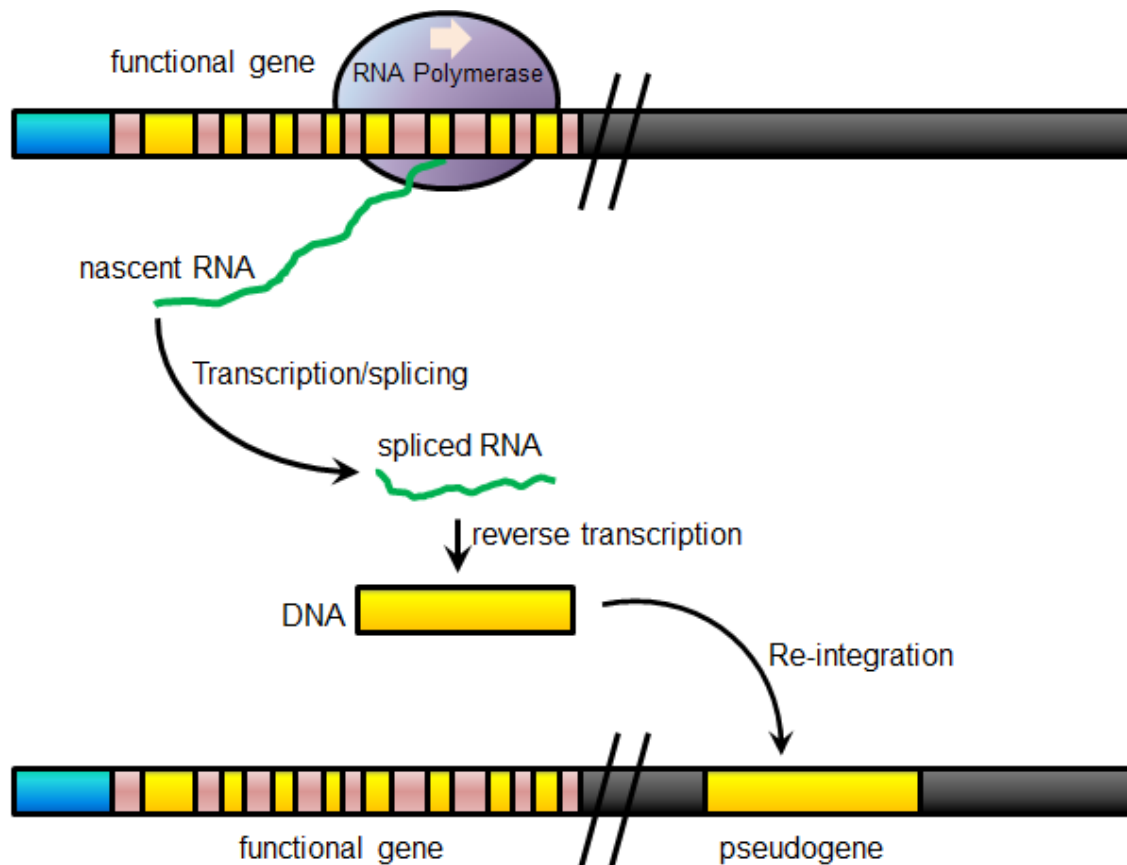


Figure (2.5). Processed pseudogenes arise from integration of reverse transcribed mRNAs. When reverse transcriptase is present in a cell, mRNA molecules can be copied into double stranded DNA. in rare instances, these DNA molecules can integrate into the genome creating pseudogenes. Because introns are rapidly removed from newly transcribed RNAs, these pseudogenes have the common characteristic of lacking introns. This distinguishes the pseudogenes from the copy of the gene from which it was derived. In addition, pseudogenes lack the appropriate promoter sequences to direct their transcription as these are not part of the mRNA from which they are derived.

Second

Extragenic Noncoding Sequences

The most important examples of these sequences are:

A. Repetitive Sequences

The genomes of higher organisms consists of much more DNA than the amount necessary to form the genes since most of this DNA don't encode any genetic function, which is it's considered as noncoding elements. Human genome is an example on this phenomenon. The human genome consists of almost 3 billion base pairs (3×10^9 bp or 3000 Mb) per haploid set of chromosomes. The amount of gene related mammalian DNA is only 30%, whereas the rest 70% is noncoding. The amount of coding DNA ratio in genes is only 3% of the total amount of DNA. The high ratio DNA (70%) consists of repetitive DNA sequences. There are distinct types of these repetitive sequences that are tandem repeats. Depending on their size and pattern, these repeats are classified into several types (figure 2.6); the classic satellite DNA, minisatellites, and microsatellites. All of these sequences constitute only 14% of the total DNA. More than half of human DNA consists of inserted repeats throughout the genome. The most important of these types are long terminal repeats or LTRs, long interspersed nuclear elements or LINEs, short interspersed nuclear elements or SINEs, and transposons (see chapter eight).

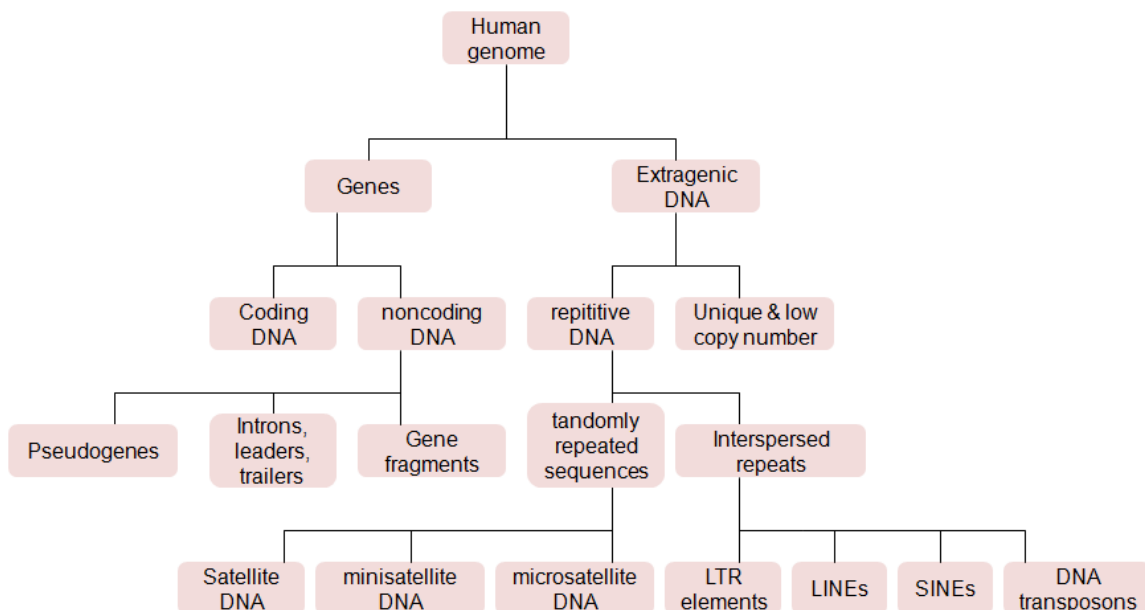


Figure (2.6). The components of nuclear human genome.

When human DNA is exposed to density gradient centrifugation using cesium chloride, the major ratio of DNA forms a band with certain density, but three additional bands (satellites) appears with lower densities because their CG content differs from that of the main DNA (figure 2.7). These three bands contain the classical satellite DNA

(macrosatellites) that consists of 100 – 6500 bps repeats. Microsatellite DNA consists of 10 – 20 bps repeats. It deserves to note that microsatellites are the most frequent repeats exist.

Long interspersed nuclear elements (LINEs) are type of transposons exists in mammals which transpose (jump) by mechanism similar to the mechanism by which retrovirus jump (retrotransposition), but it differs from retroviruses since they don't have long terminal repeats in their structures (see chapter eight). This sort of repeats constitutes more than 70% of human genome weight. As we mentioned, the length of these repeats is more than 6500 base pairs which are rich in adenine only at their 3' ends. While, short interspersed nuclear repeats (SINEs) consist of moderate size repeats with similar nuclear sequences in 300 base pairs average. Their essential structure is consisting of tandem repeats of G – C rich fragments separated by A rich fragments. The most commonly existing sequence in human beings is called *Alu* family. It is called so because they have a sensitive site for *Alu* restriction enzyme.

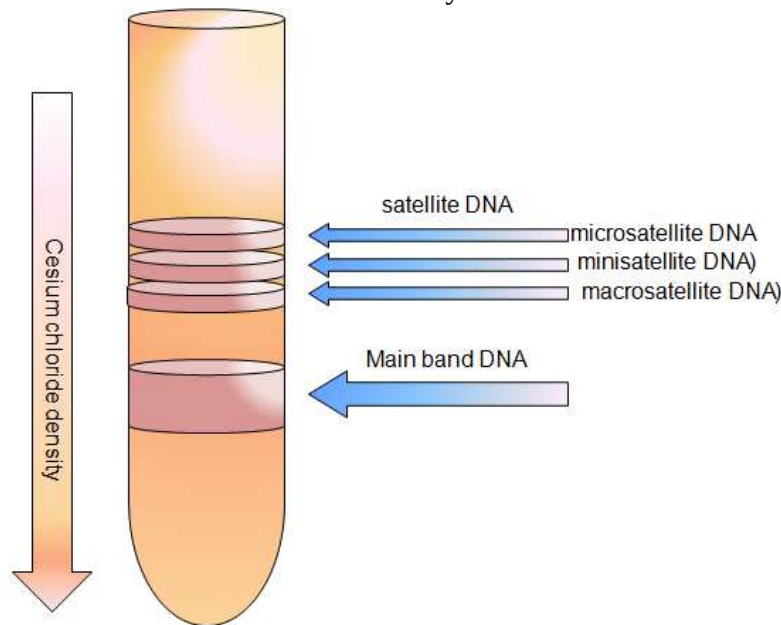


Figure (2.7). Illustration for the human genome discovered satellites according to density gradient centrifugation.

B. Unique DNA sequences

Each linear chromosome forming DNA molecule must contain on the middle part (centromere), two of terminal parts (telomeres), and replication origins. But, before we explain the meaning of telomere and centromere we have to give simple basics about the different states of chromosomes during cell cycle, in which chromosomes are replicate during interphase and become highly condensed during mitosis in order to be separated and distributed into two daughter nuclei. Highly condensed chromosomes in mitotic cells is called mitotic chromosomes (figure 2.8), this shape of chromosome is the most predominant form that can easily be seen.

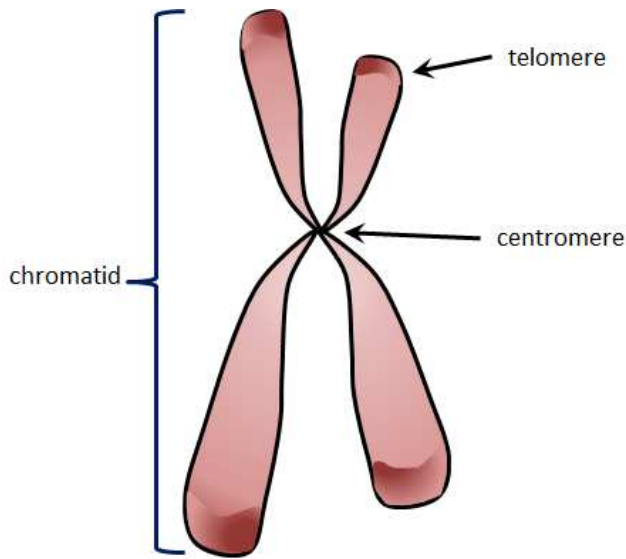


Figure (2.8). The appearance of typical metaphase chromosome under microscope. Each chromosome consists of two chromatids; each chromatid consists of one of the two identical DNA molecules. The centromere; the place at which the two chromatids bind, it is necessary to separate them in the last stages of mitosis. Telomeric sequences protect the extreme parts of chromosomes and other functions.

Indeed, all shapes of chromosomes that shown in many books are mitotic chromosomes. Instead, the appearance of these chromosomes in this form doesn't last very long. This condensed state is very important to allow replicated chromosomes to separate by mitotic spindle. However, this stage doesn't stay approximately half an hour in the whole 24 hours of human cell cycle. As for the rest 23 hours and half, the DNA exists as tangled threads. Through stages of cell cycle when the cell is not in the dividing state (interphase), chromosomes stretch and many chromatins exist as fine and long tangled threads in the nucleus. Thus, chromosomes could not be seen easily. These chromosomes in this extended state are called interphase chromosomes.

Chromosome act as significant structural unit to make one copy of it pass to each daughter cell during division. Each chromosome must be capable on replication, and the newly replicated copies must be separated later and segmented properly into two daughter cells. These main functions are controlled by three different types of DNA specialized nucleotide sequences, each of which binds with specific proteins that lead the machine that replicate and segregate chromosomes (figure 2.9).

The first type of nucleotide sequences acts as replication origin, it is the site at which DNA replication start. Eukaryotic chromosomes consist of many replication origins to ensure each the rapid replication of the whole chromosome. After replication, daughter chromosomes remain connected with each other. With the proceeding of the cell cycle, chromosomes much more condense to produce mitotic chromosomes. The presence of second specialized DNA sequence which called centromere allows only one copy of each replicated and condensed chromosome to pull out into each daughter cell during cell division. The protein complex that forms at centromere is called kinetochore; it connects the replicated chromosomes to mitotic spindle, to allow these chromosomes to pull apart from each other (figure 2.10).

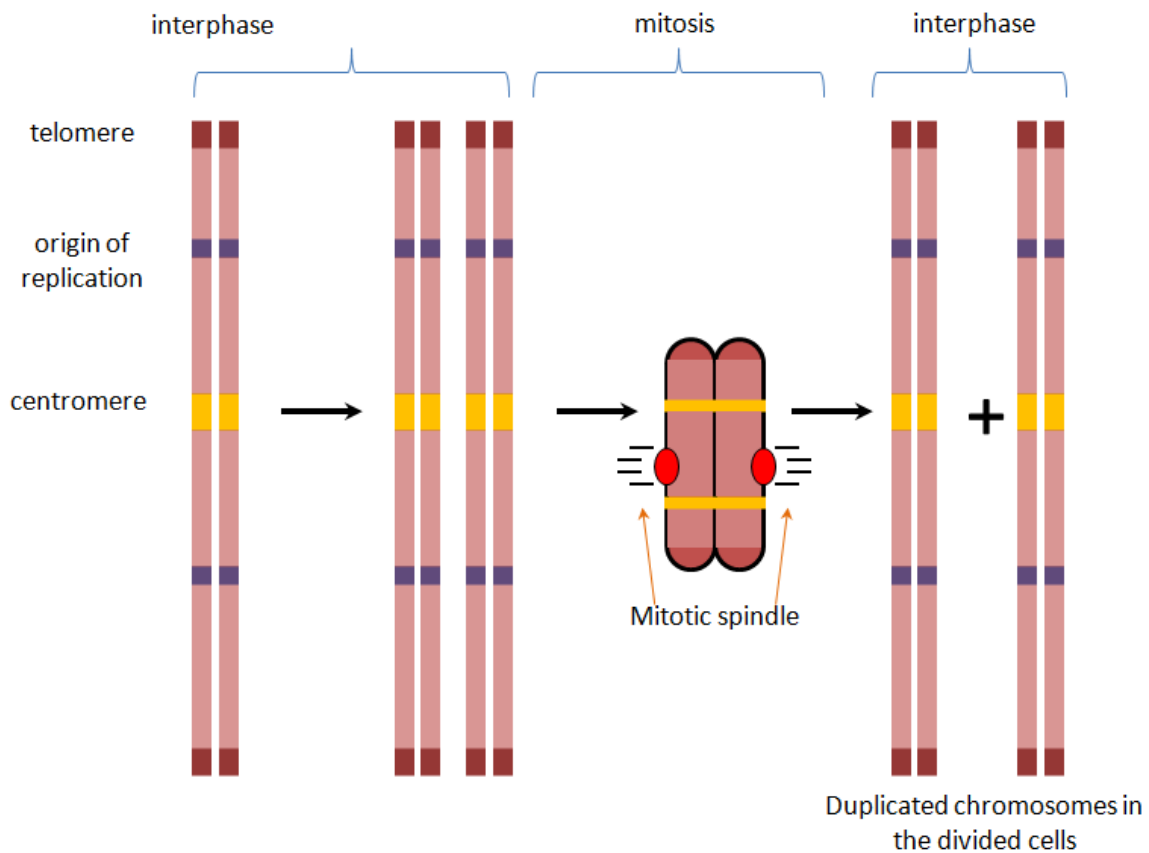
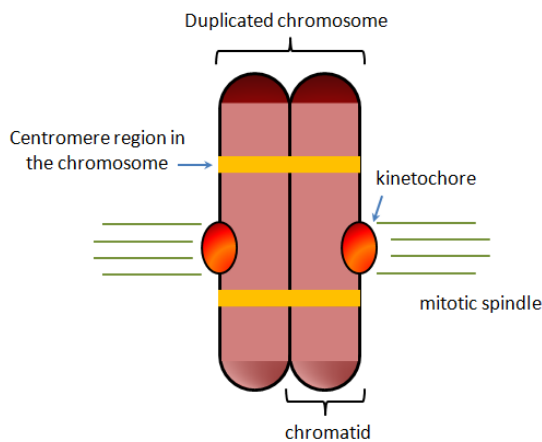


Figure (2.9). The three DNA sequences required to produce a eukaryotic chromosome that can be replicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. Shown here is the sequence of events a typical chromosome follows during the cell cycle. The DNA replicates in interphase beginning at the origins of replication and proceeding bidirectionally from the origins across the chromosome. In M phase, the centromere attaches the duplicated chromosomes to the mitotic spindle so that one copy is distributed to each daughter cell during mitosis. The centromere also helps to hold the duplicated chromosomes together until they are ready to be moved apart. The telomeres form special caps at each chromosome end.



The third specialized type of DNA sequences forms what is known telomeres (*telo* is a Greek word means "terminal", and the word *mero* means "parts"), which are the extreme parts of chromosomes. The telomere consists of nucleotide sequences enable chromosomal terminal parts to replicate efficiently.

Figure (2.10). Connection of kinetochore complex between centromere and mitotic spindles.

Telomeres perform another function through their cooperation with certain DNA binding proteins to form structures protect the chromosomal extreme parts in order not to be recognized by the cell as a broken DNA need to be repaired. In general, when telomere binds with such a knob like proteins nucleoprotein complex forms in order to make digestive enzymes incapable on breaking the extreme ends of chromosomes. This let some molecular geneticists to compare these terminal parts with the brilliant piece that exist at the edge of shoes connector (shoelace).

The Flow of Genetic Information

If the amino acids sequence that exist in a protein is dictated by nucleic acid sequences exist in DNA, how for a molecule like DNA to control protein synthesis in cytoplasm while they are localized at their places in the nucleus? small molecules replied this question since are capable of transporting themselves from nucleus (place of DNA; the source of information) into ribosomes in cytoplasm (where protein synthesis takes place). Since RNA molecules are essentially located at cytoplasm, it is rational to consider RNA as intermediate substance carrying genetic information from DNA into ribosomes. Thus, there is a path for flow of genetic information known as central dogma of molecular biology which is summarized in (figure 2.11).

According to this concept, RNA is synthesized from DNA template in a process called transcription, and protein is synthesized from RNA template in a process called translation. Genetic information existing on DNA that localized in nucleus is copied into specific nucleotide sequences in mRNA. Then, this information located on mRNA in the cytoplasm are translated into specific amino acid sequences in protein.

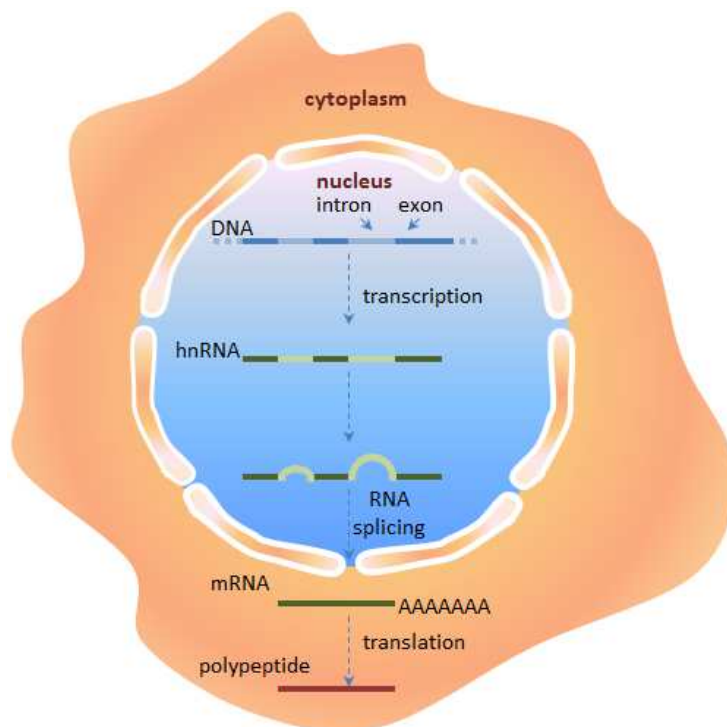


Figure (2.11). Central dogma of molecular biology. Note the classical flow of genetic information from DNA into RNA through transcription and into protein through translation.

Reverse Transcription

According to central dogma of molecular biology, cells contain both DNA and RNA, in which there is only one path for the flow of genetic information from DNA into RNA in a process called transcription, and from RNA into protein in a process called translation. It was thought during 1950s and 1960s that genetic information could be transferred from DNA into RNA and not from RNA into DNA. Viruses (which differ from cells) contain either DNA or RNA. As for DNA containing viruses, the above mentioned rule of the one path of genetic information flow can only be applied without difficulty. What happen in case of RNA containing viruses?

One of the important groups of animal viruses include retroviruses; their genetic material consist of single stranded RNA that can act like mRNA but it don't act so, instead, they act as template to synthesize double stranded DNA. In other meaning, these viruses exert reverse flow of genetic information, this means, from mRNA into DNA.

At 1963, Howard Temin showed that replication of retroviruses (RNA containing viruses) was inhibited by actinomycin D which bind with DNA only. While the replication of other RNA viruses was never inhibited by this antibiotic. As long as this data was confusing the scientific society at that time since it was not concomitant with the central dogma, that's why it was ignored until 1970 when both Howard Temin and David Baltimore published that retroviruses consist of RNA dependent DNA polymerase (or the enzyme that often referred as reverse transcriptase). This type of reverse flow of genetic information is called sometimes "Teminism" according to its discoverer. The sort of the reverse flow of genetic information might be against the central dogma. This was discovered by Temin and Baltimore independently at 1970. H. Temin and D. Baltimore received Nobel prizes for their discovery of this polymerase enzyme. This enzyme is involved in a tumor virus's infection of a normal cell and the transformation of that cell into a cancerous cell. When the viral RNA enters a cell, it brings this enzyme with it (figure 2.12). This enzyme synthesizes a DNA-RNA double helix, which then is enzymatically converted into a DNA-DNA double helix that can integrate into the host chromosome. After integration, the DNA is transcribed into copies of the viral RNA, which are both translated and packaged into new viral particles that are released from the cell to repeat the infection process.



Howard Temin



David Baltimore

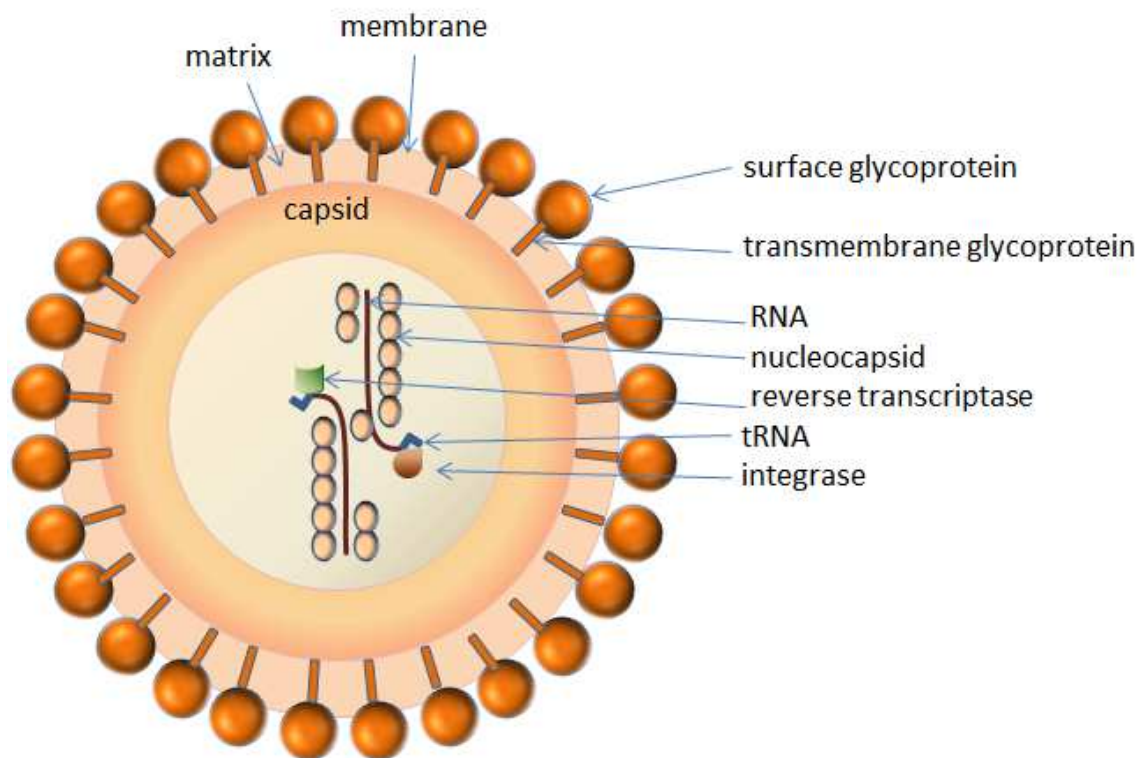


Figure (2.12). Structure of retroviruses. RT means reverse transcriptase. Retrovirus converts RNA into DNA through reverse transcriptase that born by the same the virus itself. Integrase inserts the virus in the genome of the host. NC refers to nucleocapsid which surrounds the genetic material of the virus. TM refers to transmembrane glycoprotein.

RNA Self-Replication

The second modification to the original central dogma is the verification that RNA can act as a template for its own replication. This process has been observed in a small class of phages. These RNA phages, such as R17 and MS2, are the simplest phages known. MS2 contains about thirty-five hundred nucleotides and codes for only three proteins: a coat protein, an attachment protein (responsible for attachment to and subsequent penetration of the host), and a subunit of the enzyme RNA replicase. The RNA replicase subunit combines with three of the cell's proteins to form RNA replicase, allowing the single-stranded RNA of the phage to replicate itself. Since the new protein needed to construct the RNA replicase must be synthesized before the phage can replicate its own RNA, the phage RNA must first act as a messenger when it infects the cell. Thus, protein synthesis is taking place without a preceding transcription process. The viral genetic material, its RNA, is first used as a messenger in the process of translation and then used as a template for RNA replication.

DNA Involvement in Translation

In the mid-1960s, B. J. McCarthy and J. J. Holland showed that under certain experimental conditions, denatured (single-stranded) DNA could bind to ribosomes and be

translated into proteins. The experimental conditions usually involved the addition of antibiotics that interacted with the DNA or the ribosome. Direct translation of DNA is not known to occur naturally. Even in our updated central dogma (figure 2.13), no arrows originate at the protein. In other words, protein cannot self-replicate, nor can it use amino acid sequence information to reconstruct RNA or DNA. But additional exception might be updated on the central dogma of molecular biology due to the presence of an exceptional protein called "prion" that many papers referred to its ability to reverse the direction of the central dogma. That's why Reed Wickner, one of the specialist in the prion field comments on this subject saying: "as long as nucleic acids can act as catalyzers for enzymatic reactions (see chapter four appendix), proteins can act like genes". Here, another exception is emerged which is the role of some proteins such as prion (see the appendix of this chapter) in the addition of another confusion to the central dogma of the flow of genetic information.

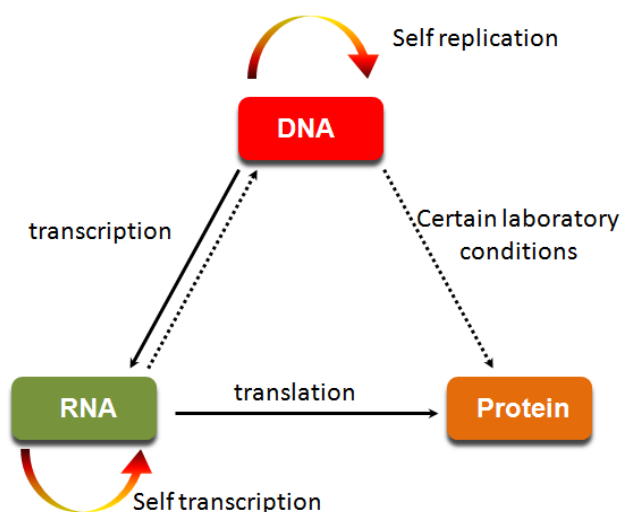


Figure 2.13. An updated version of Crick's central dogma, showing all known paths of genetic information transfer. Paths confirmed since Crick proposed the original central dogma appear as dashed red lines (reverse transcription, RNA self replication, and direct DNA translation). Direct DNA translation is known only under laboratory conditions: the process apparently does not occur naturally. There is no known information flow beginning with protein.

Chapter Two Appendix

Application of Prion and Its Interesting Role in Understanding the Central Dogma

Prion was nominated by Stanely B. Prusiner at 1984 and he received Nobel Prize in physiology and medicine at 1997 for his discovery for the features of prion proteins. The idiom prion is derived from "*proteinaceous infectious particles*" which means "infectious protein particles"; which are infectious agents may consist of singular type of protein molecules without any nucleic acid, and it can tolerate temperature, UV, and ionizing radiation in addition to their resistance to the viral damaging chemical treatments (such as the resistance to phenol or to nucleic acids digesting enzymes). The molecular weight of prion is estimated from 50000 to 100000 Dalton, this weight is one hundred folds smaller than the smallest viruses.

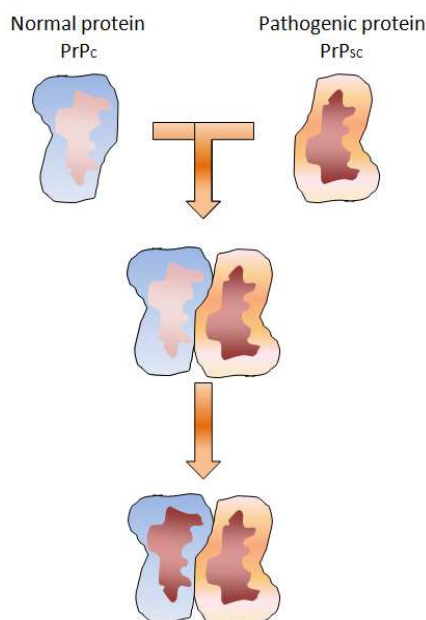
These infectious agents are connected with the "slow" viral diseases such as Creutzfeldt-Jacob disease in human, scrapie in sheep, and BSE; bovine spongiform encephalopathy (mad-cow disease) in cattle. All of these diseases are fatal, and all of them were caused by the digestion of the proteins of patient or from a mutation in the normal gene. There is no mentioned treatment for any of these diseases. The most important feature of scrapie is that nobody capable to get any biochemical evidence for the existence or involvement of any nucleic acid in this disease. At beginning, Prusiner



Stanely Prusiner

hypothesis is fiercely criticized since he assumed that prion proteins are entirely consisting of protein lacking any trace for nucleic acid and they replicate without any genes. One of the principles of modern biology is that all living things have information inherited in the form of DNA or RNA. Thus, how prion is capable of reproduction without nucleic acid? Or in another word, how the seemingly genetic material free protein is capable of causing a transmissible disease when it is eaten by mammal and human?

Prusiner assumed several mechanisms through which the infective protein is can trigger the conversion of normal proteins to be infective proteins. One of these mechanisms involves a successive series of events in which the infective protein PrP^{Sc} is connected with the normal protein PrP^C, to produce two infectious proteins of PrP^{Sc} type. Also, one of these two proteins in this complex produces two proteins, the two produce four and so on. A sort of confusion is emerged from the fact of the existence of this protein and the gene that encodes this protein in normal cells "non-infecting". Despite of that nobody knows the role played by the normal PrP protein; it is noticeably expressed by Prn^P gene which encodes it. When PrP^{Sc} (the infecting form) exist it interacts with PrP^C and causes the folding of the later into the disease causing protein form (figure 2.14). Consequently, the infection by PrP^{Sc} converts the normal PrP protein into the abnormal PrP which forms the prion.



It seems that the accumulation of PrP^{Sc} in brain is responsible about the nerve diseases caused by prion. This explanation of prion disease is called "protein only hypothesis" which is confirmed later by many experiments. Despite the ambiguity of the replication and reproduction mechanism of prion, it is thought that prion is acting somehow as a template to perform reverse translation, which means the conversion process from protein into RNA into DNA. By this way, it converts the central dogma of molecular biology upside down!

Figure (2.14). Structural changes in the normal prion PrP^C by infecting prion PrP^{Sc}. Unfavorably folded infecting prion is refold the normal prion when its connecting with it. The conformation of normal prion is changed from α helix into β sheet. There is a tendency of the substitute second conformation to be accumulated with each other, this forms brain damaging agglomerations

Chapter Two Revision Questions

Q1- Justify the following:

1. Although introns are fundamental part in the structure of eukaryotic genes, they are removed after transcription?
2. Overlapping genes phenomenon is dangerous on gene system?
3. If the amino acids sequences of protein is dictated by nucleic acid sequences in DNA, how a molecule like DNA control protein synthesis in the cytoplasm where it located in its place in the nucleus?
4. Eukaryotic chromosomes consist of many replication origins
5. Some molecular biologists resemble telomere with shoe lace?

Q2- choose the correct answer:

1. Genes are classified into two classes; structural genes and regulatory genes according to
a. function b. shape c. size d. structure
2. we have to understand that the gene is not

- a. active restore b. inactive restore c. restore close to cellular events d. not restore
3. the eye color is an example on
- a. genotype b. phenotype c. both d. neither a nor b
4. in DNA nucleotides is the information that represent the gene.
- a. phosphate b. sugar c. nitrogen base d. all
5. gene can be resembled with
- a. letter b. word c. sentence d. paragraph
6. the linear arrangement of amino acids is called
- a. primary structure b. secondary structure c. tertiary structure d. quaternary structure
7. molecule contains introns.
- a. mRNA b. snRNA c. pre-RNA d. none
8. genes are usually consisting of.....
- a. DNA b. RNA c. protein d. "a" or "b" e. none
9. prokaryotic genes consist of non introns interspersed nucleotide sequences except
- a. *Archaeobacteria* b. *Eubacteria* c. *Helicobacteria* d. *Mycobacteria* e. only "a" & "b"
10. Φ X174 phage contain
- a. 1000 genes b. split genes c. overlapping genes d. only "a" & "b" e. none
11. one of the most important reasons for retrovirus popularity in molecular biology is its containing on enzyme.
- a. DNA polymerase b. DNase c. RNA polymerase d. RNase e. Reverse transcriptase
12. The hypothesis that explains the reason behind mad-cow disease is
- a. DNA only hypothesis b. RNA only hypothesis c. Protein only hypothesis d. only "a" & "b" e. none

Q3- define the following:

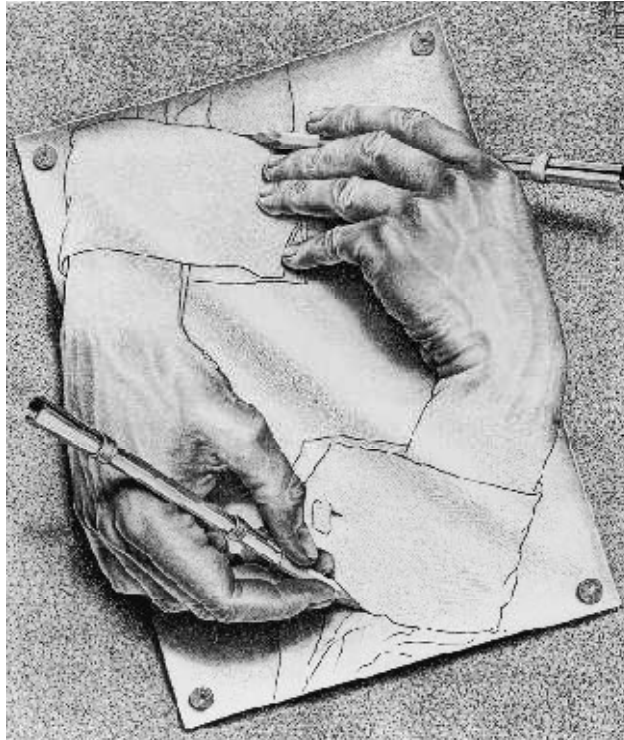
gene, polypeptide, phenotype, split genes, Teminism, central dogma of molecular biology, smart genes, structural genes, housekeeping genes, pseudogenes, C value paradox, polycistron, kinetochore, mitotic chromosomes

Further Readings

- Alberts B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fourth edition, Garland Science/USA. 2002.
- Alberts B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Bolsover S., Hyams J., Shephard E., White H., Wiedemann C. Cell biology. Second edition, John Wiley and Sons, 2004.
- Branden, C., Tooze, J., 1999. Introduction to Protein Structure (2d ed.). Garland, 1999.
- Cann A. Principles of Molecular Virology. Fourth edition, Elsevier Academic Press, 2005.
- Carter J., and Saunders V. Virology: principles and applications. John Wiley and Sons, 2007.
- Clark D. Molecular biology: understanding the genetics revolution. Elsevier, 2005.
- Cochet, F. Gannon, R. Hen, L. Maroteaux, F. Perrin, and P. Chambon. 1979. Organization and sequence studies of the 17-piece chicken conalbumin gene. *Nature* 282: 567-574.
- Creighton, T. E. Proteins: Structures and Molecular Principles (2d ed.). W. H. Freeman and Company, 1992.
- Davis T., and Kipling D. Telomeres and Telomerase in Aging, Disease, and Cancer. Molecular Biology of Human Cancers An Advanced Student's Textbook - Wolfgang A. Schulz, 2008.
- Dorit, R. Schoenbach, L. and Gilbert. W. 1990. How big is the universe of exons? *Science* 250: 1377-1382.
- Krishna R., and Wold F. 1993. Post-translational modification of proteins *Adv. Enzymol. Relat. Areas. Mol. Biol.* 67: 265-298.
- Lewis R. Human Genetics: concepts and applications. Ninth edition, McGraw-Hill, 2009.
- Lodish, H., Berk, A., Zipursky, L., and Matsudaira, P., Molecular Cell Biology (fifth edition). W. H. Freeman and Company, 2000.
- Lupsky J. R., and Stankiewicz P. Genomic Disorders. 2006, Humana Press Inc.
- McCarthy B., and Holland J. (1965). Denatured DNA as a direct template for in vitro protein synthesis. *PNAS USA* 54: 880 – 886.
- Paoletta, P. Introduction to molecular biology. First edition, WCB/McGrawHill, 1998.
- Passarge E. Color atlas of Genetics. Second edition, Thieme, 2001.
- Pierce B. Genetics: A conceptual approach, 2002.
- Schultz, G. E., and Schirmer, R. H., Principles of Protein Structure. Springer-Verlag, 1979.
- Sharp. P. 1988. RNA splicing and genes *J. Am. Med. Assoc.* 260: 3035-3041.
- Tamarin. Principles of genetics. Seventh edition. The McGraw-Hill. 2001.
- Tilghman, S. Tiemeier, D. Seidman, J. Peterlin, B. Sullivan, M. Maizel, J. and Leder. P. 1978. Intervening sequence of DNA identified in the structural portion of a mouse β -globin gene *PNAS USA* 75: 725- 729.
- Verma P., and Agarwal V. molecular biology. First Edition Schandgoup, India 2004.
- Watson J., Baker T., Bell S., Gann A., Levine M., Losick R. Molecular biology of the gene. Fifth edition, Pearson Prentice Hall, 2004.

Chapter Three

DNA Replication

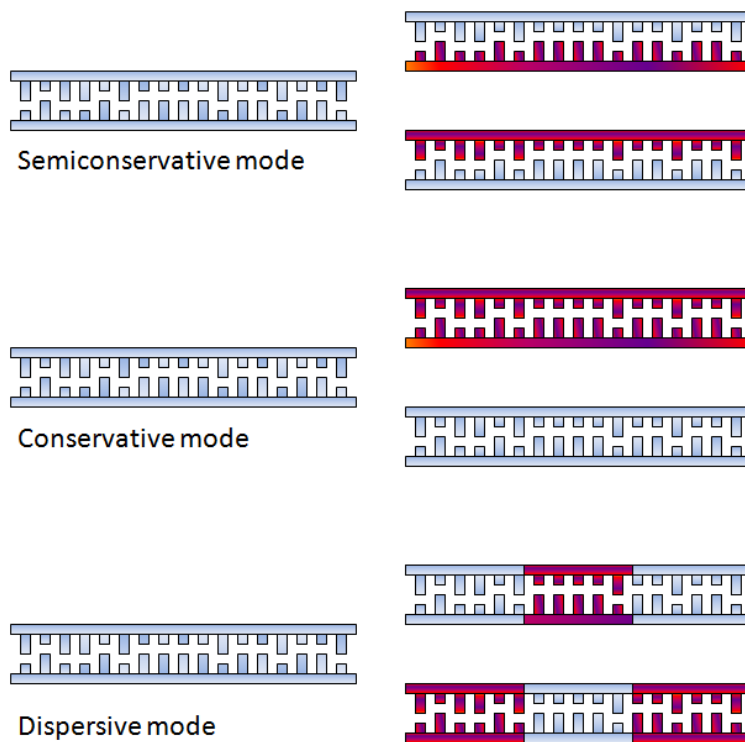


Analogous portrait simulate replication process: these two hands complement each other and the same situation for DNA molecule which contain two strands complementary to each other to form replication bubble in order to be extended through template strands and draw their path forward as in this portrait (Lewis, 2009).

Introduction

All organisms have to duplicate their own DNA with infinite precision before each cellular division. In this chapter we will study how "replication machinery" achieves this task with this precision. This can be taken place by replicating DNA with a velocity of more than 1000 nucleotides per second. The unwinding of DNA double helix is a fundamental step in replication and transcription. When DNA double helix replicates, each chain act as template to synthesize the complementary chain. Hence, base pairing mechanism suggests that the two new DNA strands are copied from the two old strands. Although this mechanism provides precise copying for genetic information, but it questions about; dose replication conservative or semi-conservative?

In the first mechanism, conservative mode, both parental strands form new double helix while the old double helix remains intact. In the second mechanism, semi-conservative mechanism, each old strand pairs with the new strand that copied from it. The second mechanism called semi-conservative since it conserve only one strand of the



parental DNA in each resulting double helix, that is, half helix is from one of the two old helices, and the another half is from one of the two resulting new strands. As for the third mechanism, dispersive mode, both two strands are dispersed into fragments, which serve as template to synthesize new fragments of DNA, and then somehow reassemble into two complete DNA molecules. In this model, each resulting DNA molecule is interspersed with fragments of old and new DNA; none of the original molecule is conserved (figure 3.1).

Figure (3.1). The three hypotheses for DNA replication, which include semi-conservative, conservative, and dispersive mechanism. Blue color refers to the parental strands, while red color refers to daughter strands.

Although semi-conservative replication doesn't require to a predictive mechanism to replicate DNA molecule, it requires to a high moiety of proficiency to invent an experiment to prove this for the first time. This invention was come from both Meselson and Stahl at 1958 when they imagined a method to prove the semi-conservative mode of replication of genetic material (genome) of *E. coli*. Their work was relied on utilization of isotopes that produce a DNA with modified densities after replication. Thus, they cultivated *E. coli* cells for several generations on a medium containing of ammonium ions (NH_4) as a source of nitrogen that DNA need (in addition to protein) in its synthesis in which all of its nitrogen have heavy isotopes ^{15}N (the normal nitrogen is ^{14}N). After cultivation of *E. coli* for several generations in this medium, they found that *E. coli* DNA was heavier than normal DNA because the existence of nitrogen atoms ^{15}N in it, in which nitrogen isotope ^{15}N was incorporated in DNA bases during replication. Consequently, the DNA of the progeny cells possesses more density compared with its normal counterpart. After this, they transferred ^{15}N saturated medium bacteria in order to grow in a



medium containing normal nitrogen (^{14}N), and they allowed cells to replicate for only one generation at beginning. After that, they isolated DNA from cells and they analyzed it by a technique permitting to identify DNA density called density gradient centrifugation that performed with cesium chloride (CsCl). The solution is placed in ultra-centrifuge which rotates with high velocities for many hours. Ultimately, equilibrium between centrifugation force and diffusion is taken place, such graduation in density occurs in a tube in which an increase of CsCl concentration is taken place from the top to the bottom of the tube. If DNA or any other material is added it will concentrate and form a band in the tube at a point in which its density is equal to CsCl density. If there is many types of DNA with different densities, different bands will be formed. In this experiment, the pure ^{15}N DNA results in single band of DNA, and the same thing for ^{14}N DNA (figure 3.2).

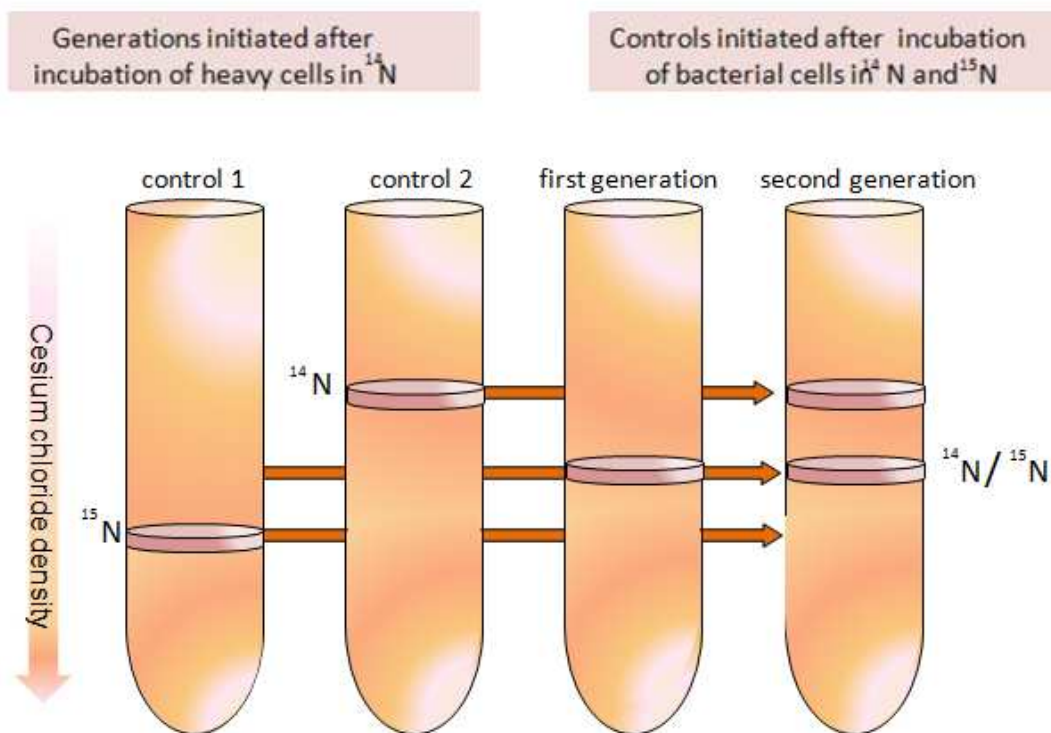


Figure (3.2). Locations of DNA bands resulting from the difference in nitrogen density in first and second generations and their comparison with heavy nitrogen control and light nitrogen control.

The only difference is represented by the fact that the more density the DNA has the more bottoms the band localizes in the tube. Thus, the position of DNA in the tube gives an indication for us to locate or to monitor DNA density. It is possible to detect bands by looking at tubes under UV using 260 nm wavelengths at which nucleic acids adsorb strongly). However, when the new generation of DNA is analyzed, it was found that it occupies a middle position exactly between the light band (new generation band) and heavy band (ex-generation band). This in itself was not surprise, because it tells us that no more

than half nitrogen atoms in new DNA are ^{14}N and the other half are ^{15}N , and the only visible band in the isolated DNA is this band that corresponds $^{14}\text{N} - ^{15}\text{N}$ DNA (figure 3.2) since the replication is semi-conservative. If replication is conservative, two bands will appear in first generation of replication which represent the original ^{15}N DNA and the new ^{14}N DNA.

This is not all the story since through successive generations, if replication mode is conservative, the original DNA will continue appearing representing by ^{15}N band. This is not happened off course. If replication mode is dispersive, the result is the appearance of multiband mode according to the degree of disperses.

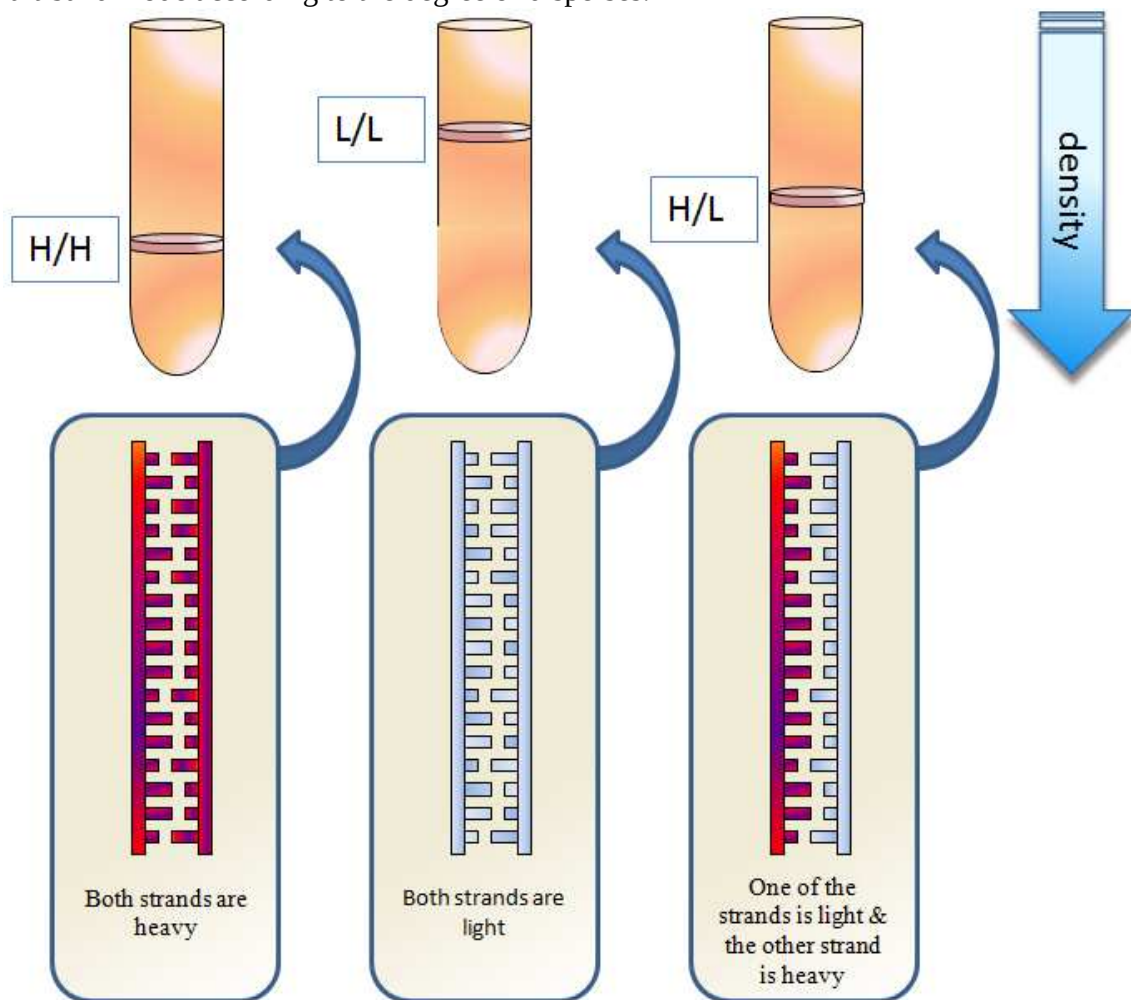


Figure (3.3). Nature of DNA strands represented by different bands in centrifugation tube.

The results obtained in figure (3.3) are completely concomitant with semi-conservative mode of replication and only semi-conservative mode, since the results in successive generations are in accordance with this deduction. Thus, the results of DNA analysis of the second generation growing in ^{14}N medium showed equal amounts of

between hybrid DNA density and light DNA density. As for third generation and other generations, there was an increase in ^{14}N ratio over ^{15}N ratio, this increase represented by increasing in ^{14}N band thickness. After undergoing these experiments, it was obviously known that DNA in both prokaryotes and eukaryotes replicates semi-conservatively. It is really deserve to be noted that some historians philosophers in genetics regard this experiment (outlined in figure 3.4) is the most performed elegant scientific experiments.

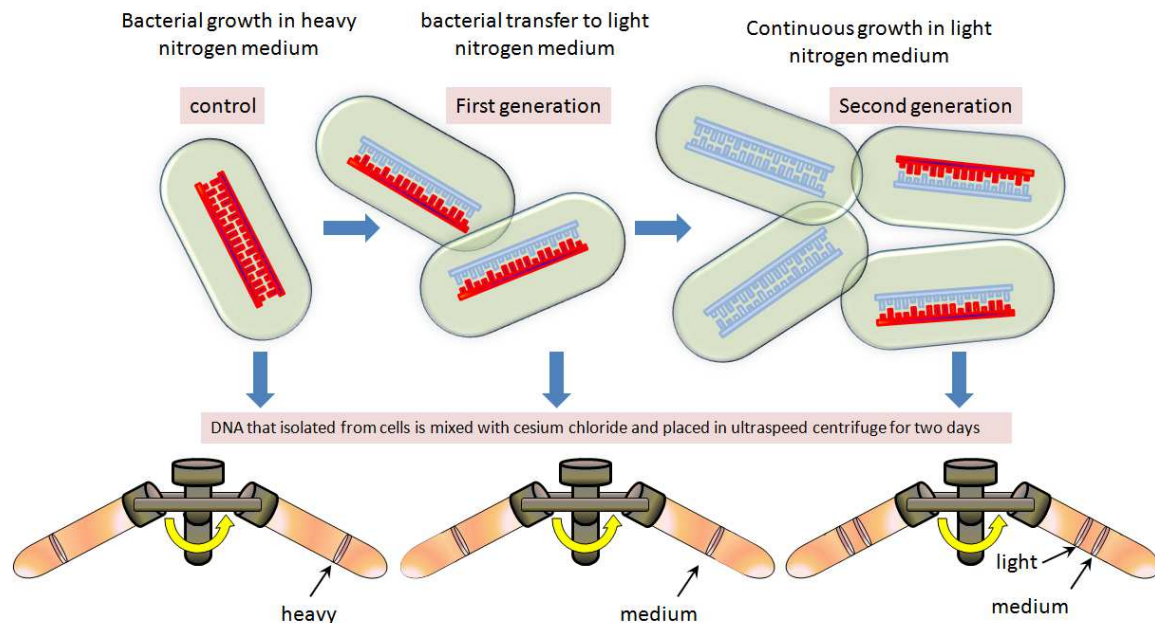


Figure (3.4). Outline for Meselson and Stahl experiment at 1958, which clarifies the practical steps of the experiment through which the right hypothesis of DNA replication was made known.

As it discussed previously, DNA templating is the principle of replication other than transcription, in which each strand is complementary to its corresponding partner (A with T, and G with C), since complementarity between DNA strands entails recognition of each nucleotide in the DNA template strand by a free (unpolymerized) complementary nucleotide to bind with it by hydrogen bonds (figure 3.5).

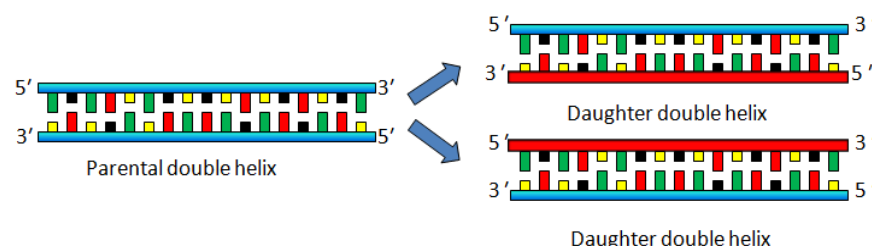


Figure (3.5). The DNA double helix acts as a template for its own duplication. Because the nucleotide A will successfully pair only with T, and G only with C, each strand of DNA can serve as a template to specify the sequence of nucleotides in its complementary strand by DNA base pairing. In this way, a double-helical DNA molecule can be copied precisely. Blue color refers to parental strands while red color refers to daughter strands.

Thus, polymerization continues from 5' to 3' with catalysis from DNA polymerase, which was discovered in 1957. The free nucleotides that serve as substrates for this enzyme were found to be deoxyribonucleoside triphosphate. The reason behind their existence as triphosphate form is to give necessary energy to get to their complementary nucleotides that located in the template strand. Each nucleotide incorporated in DNA or RNA chain in a monophosphate form with exclusion of terminal nucleotide at 5' end which is being on triphosphate form. Hence, they contain in addition to α phosphate group, on β and γ groups (figure 3.6).

Requirements of DNA polymerase

DNA polymerase require in its pure state into:

1. DNA template
2. Four deoxynucleoside triphosphate
3. Magnesium ion to synthesize DNA
4. Primer provides exposed 3' OH end, since DNA polymerase is unable to initiate replication by itself. When an exposed 3'OH end exists, this polymerase enzyme is capable to proceed in the reaction from 5' to 3' direction depending on the template strand.

This enzyme catalyzes the addition of mononucleotides into 3' OH end for the growing DNA chain. In the same time, the linkage between α and β phosphate atoms are broken, to release inorganic pyrophosphate group (PPi).

The formation of biphosphate linkage may requires to take place as nucleophilic attack, by 3' OH group, on α phosphate atom of the triphosphate nucleoside in the chain. This attack causes displacement of pyrophosphate group and formation of intra-nucleotide linkage.

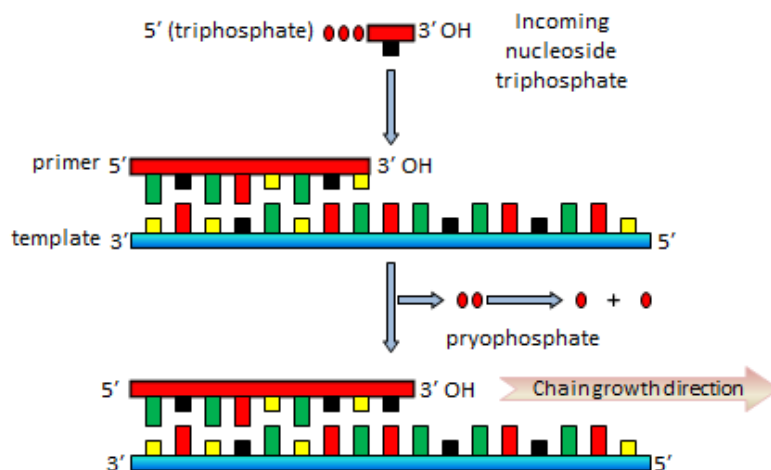


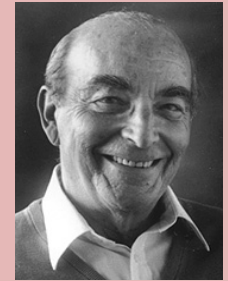
Figure (3.6). DNA synthesis catalyzed by DNA polymerase. As indicated, DNA polymerase catalyzes the stepwise addition of a deoxyribonucleoside to the 3' OH end of a polynucleotide chain, the primer strand, which is paired to a second template strand. The newly synthesized DNA strand therefore polymerizes in the 5'-to-3' direction. Because each incoming deoxyribonucleoside triphosphate must pair with the template strand to be recognized by the DNA polymerase, this strand

determines which of the four possible deoxyribonucleoside (A, C, G, or T) will be added. The reaction is driven by a large, favorable free-energy change, caused by the release of pyrophosphate and its subsequent hydrolysis to two molecules of inorganic phosphate.

The Story of DNA polymerase I and DNA polymerase III

At 1957, Arthur Kornberg and his colleges isolated from *E. coli* cells an enzyme capable to synthesize DNA on polynucleotide template. This enzyme called DNA polymerase for his ability to polymerize new DNA strand using template strand, and they called it Kornberg enzyme according to his discoverer.

It was thought at that time that this enzyme is responsible on replication of DNA in bacterial cells. But, unfortunately, when this enzyme studied with much more details they revealed to results non-concomitant with what it was revealed by Kornberg. The most famous experiments was these undertaken at 1969 in *E. coli* cells, in which the gene encoding for the enzyme discovered by Kornberg, which now known by DNA polymerase I was deleted. Although this gene was deleted but *E. coli* cells remained capable to replicate their own DNA. Ultimately, researchers revealed to the following: although DNA polymerase involved in replication, but it is not the main replicating enzyme. As for conclusion, the enzyme that thought it has a main role in replication is DNA polymerase III. The complexity of the role played in the cell is reflected by the complexity of its structure which consists of ten subunits (see below). Thus, DNA polymerase exists in *E. coli* cells in three types, DNA polymerase III which is the main replicating enzyme, DNA polymerase II which is the enzyme that is involved in what is called SOS response in which it facilitate DNA synthesis in the damaged template strands (see chapter seven), and DNA polymerase I which consists of only one subunit and involved in DNA replication by removing RNA primers at the 5' end of each Okazaki fragment and replacing it with DNA. Add to that, it has main role in DNA repair. In addition to DNA polymerase I, our discussion will be focused on DNA polymerase III which involved in DNA chain elongation in replication fork.



Arthur Kornberg

How DNA polymerase I Work in *E. coli*

DNA polymerase I consists of one subunit that can be cleaved by proteolytic treatment. The biggest cleavage product is called Klenow fragment which can be used in the synthesis reactions *in vitro*. Klenow fragment contains two activities; 5' to 3' polymerizing activity and 3' to 5' proofreading activity, the later activity is concerned with removing of mistaken nucleotides. The second cleavage product of this enzyme contains 5' to 3' exonuclease activity, and it deserves to note there is a space between the location of addition of nucleotides and location of removing nucleotides.

The three activities of DNA polymerase I are acting with each other in order to provide a unique feature for the enzyme, which collect three different activities in only one subunit, this in turn gives an ability to this enzyme to participate in many activities *in vivo* and *in vitro* in addition to replication such as DNA repair (see chapter seven) and recombination (see chapter eight). The ability of this enzyme to extend a small nick in the DNA as a result of several activities this enzyme enjoy gives a great role for it to be involved in many genetic engineering experiments (see chapter eleven). From this we can

conclude that Arthur Kornberg wasn't imagining haphazardly that this enzyme was the main replicating enzyme in the cell, but he might not realize at that time that there was a much more complex enzyme than this one which is DNA polymerase III.

How the Main Replicating Enzyme Work in *E. coli*

We will focus on DNA polymerase III since its the main replicating enzyme in the cell. This enzyme is a highly complex protein with a high molecular weight (600 KDa), and it consists of 10 different subunits. The core polymerase contains three subunits; α subunit (which represents fingers) consists of the active site for nucleotides addition since it has the 5' to 3' polymerizing activity, ϵ subunit (which represents palm) is concerned with removing of mismatched nucleotides (mistakenly added or misincorporated nucleotides) by a mechanism known as proofreading mechanism. This subunit is playing this role since it has nucleotide removing enzymatic activity which known as exonuclease activity that has a reverse direction to polymerization which is from 3' to 5', and θ subunit (which represents thumb) has however unknown function (figure 3.7).

DNA polymerase III is proceeding – as it was discussed above – from 5' to 3' since it possesses α subunit which has 5' to 3' polymerizing activity. It cannot proceed in a reverse direction because it doesn't possess another subunit with 3' to 5' polymerizing activity. But, sometimes it happens that a mismatched nucleotide is incorporated in the growing DNA chain, in this case, DNA polymerase III stop working, then it transfers the 3' end of the growing chain of the polymerizing α subunit (fingers), into the digesting ϵ subunit that has a reverse direction. Thus, the misincorporated nucleotide is removed by 3' to 5' direction. Once the misincorporated nucleotide is removed by the proofreading mechanism in a reverse direction, the 3' end of the growing DNA chain is flipped back into the polymerizing site of α subunit. Then, polymerizing process is resumed from 5' to 3' direction (figure 3.7).

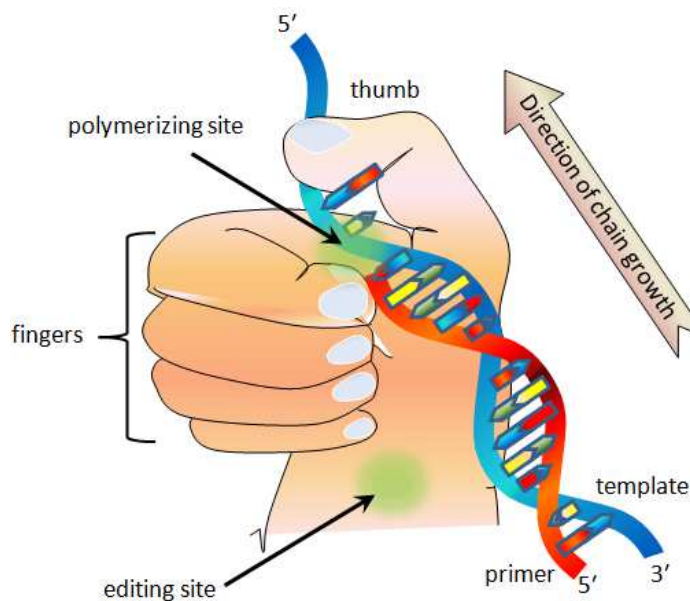


Figure (3.7). Core DNA polymerase structure in *E. coli* according to X-rays crystallography data. Simply, DNA polymerase resembles the half opened right hand, in which each of palm, fingers, and thumb grasp on DNA. Moreover, two separated locations are protruded; polymerizing site and proofreading site of the enzyme.

The main role of the residual subunits (6 subunits) is to convert the core polymerase from distributive (non-processive); the enzyme that falls down from the template strand after its synthesis of 10 to 50 nucleotides, into a processive enzyme; the enzyme that can synthesize long stretches of DNA reaching to 500 thousands nucleotides without being fallen down from the template strand. The later activity is very necessary for efficient synthesis for both leading with lagging strands. The principle of the processive nature of DNA polymerase III is the ability of β subunit on forming a shape similar to the circular cake or to the ring around double helix in order to connect and load the core enzyme (figure 3.8). Hence, one of the main answers of the question for the reason behind the intervening of a complex enzyme such as DNA polymerase III in replication despite of DNA polymerase I has the same polymerizing activity that found in DNA polymerase III is relied in the inability of DNA polymerase I to extend the DNA chain for long distances as it was found in DNA polymerase III. This is one of the main reasons in the complexity of DNA polymerase III compared with DNA polymerase I, that's why DNA polymerase III requires several subunits to do this task. DNA polymerase core molecules alone are able to synthesize only short stretch of nucleotides before they fallen down form template strand. But, this is not happened for this enzyme since the existence of a structure known as clamp, which keeps the polymerase enzyme connected firmly with DNA strand by its binding to DNA strand from one side and to the core enzyme from the another side. The presence of sliding clamp creates two conflicting features for DNA polymerase (figure 3.8). Although it binds tightly with DNA over a long stretches, it is smooth enough to move slightly from one nucleotide to another one. The reason behind these conflicting features is attributed to the conformation of β subunit with resemble a circular cake around DNA or a ring on a string (figure 3.8).

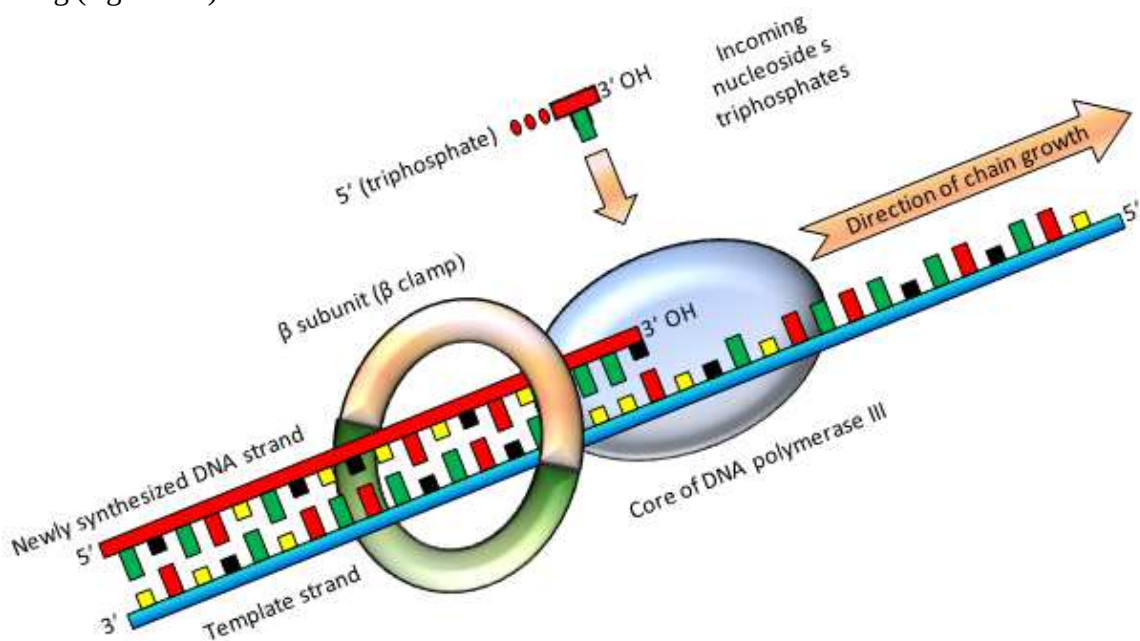


Figure (3.8). β subunit dimer restrains *E. coli* core DNA polymerase III with DNA, by this way, the processivity of this enzyme is increased more than one thousand folds.

How for such clamp has the ability of preventing polymerase from dissociation and without at the same time halting the rapid movement of polymerase along DNA molecule? Researches that taken place on this clamp revealed that it's a large ring around DNA double helix. It was noticed that only one side of the ring is connected to the backward portion of DNA polymerase, while the entire ring is sliding freely along DNA as long as polymerase moves. The clamp is never connected with itself DNA polymerase. Instead, a clamp loader or what is called γ complex, which consists of five subunits, and with the assistance of ATP hydrolysis connects the clamp with DNA polymerase (figure 3.9). Thus, the processivity of this enzyme is increased.

The γ complex intermediates two important functions; (1) loading of β subunit clamp into core polymerase at the initiation of DNA strand synthesis, in a reaction needs ATP. (2) Unloading of β subunit clamp after completing the synthesis of DNA strand (figure 3.9).

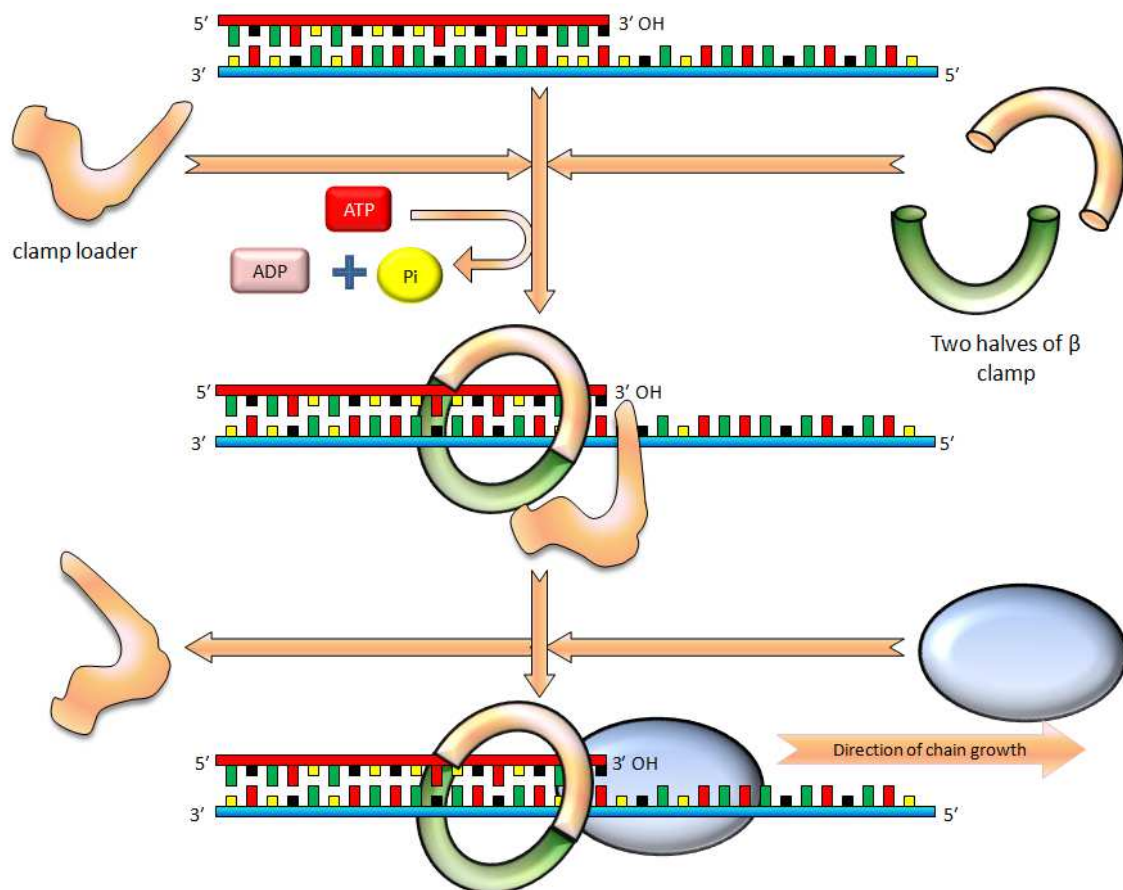


Figure (3.9). Assembly of clamp in order to keep DNA polymerase moving on DNA in the reaction simplified above. Once the clamp is assembled on polymerase, the clamp loader dissociated.

Problems Facing DNA polymerase III in Replication

One of the most beneficial ways to get into the details of replication is to resemble the main replicating enzyme DNA polymerase III with the architectural engineer who asked to replicate DNA, then this engineer look at DNA to see what he will do. Off course, DNA replication is not easy, so, many problems will appear against this brainy engineer if he conquers them he can achieve this task. These problems might be summarized into five main problems:

The first problem; it can be called unwinding problem, in which DNA polymerase is unable to melt DNA double helix (*i. e.* breaking hydrogen bonds) in order to separate strands to be copied.

The second problem; it is the initiation problem, and its represented by inability of DNA polymerases usually to synthesize DNA or RNA chains *de novo* since it requires to a previously existed DNA or RNA strand known as primer.

The third problem; it's called directionality problem and its represented by inability of all DNA polymerases to catalyze the addition of nucleotides in both directions, *i. e.* from 5' to 3' and from 3' to 5'. As long as the two strands of DNA double helix are antiparallel (from 5' to 3' and from 3' to 5') in their chemical direction, and as long as all DNA polymerases catalyze the addition of nucleotides at 3' OH end of the growing DNA chain and not at 5' end, therefore DNA strands can grow only from 5' to 3' and they are unable to grow in the reverse direction.

The fourth problem; it correlates with DNA topography (that's why I call it the topological problem). This problem is represented by the supercoiling of double stranded helix during the proceeding of replication fork forward. This problem can be understood through resembling DNA with a rope of two braids, fixed by a nail form one of its margins on the wall and the another margin held by someone to twist in order to proceed with it toward the wall. Hence, this problem appears in which when this person arrives to a stage at which he can't proceed much more toward the wall because of "supercoiling" against the twisting point (against replication fork). This impedes the proceeding and stops the process entirely.

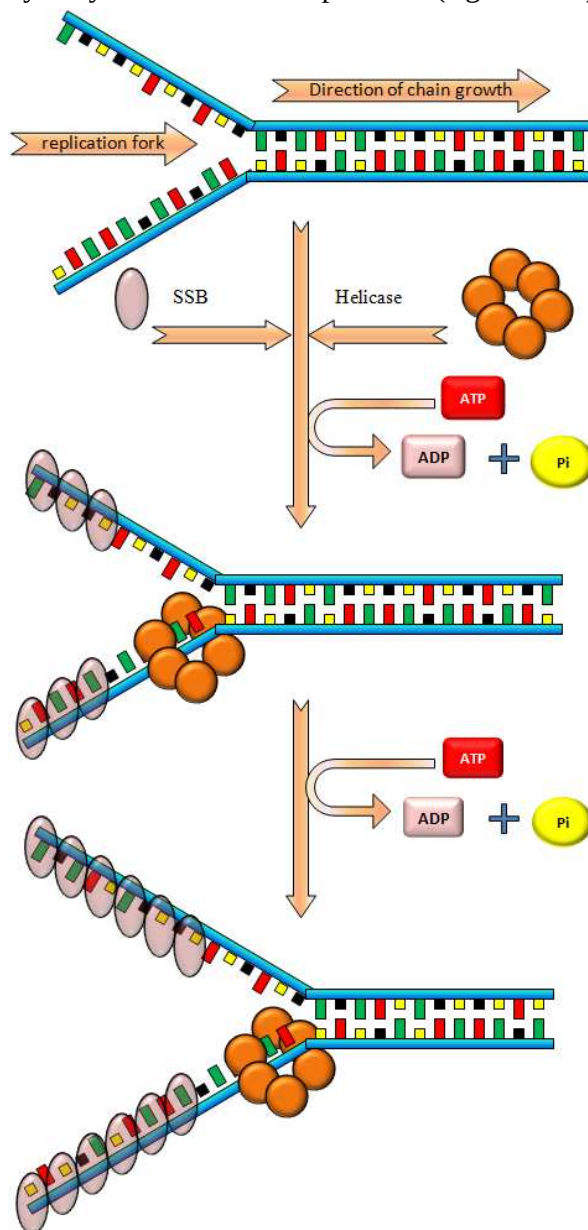
The fifth problem; it's called end replication problem. It's specialized with eukaryotes since they contain linear chromosomes instead of circular counterparts that exist in prokaryotes such as bacteria. In this chapter, we will describe how cells solve the above mentioned problems that resulting from DNA structure and from the characters of DNA polymerases, such as unwinding problem, priming problem, directionality problem, topological problem, and end replication problem.

Solving Unwinding Problem

To proceed DNA synthesis, DNA double helix must be opened upstream of replication fork to enable deoxyribonucleoside triphosphates to base pair with the template strand. However, DNA double helix is very stable under normal conditions, at these conditions base pairs bind with each other with a high strength, so, water temperature must reach boiling degree to separate them in test tube (see chapter one – DNA denaturation).

Therefore, DNA polymerase III can copy DNA duplex only when template strand exposed after its separation from its complementary strand. Hence, other replicative proteins should exist to assist opening of duplex to provide single stranded DNA template to be copied from DNA polymerase III. There are two proteins participating in this process; DNA helicase and single stranded DNA binding proteins.

1. DNA helicase; helicase is an enzyme consists of six identical subunits, and it has the ability to wound around both single DNA strands in an ATP required reaction. Helicase represents a class of enzymes that move along DNA duplex utilizing ATP hydrolysis for strands separation (figure 3.10). Helicase binds to a single-stranded segment of DNA, and then it moves along that strand melting the hydrogen bonds that link it to its complementary strand. DNA helicase uses ATP hydrolysis which enable it to change its conformation by a certain way allow it to do its function in separating DNA strands, by this mechanism, it pulls itself forward along the single stranded DNA. When it faces the double helix region, its continue moving along the strand, separating the double helix strands from each other in 1000 nucleotides per second (this is the same velocity of replication in *E. coli*). Helicase, as long as many proteins that act on DNA, is processive (see the meaning of processivity in DNA polymerase III). Because it forms a clamp around a single strand of DNA, helicase does not "fall off" until it reaches the end of that strand or it is "unloaded" from DNA by another protein.



segment of DNA, and then it moves along that strand melting the hydrogen bonds that link it to its complementary strand. DNA helicase uses ATP hydrolysis which enable it to change its conformation by a certain way allow it to do its function in separating DNA strands, by this mechanism, it pulls itself forward along the single stranded DNA. When it faces the double helix region, its continue moving along the strand, separating the double helix strands from each other in 1000 nucleotides per second (this is the same velocity of replication in *E. coli*). Helicase, as long as many proteins that act on DNA, is processive (see the meaning of processivity in DNA polymerase III). Because it forms a clamp around a single strand of DNA, helicase does not "fall off" until it reaches the end of that strand or it is "unloaded" from DNA by another protein.

Figure (3.10). The mode of action of DNA helicase and SSBs.

2. Single stranded binding proteins (SSBs); which called duplex stability removing protein. SSBs bind firmly and

cooperatively (the binding of each one facilitate the binding of other) with the exposed nitrogen bases single stranded DNA strands, that's why bases remain ready to act as templates. SSBs are unable to open the long double helix directly (figure 3.10), but they assist helicase in stabilizing the unwinding state by preventing rewinding state, by this way, they present template strand to DNA polymerase. Moreover, their cooperative binding cover and strengthen the single stranded DNA regions of lagging template strand. Thus, they prohibit the formation of short DNA molecules that are similar to hairpin structures which hamper DNA synthesis stimulated by DNA polymerase. The binding of SSBs with single stranded DNA acts on protecting it from endogenous nuclease attack.

Solving Initiation Problem

As it is mentioned above, DNA polymerase can't elongate DNA strands without previously existing primers by other enzymes. The primers used in DNA replication in prokaryotes and eukaryotes are short RNA molecules their synthesis are stimulated by primase which is a type of RNA polymerase. RNA primase doesn't act its role alone by depending on itself, but it does so by its binding with helicase that bound at that time in the DNA region to be replicated. Usually, the term primosome is used to indicate the complex that formed between primase and helicase. After its binding, RNA primase synthesizes short RNA primers complementary to both strands of DNA double helix. Once this enzyme finishes its work, it dissociates from the single stranded template strand.

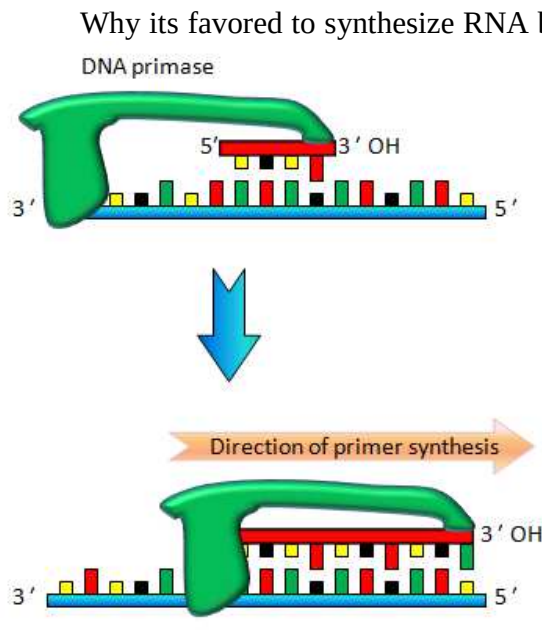


Figure (3.11). Synthesis of RNA primer. Illustration for the reaction catalyzed by DNA primase. After synthesis of short polynucleotide, primase stops to provide exposed 3' OH end to be elongated by DNA polymerase afterward.

Solving Directionality Problem

Before we get to solve this problem, we have to understand its details first then we have to think on how solve it later. The details that taken place through intracellular DNA replication might be understood by DNA polymerase, in which each old (parental) DNA strand acts as template to create complete new strand. As long as each of daughter cells inherits new DNA helix consisting of two strands one of them is old and the other one is new. Therefore, the replication of DNA double helix is semi-conservative. How such unique heroically mission is achieved? The essence of action is kept in template strand itself, which is the strand that is copied to form new strand. The information of template strand is kept despite of that the first copy has complementary sequence, and not has homologous sequence. Hence, copy of the copy produces the original template sequence again. Once copying ends, each of the resulting double helices consists of one of original strands plus its copy separated from each other.

Researches carried out in 1960s have revealed that there was short replication regions were persistently moving along the double helix of parental DNA. Since it resembles Y letter, this active region is called replication fork (figure 3.12). This region is a highly specialized area of DNA, on which DNA polymerase begins to add nucleotides. These replicative forks start to form at *OriC*, (the point of replication initiation) then; they continue growing and enlarge in shape as they continue growing. The presence of these structures is not restricted to the genome of chromosome but they are present in any molecule having replication origins sequences (therefore each molecule having these sequences is called replication entities or replicons).

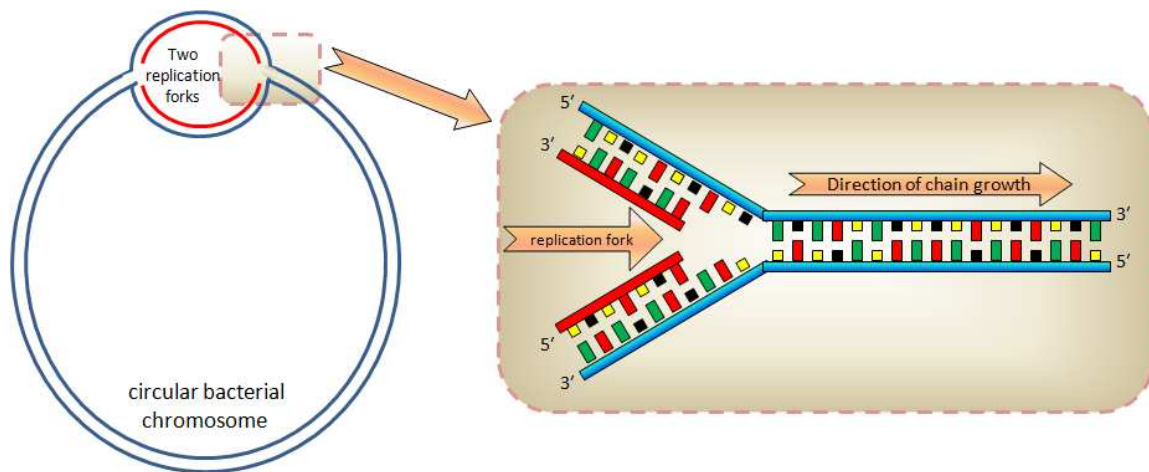


Figure (3.12). Two of replication origins moving in a reverse direction in a circular chromosome.

Initially, the simplest mechanism for DNA replication is the continuous growing of both new strands, nucleotide after nucleotide, in which replication fork moves from one end of DNA to the other end. Because of antiparallel nature of the two DNA strands, this mechanism will require one daughter strand to polymerize from 5' to 3' and the other strand

to polymerize in the reverse direction. Such replication fork will require two types of DNA polymerase. One of them polymerizes from 5' to 3', the other one polymerizes in the reverse direction (from 3' to 5') (figure 3.13 A). But this type of polymerization is not happen in replication since there is no DNA polymerase has 3' to 5' polymerizing activity. So, how DNA chain growth is achieved from 3' to 5'?

This question was answered through several experiments done by Reiji Okazaki at the end of 1960s, he noticed the presence of short segments of 1000 – 2000 nucleotides length, these segments known now as Okazaki fragments, their length is estimated in eukaryotes (such as human) to be 100 – 200 nucleotides. Totally, Okazaki fragments are polymerized from 3' to 5' then they connected with each other after their synthesis to form a long chain of DNA. The daughter DNA strand that is synthesized continuously from single RNA primer is called leading strand, its direction is from 5' to 3', this direction is the same for replication fork movement. The synthesis of this strand is slightly precede the synthesis of the other daughter strand which is synthesized discontinuously, which is known as lagging strand, which its synthesis is much more complex. As for lagging strand, the direction of nucleotides polymerization is somehow against the direction of replication fork movement. DNA synthesis in lagging strand is delayed because it has to wait leading strand to expose the template strand in order to allow Okazaki fragments to be synthesized on it. The nature of lagging strand discontinuous synthesis is the only reasonable mechanism which is compatible with the 5' to 3' polymerizing activity for DNA polymerase. This mechanism is considered complex compared with leading strand (figure 3.13 B).

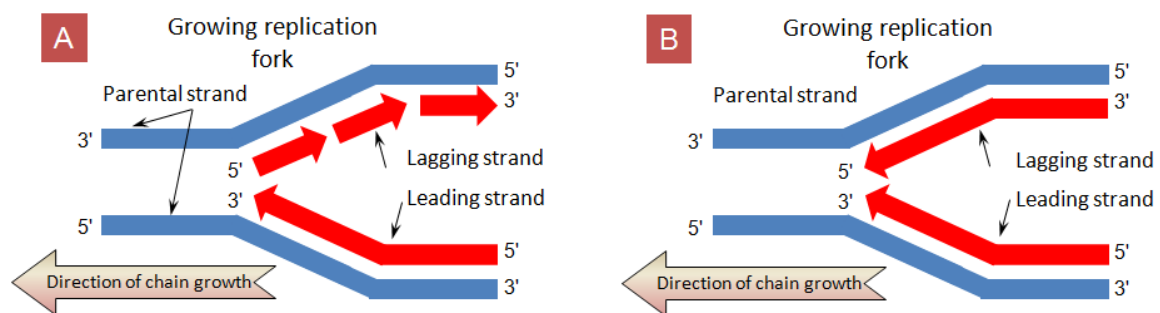
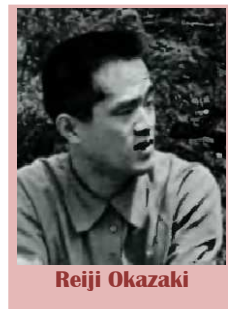
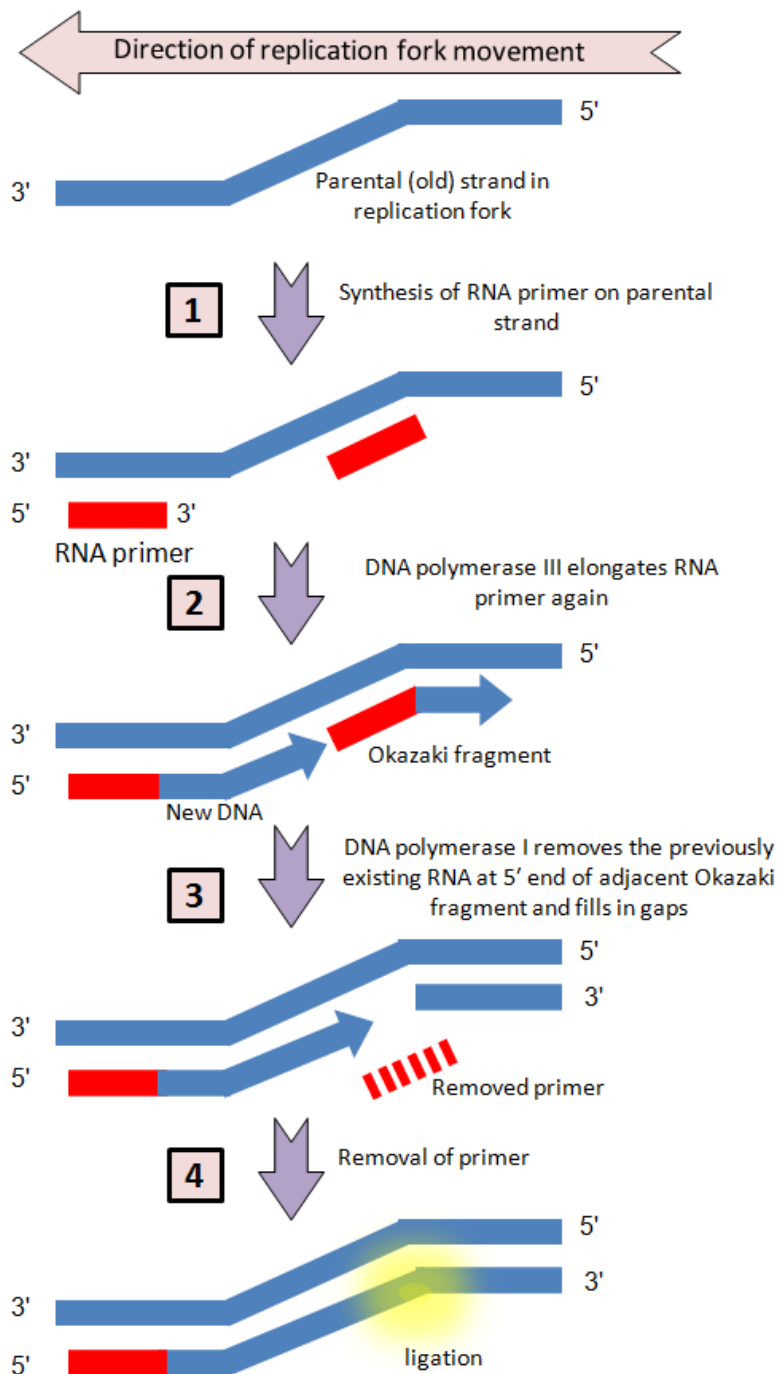


Figure (3.13). (A) The right model for DNA replication, which involves discontinuous synthesis of Okazaki fragments in the lagging strand. (B) Incorrect model for DNA replication. Although this model seems to be the simplest reasonable model for DNA replication, but the mechanism shown here is not the mechanism that used by cells in their replication. In this figure, both daughter strands will grow continuously. This require to the growth of daughter DNA chain in both directions (from 3' to 5' and from 5' to 3'). But in fact, there is no enzyme possesses 3' to 5' polymerizing activity to stimulate reaction in this direction.



Therefore, it's obvious – after Okazaki fragments discovery – to us how this mechanism solves this problem. Regarding leading strand, only one primer is only necessary at the beginning of replication, once replication fork forms, DNA polymerase III is continuously adds new nucleotides to the end of the growing DNA chain. Regarding lagging strand of replication fork, each time DNA polymerase finish Okazaki fragments synthesis (which cost several seconds), it must synthesize completely new fragment in a distant location along the template strand (figure 3.14). This mechanism is used specially to produce primers necessary to this purpose.

Figure (3.14). Okazaki fragments synthesis on lagging strand which consists of four steps: step (1) primer synthesis, step (2) primer elongation, step (3) RNA removing and filling gaps, step (4) ligation of the newly formed Okazaki fragments with each others.

In the first step: RNA primase (which uses ribonucleoside triphosphates instead of deoxyribonucleoside triphosphates) synthesizes short RNA primers on lagging strand (figure 3.14).

In the second step: since RNA primers contain 3' OH end, therefore, it can be elongated by DNA polymerase III in *E. coli* to initiate addition of deoxyribonucleoside triphosphates (dNTPs) to the 3' end of primers by the same enzyme. Thus, each lagging strand grows in a reverse direction to the movement of replication fork. These growing short fragments that containing RNA that covalently bound with DNA are known as Okazaki fragments according to their discoverer Reiji Okazaki. In bacteria and bacteriophages, the length of these fragments are from 1000 to 2000 nucleotides, and the cycle of Okazaki fragment synthesis costs two seconds to complete. In eukaryotes, Okazaki fragments are much shorter (from 100 to 200 nucleotides). The synthesis of each Okazaki fragment ends once DNA polymerase III reaches to the 5' end of the previous Okazaki fragment.

In the third step: a special rapid system is performed to repair DNA to produce a continuous chain of DNA fragments on lagging strand through removing of RNA primers and replacing them with DNA sequences. This is done in *E. coli* by DNA polymerase I.

In the fourth step: an enzyme called DNA ligase links 3' end of DNA with 5' end of the previous fragment to complete the process (figure 3.14).

Solving Topological Problem

Since DNA molecule consists of two strands wound around each other, some processes that correlates with replication and replication termination of DNA and transcription of DNA faces topological problems. Certain enzymes in the cell cause overcoiling (positive supercoiling) or undercoiling (negative supercoiling).

Positive supercoiling is taking place when the circular double helix wrapping around itself with the same direction of duplex wrapping (right handed) as in the figure (3.15), while negative supercoiling is taking place when the duplex wrapping around itself with a reverse direction of the duplex wrapping (left handed).

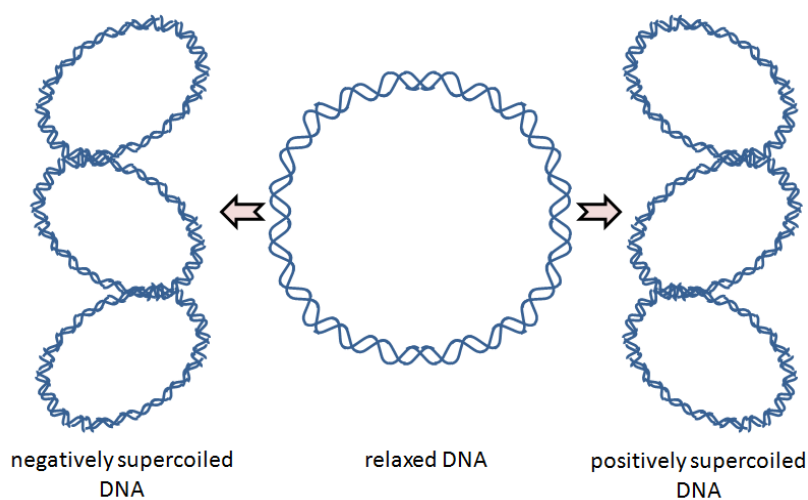
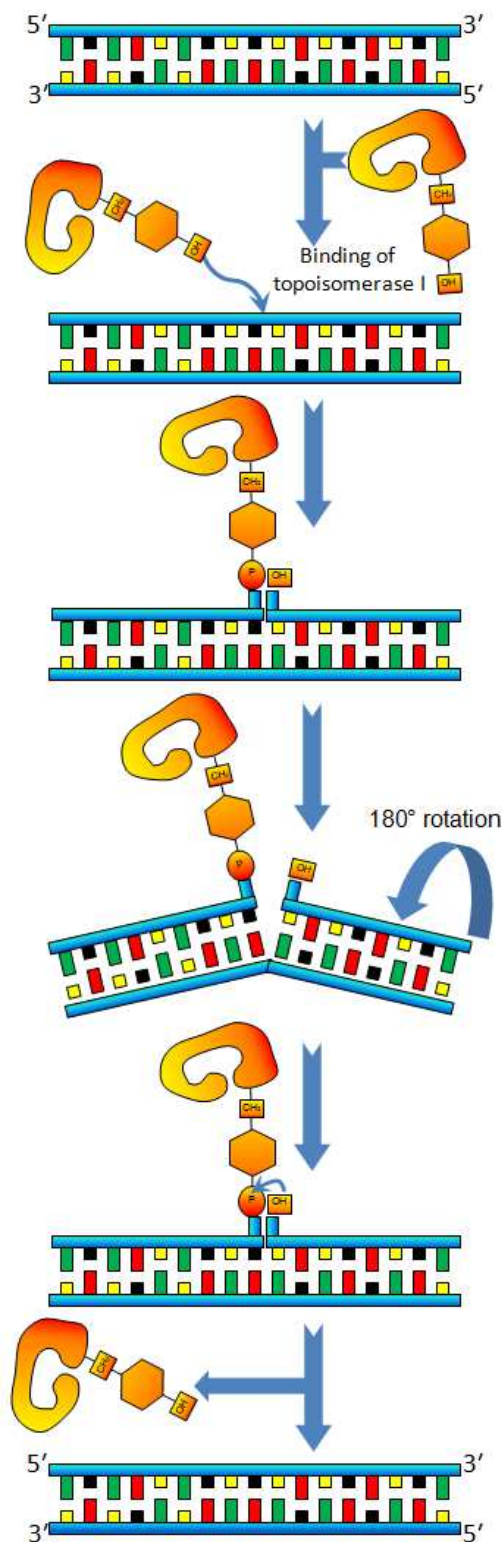


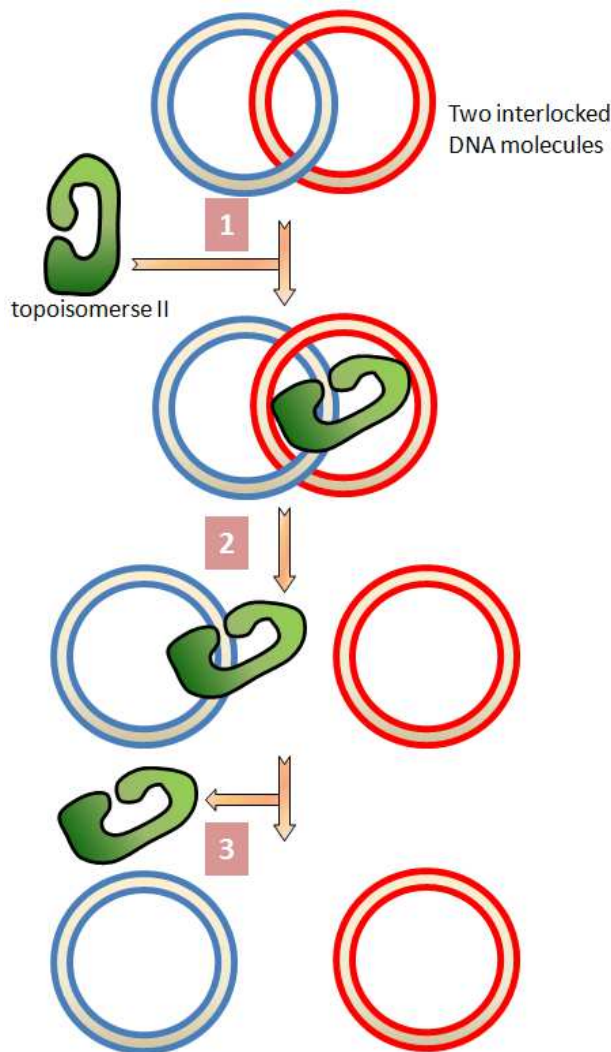
Figure (3.15). Negative supercoiling and positive supercoiling of a circular DNA molecule. Topoisomerases can act on the relaxed DNA (in the center) and gives it negative supercoils (to the left) or positive supercoils (to the right).



Negative supercoiling increases the number of coils of each duplex around the other (this is called linking number or L), whereas positive supercoiling decreases it. All the three forms noticed in figure (3.14) having the same sequences, but they differ only in linking number. Accordingly, these structures are called topological isomers or topoisomers. These enzymes that synthesize or reduce these states are called topoisomerases. These enzymes affect on supercoiling with one of two ways, either topoisomerase type I, which breaks one of the duplex strands, while its still connected with the broken ends, then it seals the nick (figure 3.16), or topoisomerase type II (such as DNA gyrase in *E. coli*) that have the same mode of action, but instead of breaking one strand it breaks both of them and passes the other duplex through the temporal gap. And as long as DNA replication proceeds, a positive supercoiling is formed before replicons. Such supercoiling is discarded by topoisomerases which either create negative supercoiling upstream of replicons to prepare for replication or to alleviate of positive supercoils after its creation.

Figure (3.16). Role of topoisomerase I in reducing DNA supercoiling by breaking one strand of duplex and passing the other strands through it. It is impossible for each end of duplex to rotate on itself without enzymatic intervention that shown in steps 1 to 5. In step 1; active site tyrosine loaded topoisomerase type I recognizes phosphate group in the duplex first. Then, (in step 2) it covalently links with this group. By this way, it breaks the phosphodiester bond in only one strand. Hence (step 3) it is possible for the ends of the duplex to rotate with respect to each other releasing accumulated tension as a result of supercoiling. Since phosphodiester bond energy is kept by the enzyme by phosphate linkage with tyrosine, therefore, it is possible for the reaction to reverse by the attack of OH group on phosphate group to pulls it to its original location again (step 4). Eventually (step 5), phosphodiester bond is formed again to repeat the same enzyme without changing and the same duplex with a reduction of its supercoiling of one degree.

In addition to the importance of topoisomerases in the above mentioned replication, they are generally important in other aspects, this importance is clarified through the experiments done on the genes that encode for these enzymes. It was appeared that mutating of these genes result in a very poor growth for the bacteria that exposed to such mutation. Topoisomerases remove tangled DNA as in chromosome condensation (see chapter one). Moreover, topoisomerase type II plays an important role in the final stage of



circular molecules replication. When the circular DNA replication is complete, two interlocked circles are formed, hence, DNA topoisomerases act on separating of these two interlocked molecules from each other (figure 3.17). so, as noticed above, this enzyme act in a way similar to topoisomerase type I but it causes a temporal breakage in both duplex strands. Thus, topoisomerase type II binds with one of the circles that generated from replication of the duplex and it causes a double stranded temporal breakage which works like a "portal" through which DNA circle can pass (figure 3.17). Then, topoisomerase type II reseals the broken strands. This enzymatic feature was exploited in DNA ligation experiments in genetic engineering (see chapter ten).

Figure (3.17). Role of topoisomerase in terminating replication of DNA circular molecules. In step (1) topoisomerase type II binds with the newly separated interlocked DNA molecules. In step (2) the same enzyme causes a breakage in both DNA strands for one circular DNA molecule to passes the other molecule through the breakage. In step (3) the enzyme reseals the broken strands for the same broken circular molecule.

Solving End Replication Problem

Circular and linear DNA molecules face different positions at the end of DNA synthesis. Circular molecules have DNA regions called regions of termination or (*ter*) in *E. coli* chromosome, they are located 180° away from origin of replication in *E. coli* circular chromosome (*OriC*). There are six additional sequences localized at both sides of meeting

point, they act as terminators when they bind with termination protein. These regions order the replisome to stop replicating DNA.

In case of circular molecules, each replication fork leaves replication origin and moves to the opposite directions away from each other and meet in the middle of the way around the circle. If one of the forks is late regarding to the other one, it will stop when it reaches its termination point and waits the arrival of its partner. Termination point inhibits helicase that used during DNA synthesis, therefore, the proceeding of two replication forks is halt, and when the two forks meet, this is the end of replication is actually happened.

In case of linear molecules, as it's found in eukaryotes such as human, animals, plants, yeasts, and fungi, all of them have linear chromosomes; replication alone is unable to replicate the ends of chromosomes entirely. There is a special way correlates with the behavior of telomere is used to ensure not to loss information from the ends of linear chromosomes.

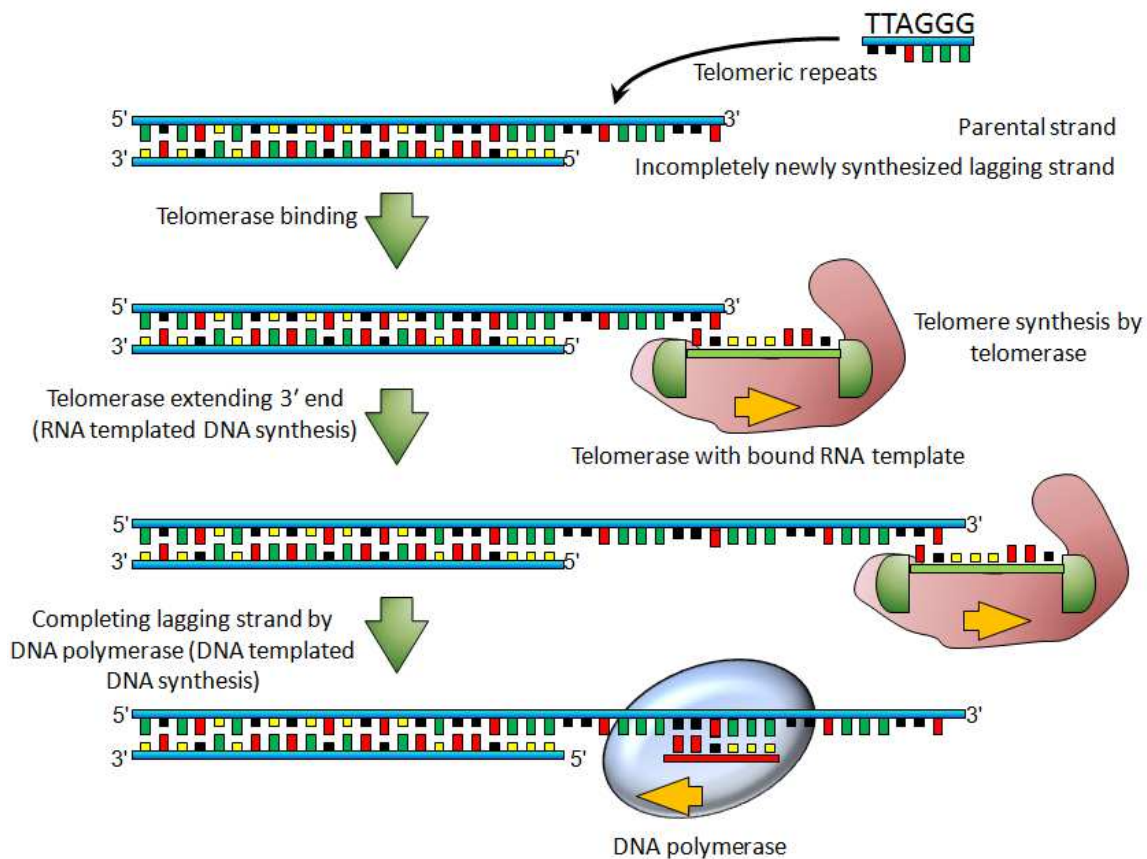


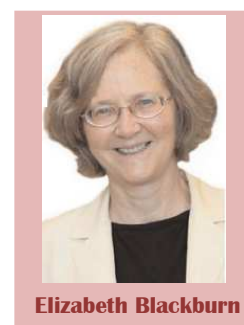
Figure (3.18). Telomere replication. This figure summarizes reactions involved in the synthesis of G-C rich sequences which constitute the ends of chromosomes (telomeres) for many eukaryotes. The incomplete, newly synthesized strand is the strand made on the lagging side of a replication fork. As indicated above, the telomerase is a ribonucleoprotein complex that carries an RNA template for synthesizing a repeating, G-rich telomere DNA sequence which is TTAGGG in humans. The lagging strand is presumed to be completed by DNA polymerase α , which carries primase as one of its subunits.

DNA replication is semi-discontinuous; this shows the difference of DNA replication mechanism that taken place in leading strand compared with that taken place in lagging counterpart. The leading strand replicates continuously, for the lagging strand to replicate, DNA polymerization begins to form many RNA primers which are then elongated to synthesize Okazaki fragments. Finally, these primers are broken and replaced with DNA sequences. The displacement of any RNA primer in leading strand leaves a gap which is usually filled in by the extension of the next Okazaki fragment. In mammals, there is a problem in replicating the extreme ends of linear chromosomes. At the beginning of 1970s, James Watson was discovered that the features of DNA replication prevent the cell from copying the linear DNA termini (telomeres) entirely. Because of the nature of lagging strand synthesis, DNA polymerase is unable to completely replicate 3' end of the duplex linear DNA strand. When Watson deduced this problem at 1972, he showed that when DNA polymerase arrives to the end of linear DNA molecule there would be a problem in completing replication (figure 3.18).

In case when telomere absent, semi-conservative DNA replication will cause in production of linear molecule which became shorter and shorter with each replication. Watson showed this shortening indirectly when he noticed unmasking of part of telomere from the activity of DNA polymerase. Thus, ends are blocked from replication with each successive cellular division. In other meaning, telomere length is reduced with each cycle of replication. For many years, it was noticed that gradual losing of telomeres may act as a triggering essential mechanism for the beginning of senescence. But the activation of telomerase (ribonucleoprotein complex, in which RNA component plays an important role in stimulating the addition of telomeric sequences in each cycle of replication, while the role of proteins is confined with assisting of RNA in this process) in malignant cells almost act on compensating this defect by adding of tandemly repeated sequences (TTAGGG) after each cycle of replication. Consequently, its possible for malignant cells to keep their survival by this way (figure 3.18).

There are many experiments paved the way to the discovery of telomerase which are began with the notion that losing of genomic sequences in each cycle of replication was can be compensated by addition of terminal sequences. Moreover, living organisms have the ability to transfer species specific terminal sequences to the DNA.

At 1985, both Carol Greider and Elizabeth Blackburn were discovered that there was more than on activity in *Tetrahymena* extract. They found an activity responsible for addition of telomeric repeats for only telomeric single stranded DNA oligonucleotides. They found too that this process was inhibited by treating the extract with RNase. Therefore, this DNA dependent activity is called terminal transferase of telomerase. This process is taken place by copying template strand sequences which is part of RNA component in this enzyme. They found later that this enzyme is composed of RNA component and protein component. This enzyme was studied extensively by using *Tetrahymna thermonphila* since a single cell in this organism has more than 40000 telomeres. Generally, telomerase mode of action can be summarized with the following steps; primer recognition, nucleotides addition, translocation from one place



Elizabeth Blackburn

to another (figure 3.18) and by this way the enzyme extends G rich 3' overhang that exist at the end of telomeres.

Summary of Replication Steps

Replication of DNA begins at nucleotide sequences called origin of replication. Helicase unwinds DNA duplex, SSBs bind cooperatively with the single stranded DNA regions to keep the unwinding state. DNA polymerase III is considered as the main replicating enzyme in *E. coli*. This enzyme is capable of adding nucleotides for 3' end of the previously existing chain of nucleotide, but it cannot initiate chain synthesis *de novo*. Therefore, RNA primase, a type of RNA polymerase, synthesizes RNA primer, which is a sequence composed of 10 bas pairs complementary for parental DNA. After primer synthesis, DNA polymerase III adds deoxyribonucleoside triphosphates (dNTPs) to synthesize new chain of DNA. Since the two DNA strands are antiparallel, their elongation must be done by two different mechanisms, in which the leading strand elongates toward replication fork by the addition of nucleotides continuously to the growing 3' end. In contrast, lagging strand that elongates away (in reverse direction) from replication fork. It is synthesized discontinuously by a chain of fragments called Okazaki fragments (figure 3.19).

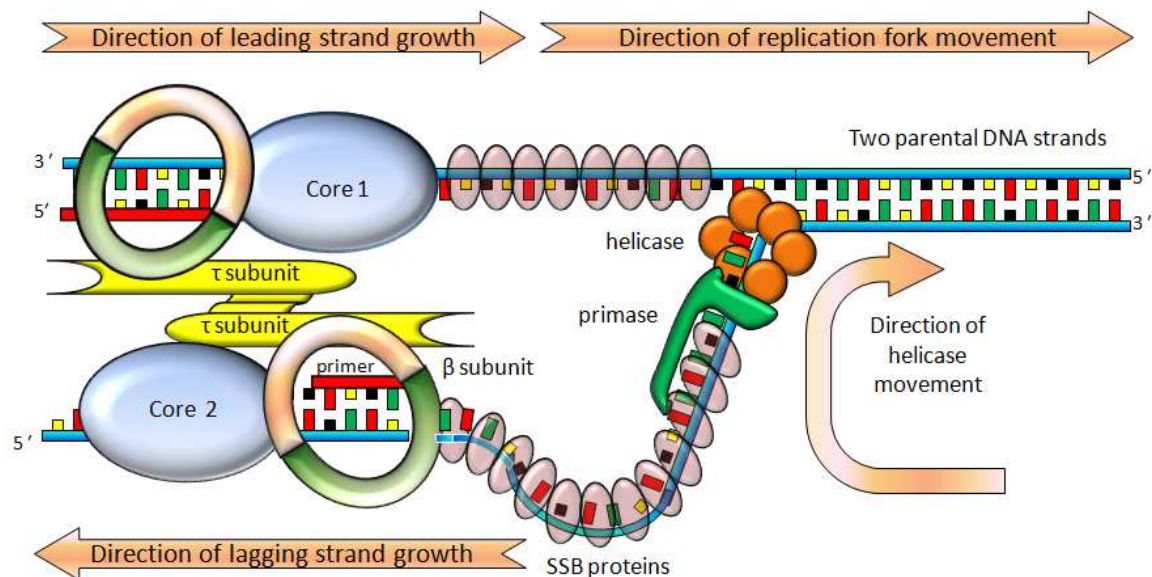
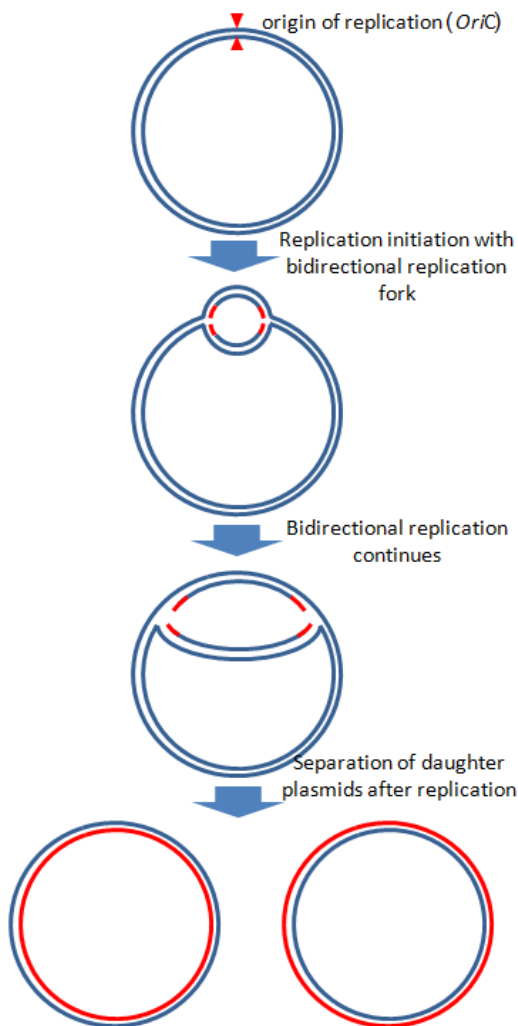


Figure (3.19). The relationship between the different replication proteins with each other at replication fork.

When DNA polymerase III arrival to RNA primer of lagging strand, it will be replaced with DNA polymerase I, the later removes RNA and replaces it with DNA. DNA ligase then binds and seals the residual nicks. Afterword, DNA is unwounded much more and new primers are added to the lagging strand, whereas DNA polymerase III jumps ahead to commence the synthesis of other Okazaki fragment. For simplicity, DNA polymerase III is composed of two separate molecules, one of them acts on leading strand, and the other

one acts on lagging strand, and they both acting with each other simultaneously. The reason behind this belongs to τ subunit which acts on dimerizing the two core molecules of polymerase with each other, which makes the synthesis of both leading and lagging strand simultaneous. Thus, replication proteins are connected with each other in a single large unit, and moves quickly along the DNA to give it the ability to be synthesized in both directions of replication fork in a compatible and efficient mode. The DNA replication is a process involves a mixture of proteins each of which was discussed alone, but in real, most of proteins are carried with each other by a large multi-enzyme complex known as replisome which moves quickly along DNA duplex. This complex can be imagined as a small sewing machine that composed of proteinaceous components that get energy from hydrolysis of nucleoside triphosphates. However, although replication complex has studied extensively in *E. coli* and many viruses, but this structure resembles in its function the structure exists in eukaryotes. The functions of units of this structure are shown in figure (3.19).



Other Replication Structures

The above discussed *E. coli* replication is represented by theta (θ) structure. This typical model is taken place in circular and linear DNA molecules altogether. And it's followed in large molecules such as bacterial genome and some plasmids as well (figure 3.20). Replication proceeds from one origin in one direction (unidirectional) or in both directions (bidirectional) until the entire plasmid is copied.

There are other models for replication circular chromosomes, and they are represented by other two styles for replicating circular chromosomes which are rolling circle and D – loop.

Figure (3.20). The autonomous plasmid replication from its bacterial chromosome. Replication begins at origin of replication (*oriC*) and continues around the circle bidirectionally. And in some plasmids, replication is taken place in a unidirectional way.

Rolling Circle Model

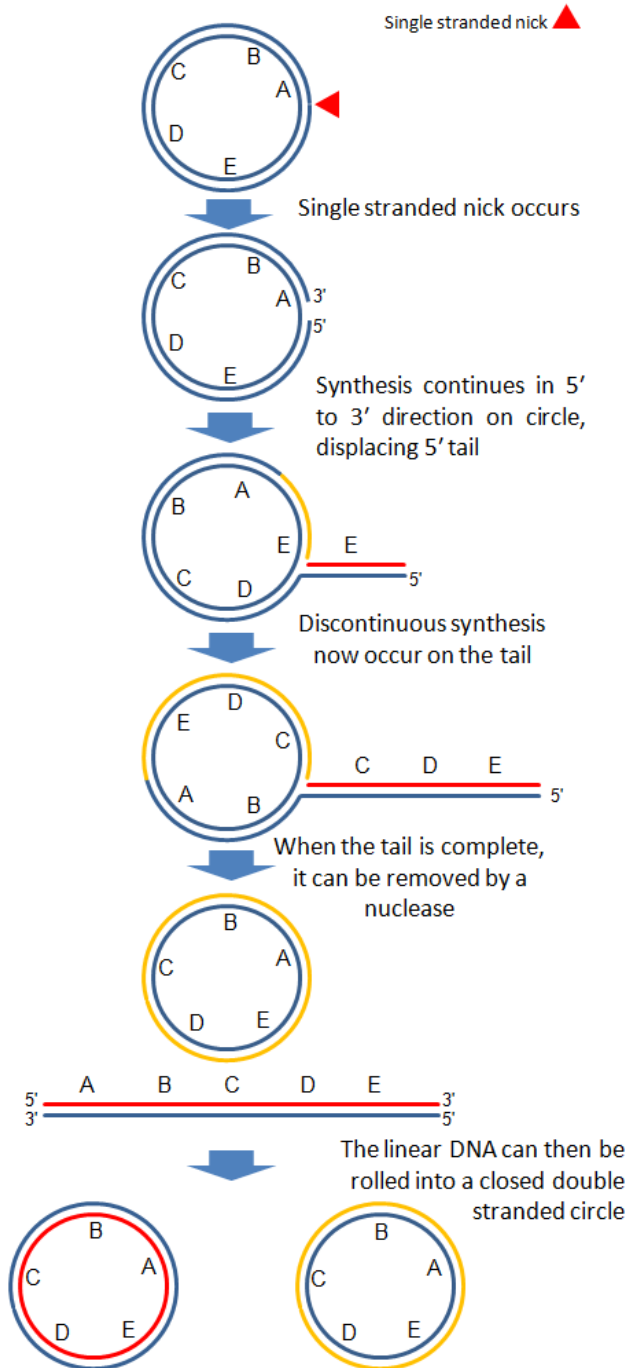
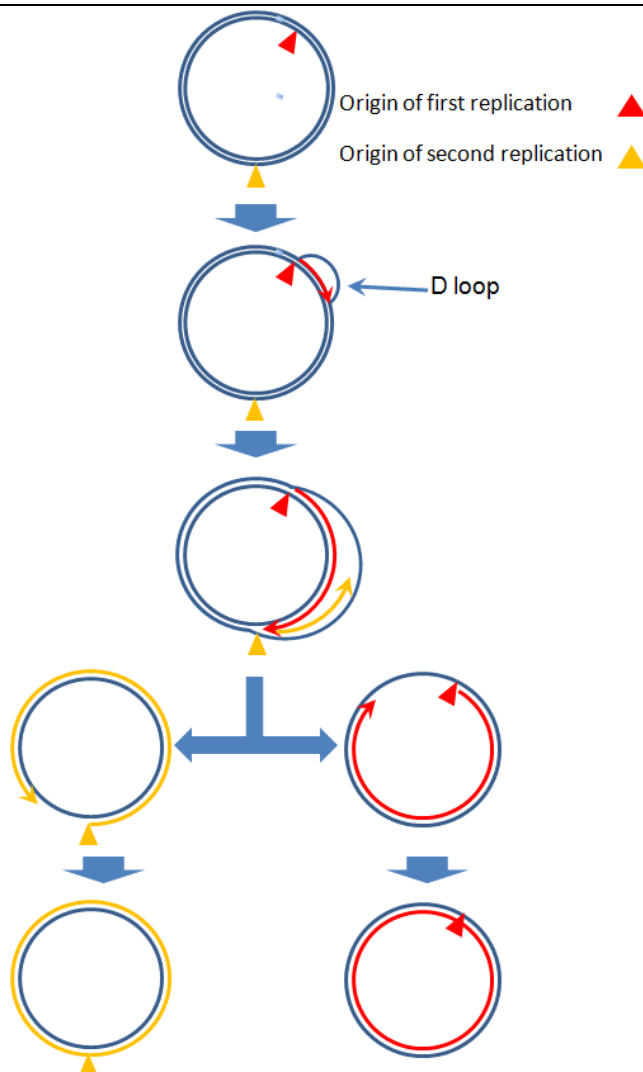


Figure (3.21). Rolling circle replication. Letters A to E are considered as chromosomal markers to show the direction of movement.

This type of replication is taken place in viral DNA, in which many phages (viruses that infect bacteria) use this way to fill in their heads with linear DNA that is replicated from the circular parental molecules. Add to that, this mode of replication is taken place in *E. coli* too during the mating as it exemplified in F plasmid or in Hfr chromosomes (see chapter nine). According to rolling circle replication, a nick (a breakage in one of the phosphodiester bonds) in one of the circular DNA molecule is made. And it carries two ends; one of them is 3' OH end, and the other is 5' phosphate end (figure 3.21).

Under the effect of helicase and SSBs, replication fork is generated. Hence, primer strands synthesis is not necessary because of the availability of 3' OH end. Consequently, leading strand synthesis continues by elongating this exposed OH end. In the same time, the parental template strand is displaced to synthesize lagging strand. The type of polymerase used in this process is DNA polymerase III. The parental strand is replicated by the usual way by primer. Such replicative model produces circle with linear extension, this is resemble the Greek letter sigma (σ) hence the name sigma replication or rolling circle replication comes.



D – Loop Model

Chloroplast and mitochondria (in eukaryotes) that contain circular DNA molecules are replicated by a slightly different mechanism. In these molecules, the origin of replication occupies a different point for each parental DNA strands, displacing the other strand in the same time at which the displacement loop (D – loop) is formed (figure 3.22). Replication proceeds until it arrive the origin of replication of the other strand. Then, replication begins on the other strand in a reverse direction. But the first model (theta structure) in mitochondrial DNA can be taken place under certain growth conditions.

Figure (3.22). D loop model in replicating mitochondria and chloroplast. D – Loops are formed from replication of mitochondria and chloroplast at different regions of duplex.

Chapter Three Appendix

Replication in Eukaryotes

Although the general features of DNA replication in eukaryotic cells are similar with those seen in prokaryotes, there are some important differences seen between them. Eukaryotic chromosomes are very large, in some cases they are larger in thousands folds than their prokaryotic counterparts. To replicate these huge linear masses of DNA in a reasonable time, they must contain several regions for replication initiation which known as origins of replication. Their number is usually more than 100 *OriC*, while in the circular *E. coli* genome there is nothing except only one origin of replication (figure 3.23). The velocity of DNA polymerases in mammals (eukaryotes) is less from ten to hundred folds than the velocity of DNA polymerases in bacteria (prokaryotes). The velocity of polymerization in mammals is about 1000 nucleotides per second. The presence of nucleosomes which hamper the movement of enzymes is the reason behind this slow. It is known that polymerases "negotiate" nucleosomes somehow during polymerization, but when DNA polymerase meets nucleosomes it necessarily requires dislocation between histones and DNA in nucleosomes structure. In all cases, this is the reason behind this low velocity.

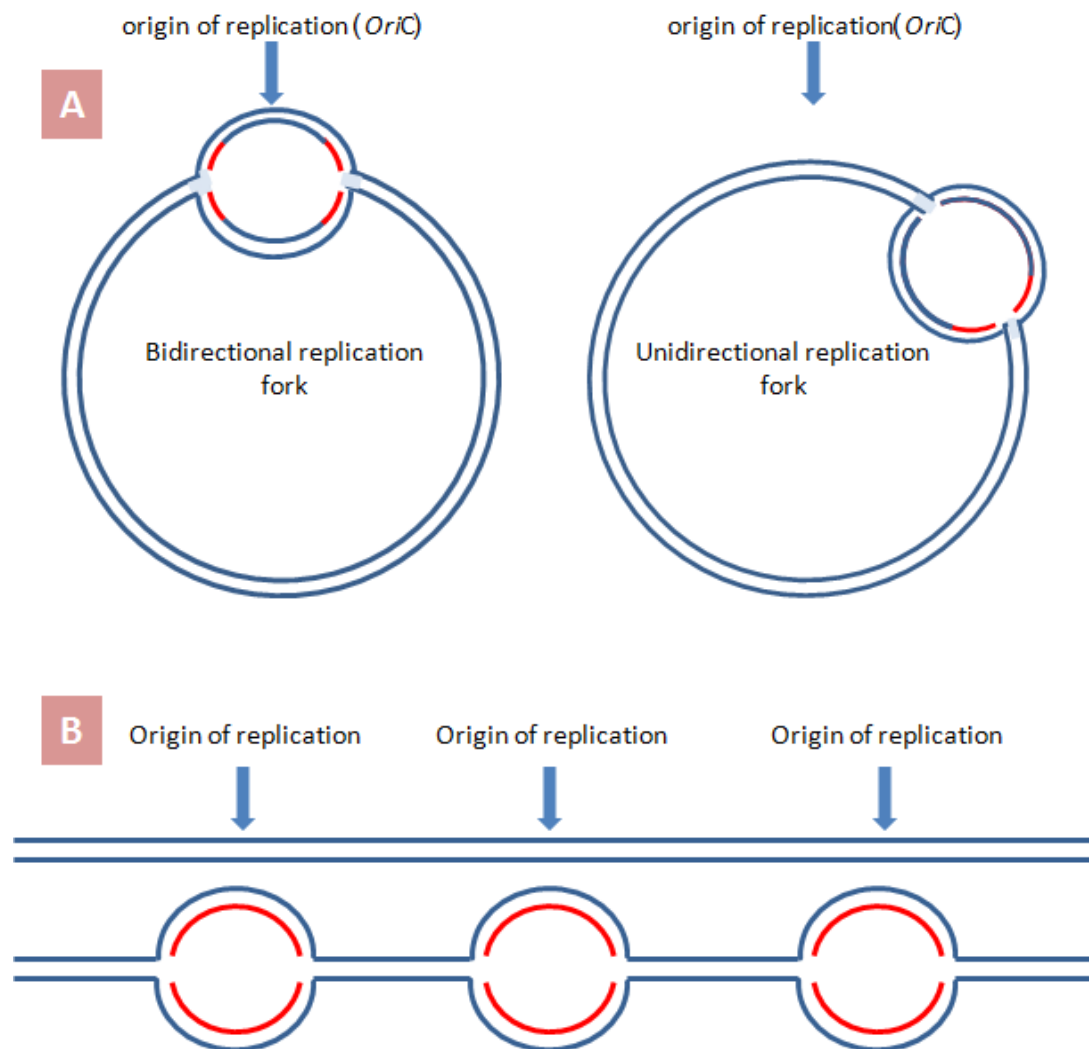


Figure (3.23). Origins and replication forks. In branch A; replication forks in a small circular DNA in prokaryotes. In branch B; part of very long linear DNA in eukaryotes.

It is deserve to note that the length of Okazaki fragments is also differ between prokaryotes and eukaryotes, since these fragments – as mentioned above – have length from 100 to 200 nucleotides in mammals and 1000 to 2000 nucleotides in bacteria. As for DNA polymerases that participating in replication in mammals, they are fallen in four main types which are; α , β , γ , and δ . DNA polymerase (δ) is the main replicating enzyme in mammals. This enzyme exists in nucleus and behaves in the same way DNA polymerase III behaves in *E. coli*. DNA polymerase δ consists of two molecules; one of them synthesizes the leading strand while the other synthesizes the lagging strand. DNA polymerase α exists in nucleus, and participates in replication but it is not the main replicating enzyme (it resembles DNA polymerase I). DNA polymerases β and ϵ exist in

nucleus as well but they don't have a role in replication, and they may be involved in DNA repair. DNA polymerase γ exists in mitochondria and it's responsible on mitochondrial genome replication (mitochondrial genome replication is circular, thus, it also mimics the circular *E. coli* genome).

Chapter Three Revision Questions

Q1- Justify the following:

- Free nucleotides that act as a substrate for DNA polymerase exist found in a deoxyribonucleoside triphosphate form?
- DNA polymerase usually requires primer?
- DNA polymerases proceed from 5' to 3' and not in the reverse direction?
- Although DNA polymerase III has connected tightly on long stretches of DNA, but its loose enough to slide sequentially from one nucleotide to another?
- DNA polymerases couldn't initiate replication *de novo*?
- RNA is synthesized from RNA primase that don't have proofreading activity, then, its replaced instead of being synthesized by DNA directly?
- DNA ligase is called by this name?
- Replication is described as semi-discontinuous?
- If we put two meters of DNA in a very small space in the cell it wouldn't be tangled, while if we put the same amount of any usual thread it would be tangled?
- Leading strand is synthesized simultaneously with the lagging strand?
- DNA helicase never "full out" until it reaches to the end of DNA strand, or until it is unloaded from DNA by another protein?
- Pyrophosphate group is released in each step of polymerization?
- DNA polymerases velocity in mammals (eukaryotes) is less ten folds than its velocity in bacteria (prokaryotes)?
- Rolling circle replication is called in this name?

Q2- choose the correct answer:

- In each step of replication, when each nucleotide is added to synthesized DNA strand, it has to release
a. nucleotide b. nucleoside c. monophosphate d. pyrophosphate
- Arthur Kornberg at 1957 was isolated..... Enzyme and gave it his name.
a. DNA polymerase I b. DNA polymerase III c. DNA polymerase III d. DNA polymerase VI
- this problem represents by supercoiling during the proceeding of replication fork forward.
a. unwinding problem b. initiation problem c. directionality problem d. topological problem
- The length of Okazaki fragments in human is about Nucleotides.
a. 10 b. 100 c. 1000 d. 10000
- Elizabeth Blackburn used in her experiments to discover telomerase organism.
a. *Tetrahymena thermophila* b. hippopotamus c. *Salmonella typhi* d. Neanderthal
- the predominant model in replication is
a. theta b. rolling circle c. strand displacement d. non
- rolling circle replication is taken place in
a. F plasmids b. R plasmids c. Mega plasmids d. mitochondria
- if one of the parental strands contains origin of replication and the other complementary one contains a different origin of replication, hence, this replication is called
a. theta replication b. rolling circle replication c. strand displacement replication d. conservative replication
- DNA polymerase III resembles in prokaryotes enzyme in eukaryotes.
a. DNA polymerase α b. DNA polymerase β c. DNA polymerase γ d. DNA polymerase δ
- the hypothesis of Is the right hypothesis undoubtedly to replicate DNA according to Meselson and Stahl.
a. conservative mode b. semi-conservative mode c. dispersive mode d. only "a" and "b"

11. Acts a major role in the repair mechanism that called SOS response.
a. DNA polymerase I b. DNA polymerase II c. DNA polymerase III d. primase
12. 3' to 5' proofreading activity is exist in Subunit of DNA polymerase III.
a. α b. β c. γ d. E
13. The velocity of DNA polymerases in mammals (eukaryotes) is less ten folds than their velocity in bacteria (prokaryotes) because the existence of which prohibits the proceeding of DNA polymerases.
a- nucleosomes b- nucleotides c- helicases d- replication fork
14. When DNA strand that containing the sequence 3'GACTAACGATGC 5' as a template in replication, the sequence of the resulting parental DNA strand will be
a. 3'CTGATTGCTACG 5' b. 5'GCATCGTTAGTC 3' c. 5'CTGATTGCTACG 3'
d. 3'CGTAGCAATCAG 5' e. 3'CGUAGCAAUCAG 5'
15. It is impossible for to play a role in replication.
a. SSBs b. OriC c. telomerase d. e. nucleosomes
16. commences replication of leading strand
a. DNA polymerase I b. DNA polymerase III c. primase d. RNA polymerase

Q3- define the following:

Dispersive mode of DNA replication, Primosome, Replication fork, Replisome, D structure, OriC, SSBs, DNA polymerase γ

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Arezi B.** and Kuchta. R. 2000. Eukaryotic DNA primase *Trends Biochem. Sci.* 25: 572-576.
- Beese L.** Derbyshire V. and Steitz T. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA *Science* 260: 352-355.
- Berk A.**, Zipursky S., Baltimore D., Darnell J., and Lodish H. Molecular Biology. Fourth edition. Paul Matuataira, USA, 1998.
- Blackburn, E. H.** (1999). The Telomere and Telomerase, How they are Interact? *Mount Sinal Journal of Medicine* 66: 292- 400.
- Bolsover S.**, Hyams J., Shephard E., White H., Wiedemann C. Cell biology. Second edition, Jhon Wiley and Sons, 2004.
- Brandy, C.** Elizabeth Blackburn and the story of telomeres – deciphering the ends of DNA. The MIT Press Cambridge, Massachusetts London, England, 2007.
- Brautigam C.** and Steitz T. 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes *Curr. Opin. Struct. Biol.* 8: 54-63.
- Cooper J.** and Hausman R. The cell, A Molecular Approach. Fourth edition, Sinecure Associates, 2007.
- Feng, J.**, Funk, W. D., Wang, S. L., Weinrich, A. A., Avilion, C. P., Chiu, R. R., Adams, E., Chang, R. C., Allsopp, J. and Yu, H. (1995).The RNA Component of Human Telomerase. *Science* 269: 1246-1241.
- Fitzgerald-Hayes M.** and Reichmann F. DNA and Biotechnology. Third edition, Elsevier/ 2010.
- Greider, C. W.**, and Blackburn, E. H. (1987). The Telomere Terminal Transferase of Tetrahymena Is a Ribonucleoprotein Enzyme With Two Kinds of Primer Specificity. *Cell* 51: 887- 98.
- Greider, C. W.** (1998). Telomeres and Senescence, The History, The Experiment, The Future *Current Biology* 8: 178- 181.
- Hames B.**, and Hooper N. Instant notes in biochemistry. Third edition, BIOS Scientific Publishers Limited, 2005.
- Hübscher U.**, Nasheuer H., and Syväoja J. 2000. Eukaryotic DNA polymerases: A growing family *Trends Biochem. Sci.* 25: 143-147.
- Kornberg, A.**, and Baker, T. A., *DNA Replication* (2d ed.).W. H. Freeman and Company, 1992.
- Lewin B.** Genes. VII, Oxford University Press, 2000.
- Lewis R.** Human Genetics: concepts and applications. Ninth edition, McGraw-Hill, 2009.
- Murray R.**, Bender D., Botham K., Kennelly P., Rodwell V., Weil P. Harper's Illustrated Biochemistry. 28th edition, McGrawHill Medical, 2009.
- Steitz T.** 1999. DNA polymerases: Structural diversity and common mechanisms *J. Biol. Chem.* 274: 17395-17398.
- Tamarin .**Principles of genetics. Seventh edition. The McGraw–Hill. 2001.
- Verma P.**, and Agarwal V. molecular biology. First Edition Schandgoup, India 2004.
- Watson J.** A Passion for DNA; genes, genomes and society. Cold Spring Harbor Laboratory Press, 2001.

Chapter Four

Gene Transcription



“Monk Transcribing Manuscript,” ca. 1470. Jean Mielot (author of *Miracles de Notre Dame*) at his desk using a quill and scraping knife (Bibliothèque National de Paris/Mary Evans Picture Library/London)

Introduction

The process of RNA synthesis from DNA is called transcription, it is understood better in prokaryotes. Although the regulation of RNA synthesis in mammalian cells is different from prokaryotes, RNA synthesis in prokaryotes is applicable in eukaryotes even enzymes and transcription regulatory signals are different.

Ribonucleotide sequences in RNA molecule are complementary to only one strand in DNA duplex. RNA is analogous in its sequence to one of DNA strands except the presence of uracil instead of thymine, which is called coding strand, while it's complementary to the other strand, which provides template for its synthesis.

The main enzyme responsible about transcription is RNA polymerase. The DNA dependent RNA polymerase or what is known as RNA polymerase is responsible about polymerizing ribonucleotides into a sequence complementary to the coding strand of the gene (figure 4.1). This enzyme binds with specific regions known as promoter, which is localized on the coding strand. This is followed by initiation of RNA synthesis at transcription start point. This process continues until arriving to the termination sequence. Transcription unit is the region located between promoter and terminator. The resulting RNA which is synthesized from 5' to 3' is called the primary transcript. In prokaryotes, the primary transcript represents several genes, while in mammalian cells; it usually represents the product of singular gene.

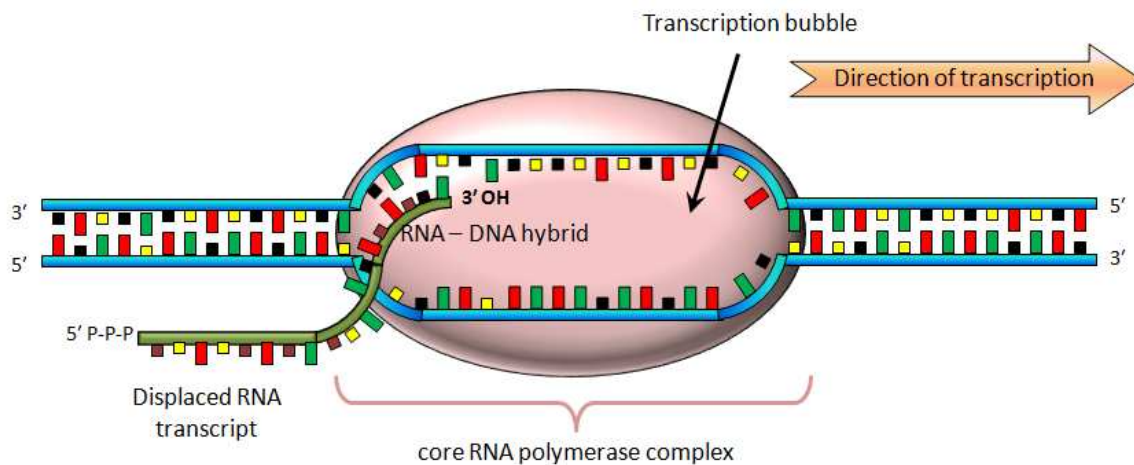


Figure (4.1). Displacing of the newly synthesized RNA chain in DNA duplex. The newly synthesized RNA copy is displaced after its generation behind transcription bubble. Green color of ribonucleoside triphosphates refers to G, yellow to C, red to A, and violet to U respectively.

From DNA to RNA

Transcription and translation are tools by which cells can express their genetic content for its genes. Since RNA is capable to direct synthesis of many analogous protein molecules, it can synthesize large amounts of proteins quickly in case they needed. But each gene can be transcribed and translated with a different efficiency giving the cell the ability to manufacture large amounts of some proteins and small amounts of other proteins. Moreover, as we will see later, the cell can change (or regulate) all of its genes according to their needs – this is often done through controlling its RNA synthesis (see chapter six).

The first step that is done by the cell is to copy a specific region of DNA – gene – into RNA sequence. Although information in DNA are copied into another chemical form (RNA), but they still initially written with the same language as it was before, this is the nucleotide sequence language. Hence, this process is called transcription.

As it was shown in DNA, RNA is a linear polymer consists of four different types of nucleotides connected with each other by phosphodiester bonds. RNA differs from DNA in four main points; (1) the nucleotides of RNA are ribonucleotides, i. e., they have ribose (hence it's called ribonucleic acid) instead of deoxyribose, (2) although RNA, as it's the case for DNA, contains the bases A, G, and C but it contains U instead of T that exists in DNA, (3) although of these small differences, DNA and RNA are excitingly differs from each other in their general structure. While DNA exists in cells as a double helix, RNA exists as a single strand. Thus, RNA chains are folded into several conformations as we will see later. (4) since RNA molecule is single stranded and complementary to only one strand of the two strands of the gene, it is not necessary for G content to equal C content, and A content to equal T content. Hence, two questions emerge, the first one; why RNA contains U instead of T while the reverse for DNA? The second question is: why DNA is deoxygenated at 2' and RNA is none deoxygenated at the same location? To answer the first question you have to review chapter seven. As for the second question reply, it can be

summarized by two reasons; the first one is referred to the utilization of this location as a natural marker provided by the cell for all cellular enzymes that interact with these nucleic acids, this consequently gives different and specialized cellular functions for both nucleic acids. The second reason for the presence of additional oxygen atom in RNA is acting on increasing its reactivity and for the same time decreases its stability compared with DNA molecules (note that if RNA molecule is placed in room temperature for a short period it will suffer from autolysis, whereas this phenomenon is never taken place for the DNA molecule at the same conditions). For this reason, as compared with RNA, DNA is considered as the most suitable molecule to act as reservoir for genetic information for the long term.

All RNA forms are synthesized in the cell by copying DNA, which is a process that has many features in common with DNA replication. Transcription begins by unwinding a small portion of DNA duplex to expose nitrogen bases on each DNA strand. As it was shown in DNA replication, nucleotide sequences in RNA chain is determined by base pairing according to Watson – Crick base pairing rule between incoming nucleotide and DNA template. When incoming nucleotide is suitable for template nucleotide, it will covalently bind with the growing RNA chain by enzymatically catalyzed reaction that loses pyrophosphate during each addition of nucleotide. Thus, the growing chain is entirely complementary for the DNA strand that is used as template.

Transcription, however, differs from replication in many important aspects. In contrary with the newly synthesized DNA strand, RNA strand doesn't stay in hydrogen bonded state with the template strand. But instead of this, and behind the region after which nucleotides are added, RNA chain is removed and replaced with DNA duplex. Thus, the transcription resulting RNA molecules from DNA template are produced in a single stranded form (figure 4.1). Add to that, since RNA molecules are copied from a specific region of DNA, RNA molecules are quite smaller than DNA molecules. DNA molecules in mammalian chromosome can be more than several million of nucleotides, in contrary, most of RNA molecules have no more than several thousand of nucleotides, and many of them are smaller significantly.

Many RNA molecules can be synthesized from the same gene in very short period of time, in which additional RNA molecules are started to be synthesized before the synthesis of the first RNA molecules is complete. When RNA polymerases follow each other in a hard work manner, each one of them moves with very high velocity. Consequently, thousands of copies of single genes are synthesized in very short period of time as we will see later.

Although RNA polymerase initially catalyzes the same chemical reaction that catalyzed by DNA polymerase, in which both enzymes catalyzes polymerization reaction from 5' to 3', but there are some differences between the two enzymes. First, RNA polymerase catalyzes ribonucleotides polymerization but not deoxyribonucleotides. Second, in contrary with DNA polymerase that involved in DNA replication, RNA polymerase can initiate RNA polymerization without primer. Instead of this, RNA polymerase starts *de novo* transcription at specific regions at the beginning of the gene (promoter). Such differences can be existed since transcription is not as precise as replication. In contrary to DNA, genetic information is not stored in the cells during

transcription. Therefore, RNA polymerase commits one mistake for every ten thousands nucleotides that copied to RNA (compared with the mistake committed by DNA polymerase with average of one nucleotide for every ten million nucleotides). Consequently, it's so obvious that mistake taken place in transcription is less important than that taken place in replication.

Promoter

Promoter is a short, identifiable sequence, located upstream of transcription start point, and its necessary for transcription. The promoter sequence are so important in such away the substitution of single base in TATA box (one of the most important promoter sequences in eukaryotes) acts as lethal strong mutation. The promoter region contains the information which tells RNA polymerase when it begins, how many times it synthesizes RNA chain, and how it binds carefully.

1. In Prokaryotes

The typical promoter in prokaryotes consists of the components; the first component is composed of hexamere that located -35 upstream of transcription start point (the negative signal indicates regions located upstream of the transcription start point, the more these sequences approach to this point, the more the number they carried decreases until zero point or transcription start site, thus -10 location is closer to transcription start site than -35 location). The second component is another hexamere located -10 upstream of the transcription start point (figure 4.2).

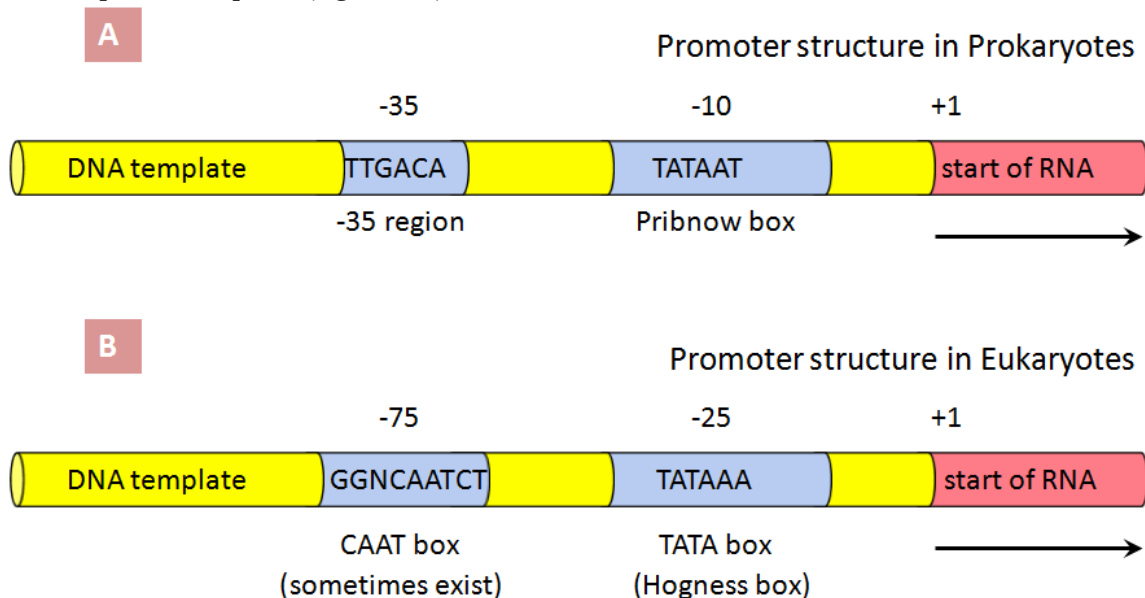


Figure (4.2). Promoter regions that are necessary to initiate transcription in both (A) prokaryotes and (B) eukaryotes. This figure shows the consensus sequences. The first copied nucleotide is numbered +1. The adjacent nucleotide on 5' side is numbered -1. Black line indicates to the synthesis of RNA strand.

The function of -35 sequences is to provide the signal that enables RNA polymerase to recognize promoter sequence, therefore this region called recognition domain. The function of -10 sequence, or what so called Pribnow box according to its discoverer, is became clear through observing DNA unwinding at this region when RNA polymerase arrive to it, therefore this region is called unwinding domain. The -10 region is exclusively consists of A = T base pairs which help on melting DNA duplex into a single strands. The necessary energy to break A = T base pairing is low compared with G $\equiv$ C breaking energy. Thus, A = T pairing requires the minimum energy to break it.

2. In Eukaryotes

Promoter sequences in eukaryotes are larger, more complicated, and more variable compared with these found in prokaryotes. Eukaryotic genes contain promoter regions called TATA box or Hogness box which has the consensus sequence TATAAA that is located at -25 (located 25 base pairs upstream of the transcription start point) (figure 4.2). Promoter regions in eukaryotes also contains CAAT box that has the consensus sequence GGNCAATCT which is located at -75 base pairs.

What increases the complexity of promoter in eukaryotes is the presence of another structure called enhancer. Vertebrates DNA contains enhancers, which activates promoter, the later activity is increased enormously in the presence of enhancers. Enhancers increase the concentration of transcription factors beside promoter. Enhancers localize on both locations; upstream and downstream of transcription start point, they have great affect on transcription that estimated 10 kb (one kb = 1000 bps) away from promoter, thus, they have the ability to operate large distances of DNA.

The difference between enhancer and promoter is not by mode of action but its operational. So, enhancers are capable to catalyze transcription for huge distances, while promoters are capable to catalyze only distances in the immediate vicinity. Enhancers are capable to catalyze gene transcription in both directions, whether it's located upstream or downstream of the gene. This was experimentally proved, in which when enhancers removed and placed in reverse direction to their original direction, it was noticed they are acting with the same efficiency. It was also noted that when promoter don't have these unique features, it don't have the ability to catalyze gene transcription without its being located upstream of it. But, enhancers are described as promiscuous, because in addition to their major function in catalyzing very far distant locations, they can stimulate promoter even in the immediate vicinity. The mode of action of many enhancers is unknown, but some suggest that binding of transcription activator proteins with enhancers causes a looping out between enhancer of one site and promoter from the other site, this is bring both of them to each other to become very close. By this way, transcription activator proteins are able to interact directly with transcription apparatus that is located on promoter.

What increased the complexity of gene regulation in eukaryotes furthermore is their containing on (as in mammals) elements called silencers, which act against enhancers work entirely. These elements resembles with enhancers in terms of their location and direct independent mechanism.

RNA polymerase

1. In prokaryotes

The *E. coli* RNA polymerase catalyzes the synthesis of all types of RNA (including mRNA or messenger RNA, tRNA or transfer RNA, and rRNA or ribosomal RNA), excluding RNA primer that binds at 5' end of Okazaki fragments where its synthesized from a different types of RNA polymerases called RNA primase (see chapter three). There are no different types of RNA polymerases compared with DNA polymerases but there is only one type of this enzyme which is involved in the synthesis of all three types of RNA. RNA polymerase holoenzyme consists of five subunits. Both β and β' subunits are very big subunits. As for the two α subunits, they are identical subunits. Units that constitutes the core enzyme are called permanent subunits that the enzyme needs them in all three transcriptional stages (initiation, elongation, and termination). This differs for σ subunit which the enzyme need it only in initiation (figure 4.3).

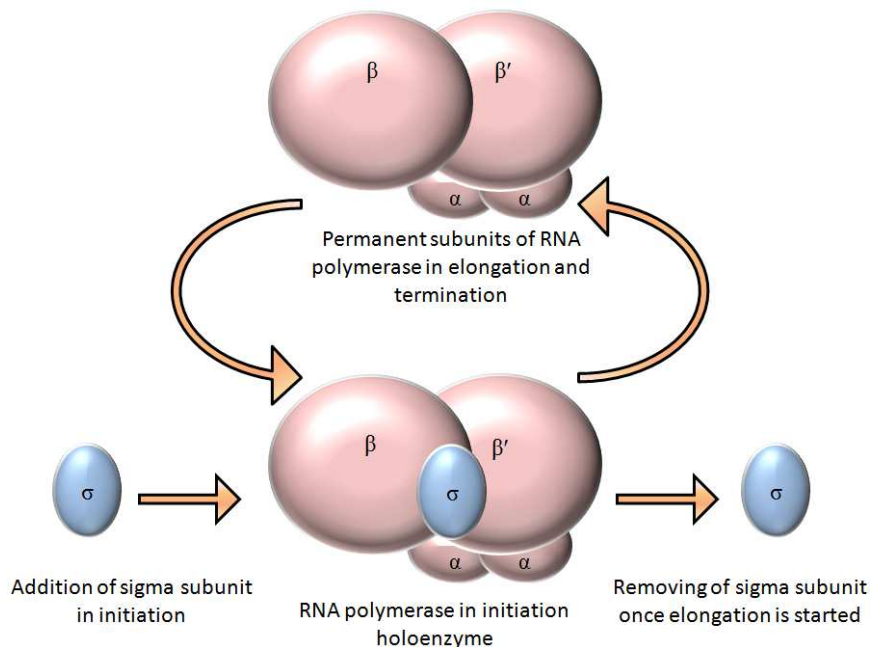


Figure (4.3). The components of RNA polymerase in *E. coli*. RNA polymerase catalyzes nucleotides polymerization into RNA sequence complementary to the template strand of the gene. *E. coli* RNA polymerase consists of core complex which in turn consists of β and β' subunits. The holoenzyme consists of σ subunit that connected with the core related subunits ($\alpha_2\beta\beta'$).

Although the precise role of several core enzyme subunits was not fully identified, but it can say that bacterial RNA polymerase consists of the following subunits:

1. β' (beta prime) subunit: is positively charged subunit, which is a polypeptide thought to be involved in the binding with DNA.
2. β (beta) subunit: is the location of the binding of many transcription inhibitors, and its thought it contains most or all active sites for the formation of phosphodiester bond, it has the ability to form a cross linking with DNA. Each β and β' subunits constitute catalytic center, which undergoes the melting of DNA duplex.
3. Two α (alpha) subunits: they are important to reconstitute the enzyme from separate subunits. Thus, they are necessary to protect the core enzyme. It's suggested that this subunit plays a role in the recognition of promoter, since when T₄ phage infects

E. coli cells; α subunit is modified by phage itself. This modification leads to reduction of the *E. coli* enzyme affinity in promoter recognition. By this way, T₄ phage can mislead *E. coli* (instead of binding with bacterial RNA polymerase to synthesize bacterial genes, *E. coli* promoter binds with T₄ RNA polymerase and synthesizes the viral genes). Moreover, α subunit represents a position on which several necessary transcription factors bound.

4. σ (sigma) subunit: to identify the function of this subunit we must know what is going on when this subunit is absent. σ free core enzyme characterizes in its entire ability to stimulate nucleoside triphosphates polymerization into RNA. Thus, this enzyme doesn't require this subunit in its essential catalytic activity. But the very interesting thing is that despite the ability of core enzyme to bind with DNA, this binding is not specific, *i.e.* this enzyme is unable to recognize the specific sequences at which transcription begins. This was identified when σ subunit is removed from RNA polymerase, it was noticed that this enzyme binds non-specifically with DNA sequences with low affinity. That's why, σ subunit is necessary to identify the right regions for transcription initiation, which are the promoter regions that the core enzyme was unable to recognize. When σ subunit binds to core enzyme, it directs it to the promoter through its specific binding with the promoter sequences -10 and -35. This binding leads to initiation of transcription at the beginning of the gene.

There is more than one class of σ factor, each one of them specified to a different type of promoter. These variations in σ factor appear in some cases. For example during changing the style of life (such as increasing temperature), *E. coli* don't suffer from exiting changes. Instead, they use only alternative σ factors to respond to the general environmental changes. The way of nominating σ factors is either by molecular weight (MW), or by the name of the gene that encodes them). σ^{70} is responsible on transcription most of genes under normal conditions. The alternative σ factors are σ^{32} , σ^{54} , and σ^E , and others, which are activated to respond to different environmental changes. When temperature increased (heat shock) the usual proteins synthesis stops, and in the same time, a new group of proteins are synthesized and produce from genes called heat shock proteins genes, which play important role in cell protection against environmental stress. The *rpo H* gene is a regulator needed to switch on the heat shock response. The product of this gene is σ^{32} that acts as alternative σ factor, which causes transcription of heat shock proteins. The heat shock response is accomplished by increasing the amount of σ^{32} when the temperature increases, and decreasing its activity when the temperature change is reversed. The basic signal that induces production of σ^{32} is the accumulation of unfolded (partially denatured) proteins that results from increase in temperature. The σ^{32} protein is unstable, which is important in allowing its quantity to be increased or decreased rapidly, σ^{70} and σ^{32} can compete for the available core enzyme, so that the set of genes transcribed during heat shock depends on the balance between them.

There is another group of temperature regulating genes, which are regulated by σ^E , which is resulting from the accumulation of partially denatured proteins that existing in periplasmic space and outer membrane. There is little information provided about this factor and about the genes that regulate it. There is another σ factor produced under

nitrogen starvation conditions. *E. coli* cells contain a small amounts of σ^{54} which is activated when ammonia is absent from the culture medium. In these conditions, genes allow to get benefits from the alternative nitrogen sources. There is another state of evolutionary reservation for σ factors which is represented by σ^F that is exist in small amounts and causes RNA polymerase to transcript genes that involved in chemotaxis and flagellar structure.

2. In Eukaryotes

In eukaryotes, the matter is much more complicated, in which in addition to RNA polymerase that exist in mitochondria, three main classes of polymerases are exist as follows;

- RNA polymerase I, which synthesizes most types of rRNA
- RNA polymerase II, which synthesizes mRNA
- RNA polymerase III, which synthesizes tRNA.

The characters of mammalian RNA polymerases are clarified in list (4.1). Each of these enzymes is responsible on copying a different group of genes (see above). These enzymes are much more complicated than prokaryotic RNA polymerases. Each one of these enzymes has large subunits and a number of smaller subunits (14 small subunits for example exist in RNA polymerase II). There is a peptide toxin that has been extracted from the mushroom *Amanita phalloides* (figure 4.4). Amanitin substance is a differential inhibitor for RNA polymerases that exist in eukaryotes, by this way, they are considered as a research tool of great importance in differentiating among them.

RNA polymerase III is considered the most sensitive enzyme for amanitin which prevent enzyme transport during transcription.



Figure (4.4). *Amanita phalloides* mushroom from which amanitin toxin is extracted to differentiate among the main RNA polymerases in eukaryotes.

List (4.4): Naming and characterising of mammalian RNA polymerases

Type of RNA polymerase	Sensitivity for α amanitin	Main products
I	Non-sensitive	rRNA
II	Highly sensitive	mRNA
III	Moderately sensitive	tRNA/5S rRNA

Description of Transcription

To simplify studying of transcription, it was divided into three main stages as follows:

Initiation Stage

Transcription begins when RNA polymerase binds with σ factor to produce RNA polymerase holoenzyme. σ factor allows RNA polymerase to specifically bind with promoter sequence of the gene. The process of RNA synthesis is shown in figure (4.5), which includes binding of RNA polymerase holoenzyme with template strand at promoter region. When σ subunit binds with the core enzyme it directs it into promoter by the specific binding with -10 and -35 promoter sequences, this leads to the initiation of transcription at the beginning of the gene. The primary binding between RNA polymerase holoenzyme and promoter is called closed promoter complex because DNA is not unwound, then, RNA polymerase holoenzyme commences unwinding the duplex with 15 base pairs ratio around transcription initiation site to form open promoter complex. The unwinding of the two strands of DNA duplex provides template strand on which RNA polymerase acts in the transcription.

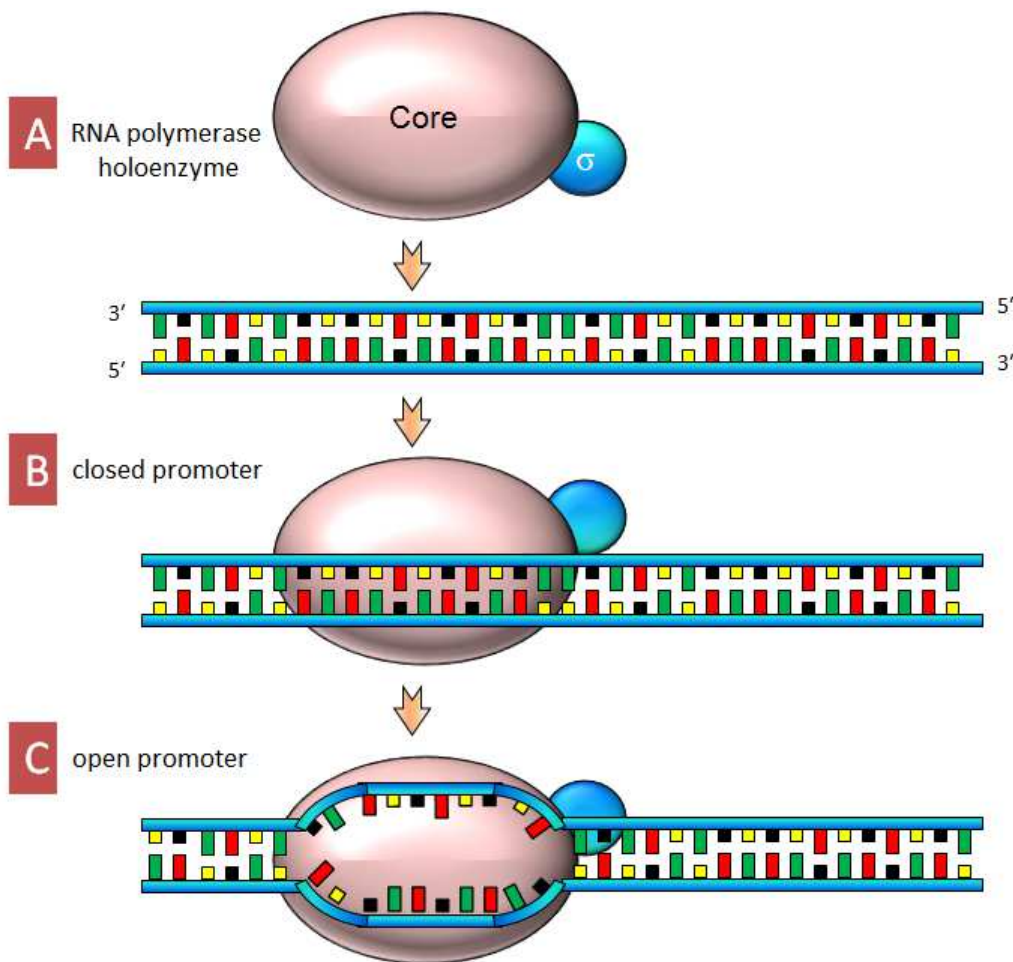


Figure (4.5). A, B, and C. RNA polymerase plus σ factor binding with DNA. Hydrogen bonds that connect the two DNA strands are broken to form promoter complex as a preparation for RNA synthesis.

Elongation Stage

Once σ factor released, elongation stage begins. After addition of the first nucleoside triphosphate (usually purine), RNA polymerase leaves promoter and moves forward along DNA template strand and continues in elongating RNA chain. When it moves ahead, RNA polymerase unwinds the duplex in front of it and rewind the duplex behind of it. Topoisomerases – as in the case of replication – alleviate positive supercoiling generated ahead of transcription bubble, and alleviate negative supercoiling behind transcription bubble (figure 4.6).

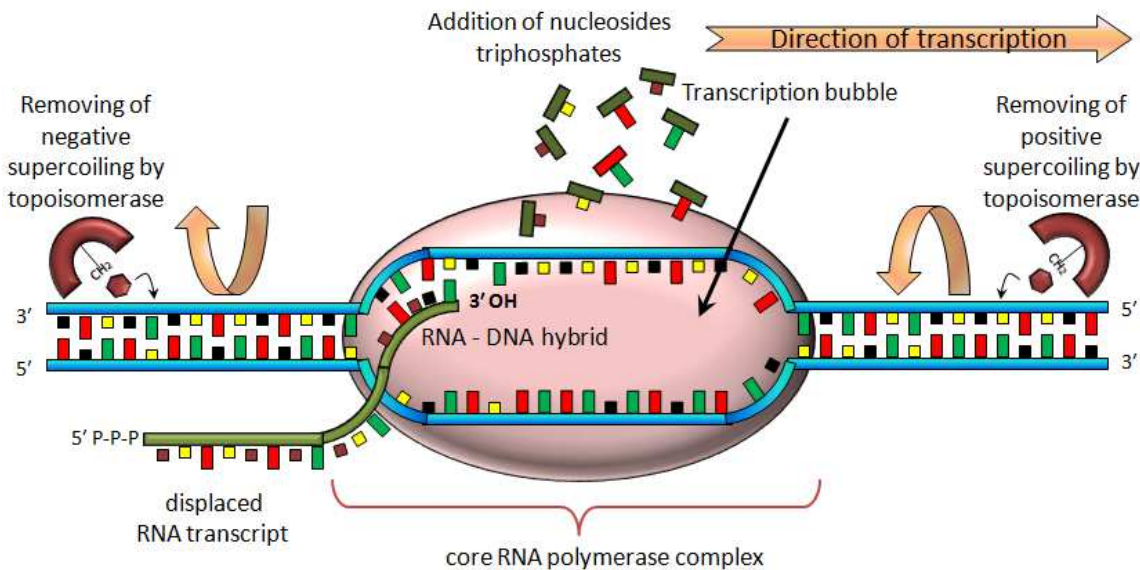
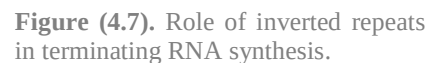


Figure (4.6). Summary of transcription elongation, (remember that RNA polymerase shown in this figure is only core enzyme and not holoenzyme because σ factor is removed once elongation is initiated).

Nucleotides are polymerized in a sequence dictated by the coding strand and deciphered by Watson – Crick base pairing rule. The necessary energy required for RNA synthesis is provided by ribonucleoside triphosphates, because they are considered as the source of energy, which are the building blocks, in addition of their ability to constitute the informational components of the resulting RNA chain (remember this is taken place in replication as well, except for nucleotides being in NTP form instead of being in dNTP form as it happened in replication). Pyrophosphate is released in every polymerization reaction as in the case of DNA synthesis. In elongation, DNA duplex unwinding continues to form transcription bubble of 17 nucleotides moving in 5' to 3' direction. Researchers are estimated the rate of mRNA synthesis, they found it's synthesized with 20 to 50 nucleotides per second, with an error rate estimated as one nucleotide per each one thousand. This error rate is relatively high compared with what is found in replication, but this error rate in mRNA synthesis "can be tolerated" because 1) most of genes are copied repeatedly, so, there is a little possibility of repetition of the same error, 2) there are some redundant genetic codons in which more than one codon encodes the same amino acid (degeneracy of the genetic code). Though nucleotides are changed, the resulting amino acid may not be



83

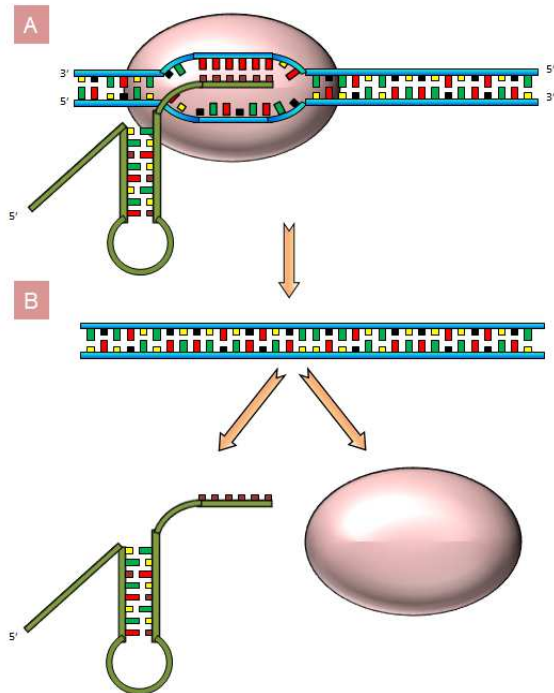


Figure (4.8). *Rho* independent termination of RNA synthesis. (A) RNA polymerase synthesizes poly – U at 3' end of RNA chain. (B) the mRNA and enzyme pull away from the DNA template.

Rho-dependent termination (figure 4.9), which is the less common and more complicated than *rho* – independent

termination, also uses a hairpin RNA formation, but dissociation of the DNA-RNA hybrid needs the assistance of the protein *rho* and no poly-U follows the hairpin as in the *rho* independent method.

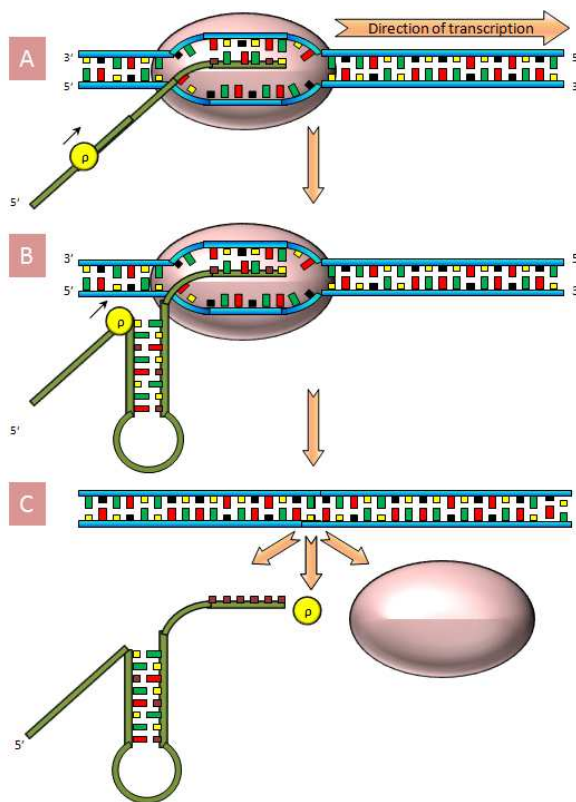


Figure (4.9). *Rho* dependent termination of RNA synthesis. (A) *Rho* attaches RNA as its being synthesized. (B) RNA polymerase pauses after stem-loop is synthesized, allowing *rho* to catch up. (C) By some unknown mechanism, *rho* causes the release of RNA and enzyme.

Rho appears to have helicase activity, so it attaches itself to the RNA while it is still under construction but after σ factor is released. *Rho* moves along the RNA behind the RNA polymerase (it does require in this movement ATP hydrolysis, to provide the energy to move the protein factor along the RNA). The formation of the stem-loop apparently stalls the

polymerase at the loop or shortly after the loop forms. The stalling allows rho to approach the enzyme, *rho* then – using its helicase activity – brings about release of the polymerase and mRNA from the DNA template, *rho* may be needed because of the lack of the poly-U sequence.

After termination of synthesis of the RNA, the enzyme separates from the DNA template and probably dissociates to free core enzyme and free factor. With the assistance of another σ factor, the core enzyme then recognizes a promoter at which the synthesis of a new RNA molecule commences. More than one RNA polymerase may transcribe the same template strand of a gene simultaneously, but the process is phased and spaced in such a way that at any given moment each one is transcribing a different portion of the DNA sequence. As long as RNA polymerase leaves promoter, another RNA polymerase binds with it, this gives a shape like a blade (figure 4.10).

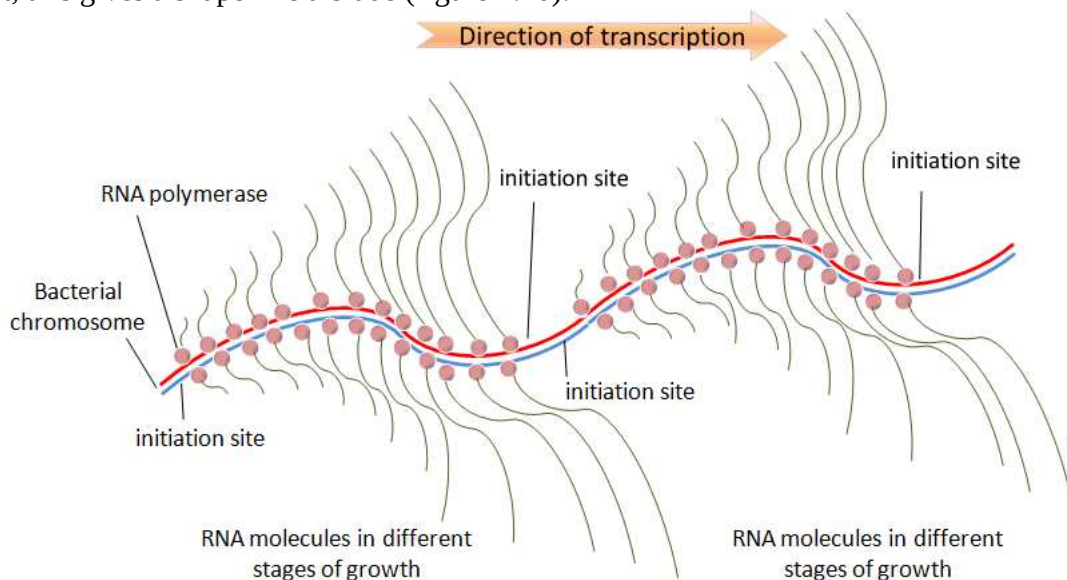


Figure (4.10). Illustration of multiple copies of amphibian ribosomal RNA genes in the process of being transcribed. Note that the length of the transcripts increases as the RNA polymerase molecules progress along the individual ribosomal RNA genes from transcription start sites to transcription termination sites. RNA polymerase (not visualized here) is at the base of the nascent RNA transcripts. Thus, the proximal end of the transcribed gene has short transcripts attached to it, while much longer transcripts are attached to the distal end of the gene. The arrows indicate the direction (5' to 3') of transcription. This figure shows also the ability to synthesize RNA with large amount incase its needed.

RNA Processing

Most of the newly synthesized RNA molecules must be modified with variable means in order to be converted to their functional forms except the bacterial mRNA which represents an exception for this rule, it used immediately as a template to synthesized proteins while transcription being taking place on it. The primary transcripts of tRNA and rRNA in prokaryotes and eukaryotes must undergo series of processing. The primary transcripts of eukaryotic mRNA also undergo extensive modifications which include removing introns by splicing. Then, it transported from nucleus to cytoplasm to act as

template for protein synthesis. Here, we will focus on processing that taken place on eukaryotic mRNA instead of tRNA or rRNA since mRNA is the molecule that carries the necessary genetic information in protein synthesis.

Processing of Messenger RNA

In contrast to rRNA and tRNA processing, mRNA processing represents a major difference between prokaryotes and eukaryotes. In bacteria, ribosome has the ability to act on mRNA directly, where translation on nascent mRNA chain starts in the time at which transcription still act on other portions of incomplete mRNA. Thus, transcription and translation are coupled in the same time (figure 4.11).

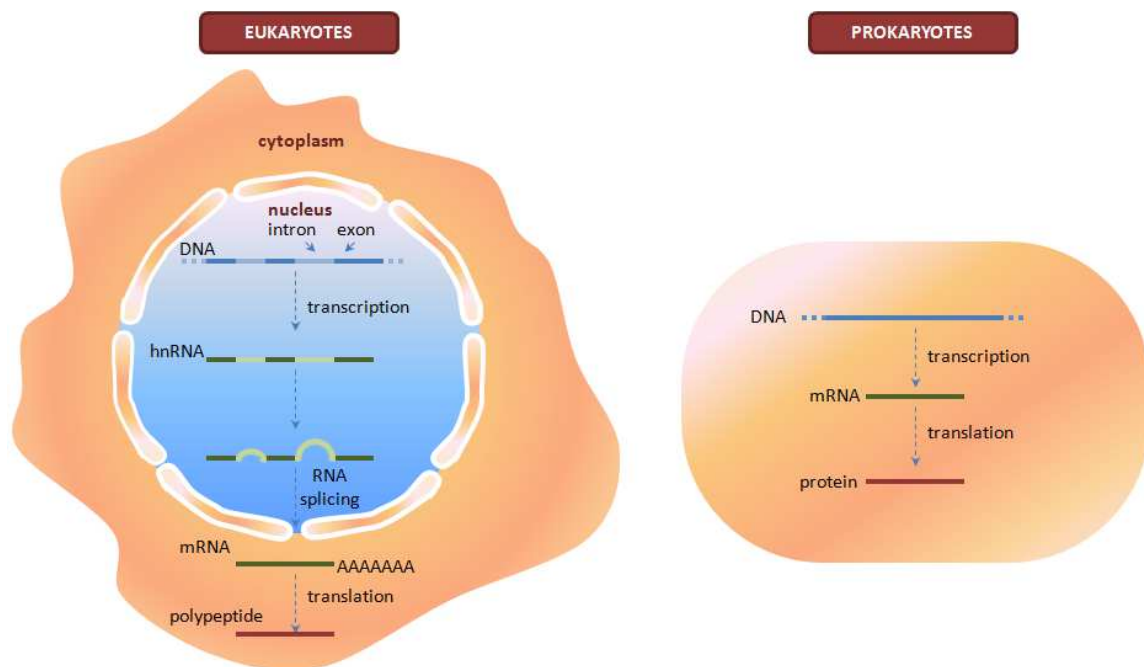


Figure (4.11). Coupling of transcription and translation in prokaryotes (right) and their separation in eukaryotes (left).

In eukaryotes, the synthesized mRNAs must be transported from nucleus to cytoplasm before its being used as template for protein synthesis. Moreover, the primary transcripts (pre-mRNAs) in eukaryotic cells must be modified before they are being exported from nucleus to cytoplasm (figure 4.12).

The 5' end of pre-mRNA is immediately modified after its synthesis by adding the cap. This cap facilitates the connection of eukaryotic mRNAs with ribosomes at translation initiation (see chapter five). Add to that, it stabilizes mRNA, and its noticed that mRNAs that don't have cap have a less half life.

Most of eukaryotic mRNAs have poly (A) tail at their 3' end (figure 4.12), that is added by a processing reaction called polyadenylation. In which an enzyme called poly A polymerase – that don't require template – adds poly (A) tail of 220 to 250 nucleotides in length at the end of transcript. This enzyme starts adding poly (A) beside the sequence ...AAUAAA... that located beside 3' end of the nascent mRNA. The function of this tail is not exactly known, but it seems that after transfer of mature mRNA to cytoplasm, the tail length is shortened somehow to be only in 100 to 150 nucleotides. Hence, the sensitive role of the tail in protecting genetic material against the attack of endonucleases and exonucleases become obvious during its transfer from nucleus to cytoplasm. The most exciting modification of pre-mRNA is the removing of introns by splicing. As it is discussed above, the coding sequences for most eukaryotic genes are interspersed with noncoding sequences (introns), which excised from the mature mRNA. Then, the non-excised portions (exons) are connected to each others to form the so called mature mRNA (figure 4.12).

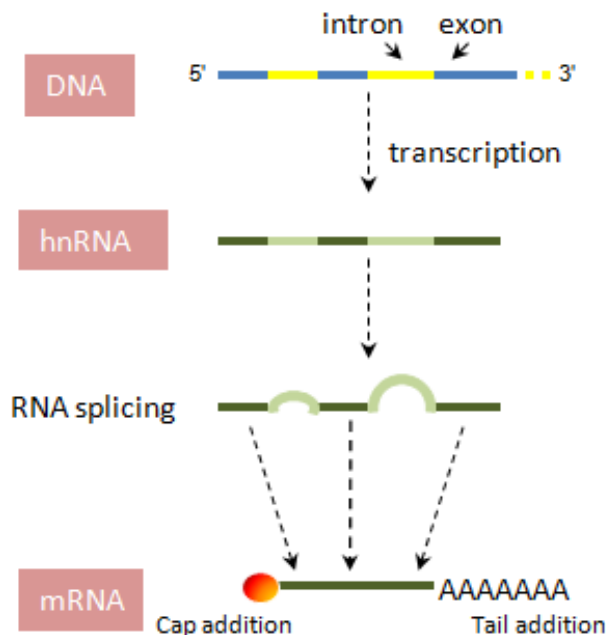


Figure (4.12). Illustration for mRNA processing in eukaryotic organisms. It's obvious from the figure after RNA polymerase II completes transcription of the first nucleotide of the first exon of the gene; a cap is added to the 5' end of RNA. RNA polymerase II is terminated at termination site of the last exon, and approaches a site called poly (A) site, which is located at 3' end of primary transcript, at which an Adenine extension is added (Poly (A) tail contains 250 nucleotides in mammals). Then, introns are removed and exons connected with each others in splicing.

Splicing of Hetero-nuclear RNA (hnRNA)

Introns are removed from hnRNA after transcription immediately and exons are connected to each others to form a continuous noncoding sequence. This process is called splicing, which is taken place by protein – RNA complexes in the nucleus that is called spliceosomes. The type of RNA that binds with the proteins to form spliceosome is called small nuclear RNA or snRNA (see types of RNA), which is localized in different five forms (U1, U2, U4, U5, and U6). Each one of these forms consists of multiple proteins and one molecule of snRNA (figure 4.13).

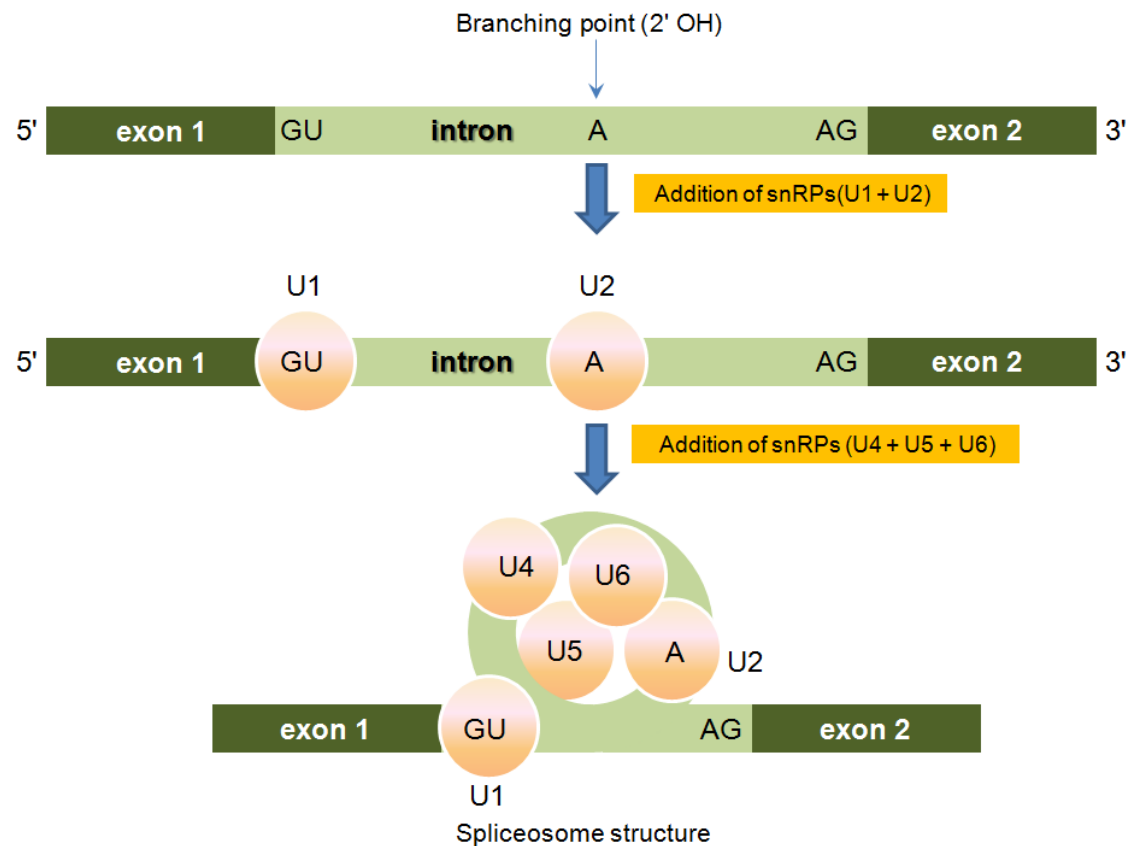


Figure (4.13). The mode of spliceosome formation.

In order to ensure not to destroy the message of mRNA, splicing has to occur in a very precise fashion. The start and the end of the hnRNA introns are recognized by a characteristic sequence at the 3' end. Another important structure is the so-called branching point inside the intron. Its sequence is less conserved than the terminal splicing sites, but it always contains one adenosine residue (A). During splicing, the 2'-OH group of this residue - supported by the spliceosome - attacks the phosphoric acid diester bond at the 5' end of the intron and cleaves it (figure 4.14 step A and step B). Simultaneously, an unusual 2' to 5' bond is formed inside the intron, which thereby takes on a lasso shape (figure 4.14). In the second step of the splicing process, the free 3'-OH group at the end of the 5' terminal exon attacks the A-G bond at the 3' end of the intron. As a result, the two exons are linked and the intron is released, still in a lasso shape. The last step is the diluting out and vanishing of the lasso shape.

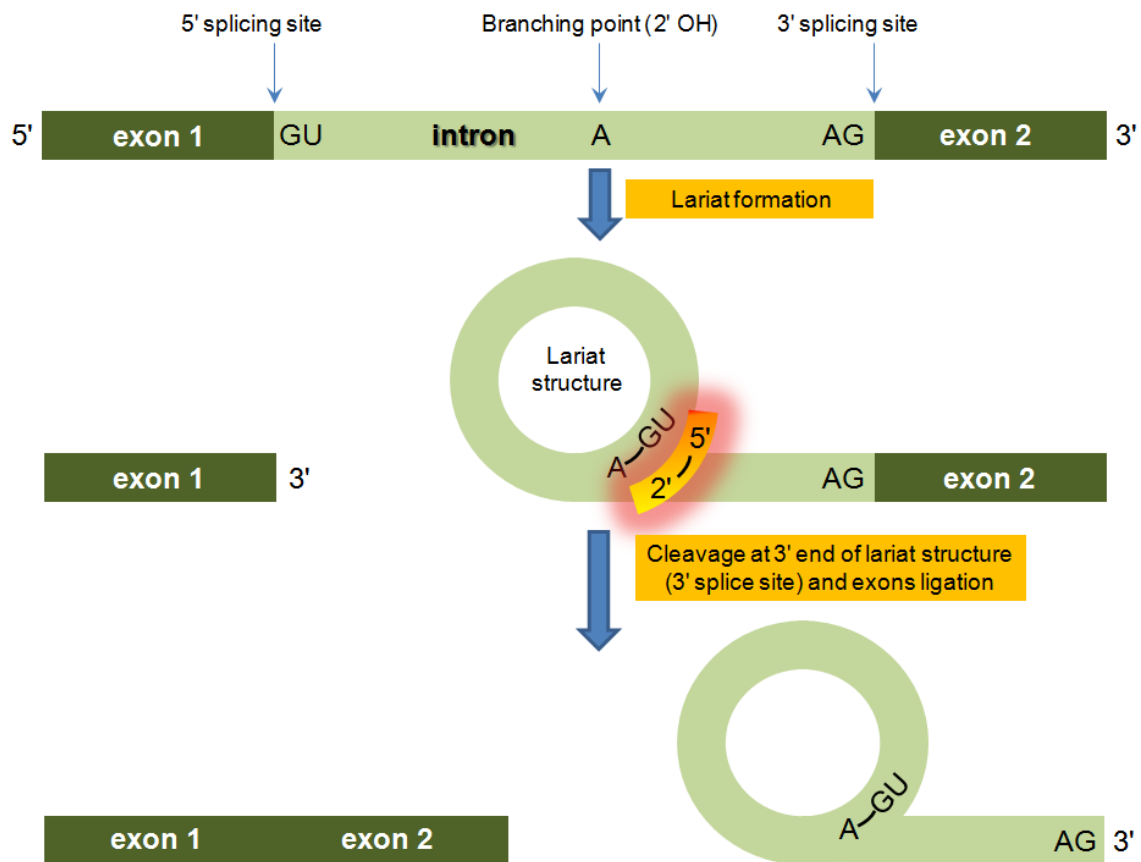


Figure (4.14). Splicing; in which introns are removed and exons are connected with each other by formation of lasso shape in each intron before its being removed.

RNA Types

There are three main types of cellular RNA involved in protein synthesis which are: mRNA, tRNA, and rRNA. Moreover, other types of RNA acting as primers to synthesize DNA are available, which are produced from RNA primase (review chapter three). Add to that, other types most important of them is snRNA are available as well.

Messenger RNA

The sequence of mRNA contains the genetic information that translated into amino acid sequences in protein. Prokaryotic cells have unstable mRNA molecules which end with a half life average about 1 – 3 minutes, while in eukaryotes, mRNA is recognized with more stability, so, the typical mRNA age in eukaryotes reaches to 3 hours, why? Eukaryotic mRNA must be transported from nucleus to rough endoplasmic reticulum in cytoplasm where proteins are synthesized. This process needs gathering of other proteins with mRNA to help it in this transportation. This process costs some time compared with prokaryotic mRNA which its translation begins while transcription still acting on it. The mRNA molecule constitutes about only 5% out of the total cellular mRNA.

Although mRNA lengths are variable, it can be said that the amount of genetic information that carried by mRNA in prokaryotes are sufficient to code for more than one gene, while this genetic information that carried by eukaryotic mRNA are sufficient to code for only one gene (figure 4.15). And for sure if mRNA molecule for only one gene (as in eukaryotes), this necessarily lead to production of one protein, and this case called monocistronic. If mRNA molecule is for many genes (as in prokaryotes), this lead to production of several proteins, and this case is called polycistronic.

In prokaryotes, as it is clarified above, mRNA is engaged with protein synthesis even before of its being completed entirely. The rapid utilization of the nascent mRNA copy reduces its processing chances. By contrast, in eukaryotic organisms, where transcription is separated from translation acutely, the nascent mRNA copy undergoes an extensive processing before its being transported into cytoplasm to be used as template for translation.

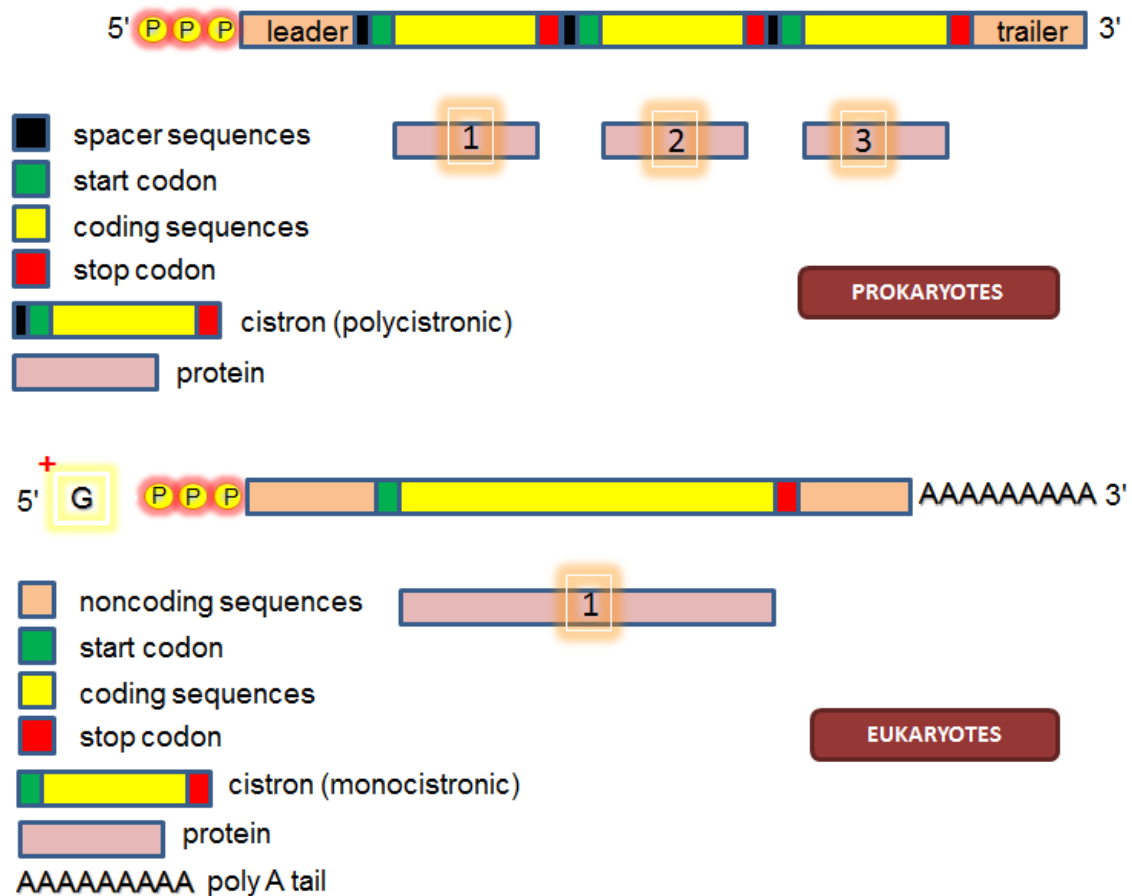
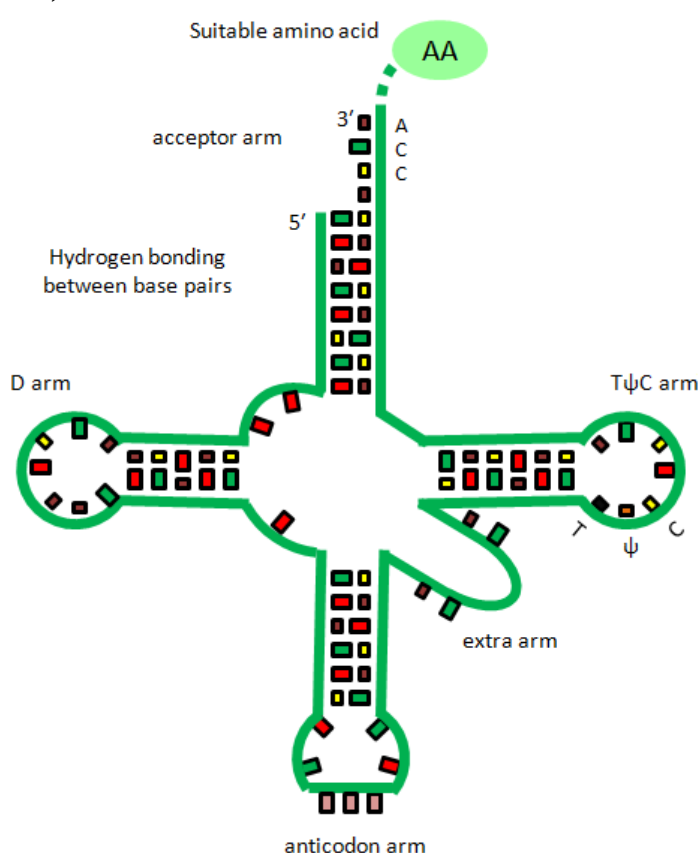


Figure (4.15). Comparison between prokaryotic and eukaryotic mRNA structure. Although both mRNAs are synthesized with a triphosphate group at the 5' end, the eukaryotic RNA molecule immediately acquires a 5' cap, which is part of the structure recognized by the small ribosomal subunit. Protein synthesis therefore begins at a start codon near the 5' end of the mRNA. By contrast, in prokaryotes, the 5' end has no special significance, and there can be multiple ribosome-binding sites (called Shine-Dalgarno sequences) in the interior of an mRNA chain, each resulting in the synthesis of a different protein.

Transfer RNA (tRNA)

The second main type of RNA is tRNA, which is the smallest type of RNA in its length, which is estimated about 76 nucleotides in prokaryotes. Its function is summarized by transfer of amino acids into mRNA template, in which amino acids bind by a specific manner. What makes this molecule competent to take the responsibility of this crucial task is its containing of the both positions; amino acids binding position which is known as acceptor arm that is found at 3' end, and anticodon arm of position that recognizes the corresponding codon (see chapter five) of three bases that exists in mRNA molecule (figure 4.16).



There are about 20 different types of tRNA in each cell. Each amino acid enzymatically binds to 3' end of one or more than one tRNA molecule by enzymatic reaction will be discussed in the next chapter. Though tRNA molecules are different from each other by their nucleotide sequences, they are many common features among them. All tRNA molecules fold on themselves in a similar way to give a structure resembles cloverleaf (figure 4.16). The tRNA molecules constitute more than 15% of the total cellular tRNA.

Figure (4.16). Transfer RNA fine structure.

All tRNA types consist of four main arms. Acceptor arm consists of stem of nitrogen bases coupled with each other. This arm ends with the sequence (3' to 5') CCA, in which the carboxyl group of amino acids binds with hydroxyl group that exist in A residues at 3' end of acceptor arm of tRNA molecule. The second arm consists of stem with nitrogen bases coupled with each others in addition to unpaired loop (figure 4.16). Anticodon arm involved in codon recognition (the trinucleotides) that exists in the mRNA molecule. The third arm is called D arm because the existence of dihydrouridine, which is the arm with which aminoacyl-tRNA synthetase binds during its addition of the suitable amino acid for tRNA molecule (see chapter five). The fourth arm is called TψC arm because it has the

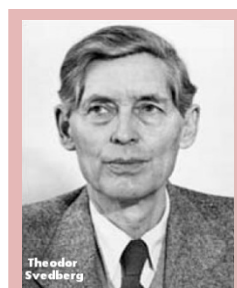
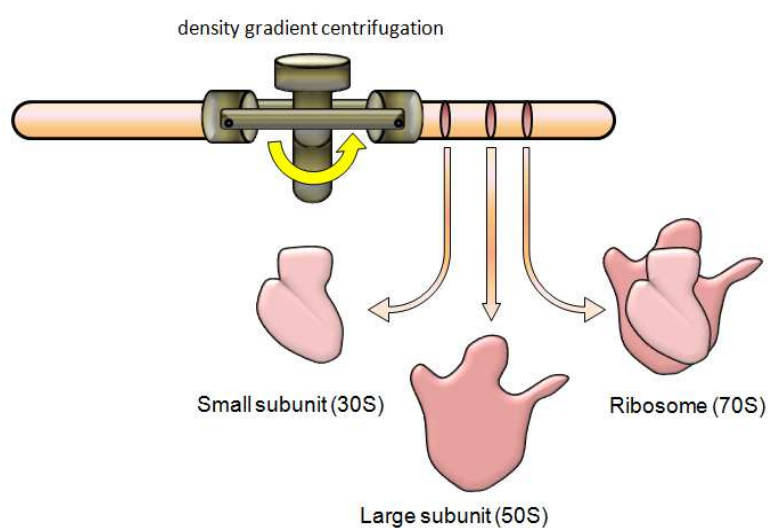
sequence thymine (T), pseudouridine (ψ), and cytosine (C). T ψ C arm is involved in aminoacyl tRNA binding with ribosome surface during protein synthesis. There is additional arm (extra arm) which is considered as the most variable arm among tRNA molecules, therefore it provides a principle for classification, and thus it's called sometimes for this reason variable arm.

It's obvious from the discussion above that there is a characteristic feature for tRNA molecule which is the high ratio of ribonucleotide bases that differ from the main four usual bases as in dihydrouridine and pseudouridine which are resulted because that are resulting from the extensive enzymatic modification that tRNA molecule submitted for by several participating enzymes in the processing of this molecule. And because of the ability of this molecule to form secondary and tertiary structures through many folding that relatively similar to proteins folding, Francis Crick was commented on this molecules saying "RNA is a molecule try to be a protein". The same thing can be said about rRNA (see appendix of this chapter).

Ribosomal RNA (rRNA)

Molecules of rRNA are the structural and functional component of ribosomes. Ribosomes are large bodies of rRNA and proteins which constitute the machinery that harbors most of the reactions that are related with protein synthesis.

Ribosome has a much known structure whether in prokaryotes or eukaryote which is of two subunits each one consists of RNA and protein. Both large and small subunits are composed of ribosomal RNA (rRNA) and many ribosomal proteins. The large ribosomal subunit contains the peptidyl transferase center that is in charge on peptide bonds formation. The small ribosomal subunit contains the decoding center, in which tRNA molecules read or decode codons in mRNA. As its common, ribosomal subunits are called large and small according to the speed of their sedimentation when they are exposed



Theodor Svedberg

to a centrifugation force. The unit used in measuring sedimentation velocity is known as Svedberg unit which is named after to Theodor Svedberg, the ultraspeed centrifuge inventor (figure 4.17).

Figure (4.17). Sedimentation by ultraspeed centrifuge in order to separate ribosomal subunits and the entire ribosome in prokaryotes.

In bacteria, the ribosomal subunit that has 50 Svedberg units sedimentation rate is known as 50S, whereas the small ribosomal subunit that are known as 30S. The intact bacterial ribosome is called 70S ribosome. Note that 70S ribosome is less in sum of 50S plus 30S! The explanation of what is seemed as apparent contradiction in this issue is the velocity of sedimentation is determined by shape and size. Hence, it is not a measurement for the mass. The eukaryotic ribosome is somewhat bigger than that in prokaryotic counterpart; it consists of 60S and 40S that constitute with each other 80S ribosome. This means that the most abundant cellular RNA is rRNA molecules.

In turn, large ribosomal subunit 50S in prokaryotic ribosomes consists of 23S and 5S in rRNA molecules. While the small ribosomal subunit consists of rRNA molecules of only 16S rRNA. As for eukaryotic ribosomes, they are similar in structure to these found in prokaryotes, although they are relatively larger (80S), and contains of larger rRNA and includes 5S rRNA, 18S rRNA, and 25S rRNA and others. Thus, there are three recognized types of rRNA molecules in prokaryotes (5S, 16S, and 23S). The molecules of rRNA constitute 80% of total cellular RNA, so, this RNA is the most available type of RNA existing in the cell.

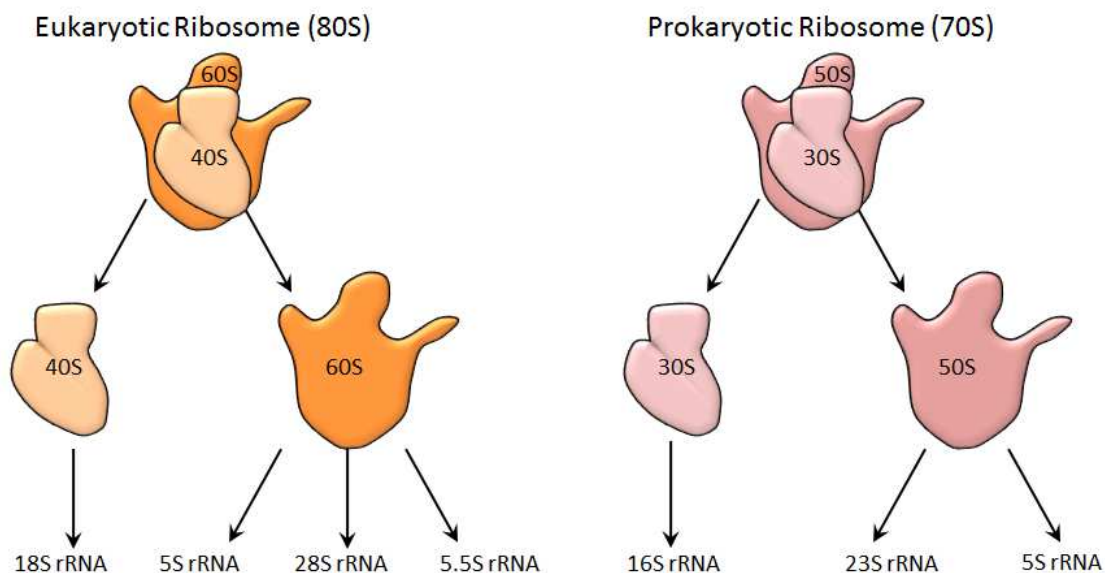


Figure (4.18). Comparison between the structure of ribosomes in prokaryotes and eukaryotes. Ribosomal components are initially classified according to S values which refers to their sedimentation rate, although differences are exist in the numbers and sizes of protein components rRNA. Both types of ribosomes have the same structure and act in a very similar ways

Small stable RNA (ssRNA)

SsRNA are small molecules exist in eukaryotic cells. Most of these molecules form complexes with proteins to form what is called ribonucleoproteins that are distributed in nucleus or cytoplasm or both.

Small nuclear RNAs (snRNA) are subgroup of this RNA which involved in mRNA processing and gene regulation. It's of several types (see above) involved in removing and

processing of hnRNA into mRNA. As it's explained in this chapter, with the escorting proteins, this type of RNA form a spliceosome complex, that removes introns and attach exons with each others. In this complex, the protein components don't play a fundamental role but only accessory role, whereas the main role of catalysis of such enzymatic reactions are represented by the RNA component, these molecules called ribozymes (see appendix of this chapter).

Chapter Four Appendix

The Ribozyme

It was widely believed for many years that only proteins could be enzymes. An enzyme must be able to bind with a substrate and, carry out a chemical reaction, release the product and repeat this sequence of events many times. Proteins are well suited for this task since they are composed of many different kinds of amino acids and they can fold into complex tertiary structures with binding pockets for the substrate and small molecule co-factors and an active site for catalysis. Now it's known that RNAs, which as it have seen can similarly adopt complex tertiary structures, can also be biological catalysts. Such RNA enzymes are known as ribozymes, and they exhibit many of the features of a classical enzyme, such as an active site, a binding site for a substrate, and a binding site for a co-factor, such as a metal ion.

The discovered ribozymes is RNase P, a ribonuclease that is involved in generating tRNA molecules from larger precursor RNAs. RNase P is composed of both RNA and protein. Although proteins exist, RNA alone is the catalyst, and the protein components in RNase P facilitate the reaction only by keeping the negative charges in RNA. Many other RNA molecules are involved in intron removing reactions from precursor molecules for each mRNA, tRNA, and rRNA by the above mentioned splicing. Ribozyme that is involved with such reactions undergoes a characteristic shape called hammerhead; which is composed of three base paired stems surrounded by a central part composed of nucleotides non-complementary to each others that are necessary for catalysis (figure 4.19). There are many other examples for the existence of RNA molecules have catalytic ability as in telomere addition reaction that is by RNA component in telomerase (see chapter three), elongation reaction that catalyzed by terminal transferase (see chapter five).

Ribozyme discovery has changed scientists view about the origin of life. We can figure out how there was a primitive form of life that completely depending on RNA. In this world, RNA acts as genetic material and as enzymatic machinery. This RNA world may precede the life we know today, in which the transfer of genetic information depends on DNA, then RNA, then protein. The hypothesis which saying that protein

world was emerged from RNA world is originated from the discovery of ribosomal component that is responsible on peptide bond, is peptidyl transferase which is only RNA molecule (see chapter five). In contrast to the ribozyme RNase P, hammerhead, and other know ribozymes that are acting on phosphorous centers, the ribozyme peptidyl transferase is acting on the carbonyl center to synthesize the peptide bond. Thus, the RNA chemistry is bounded with the most important reaction in the proteins world, which is the peptide bond formation. The ribosomal ribozyme may be just remains for a form of very old life in which all its enzymes are mere RNA molecules.

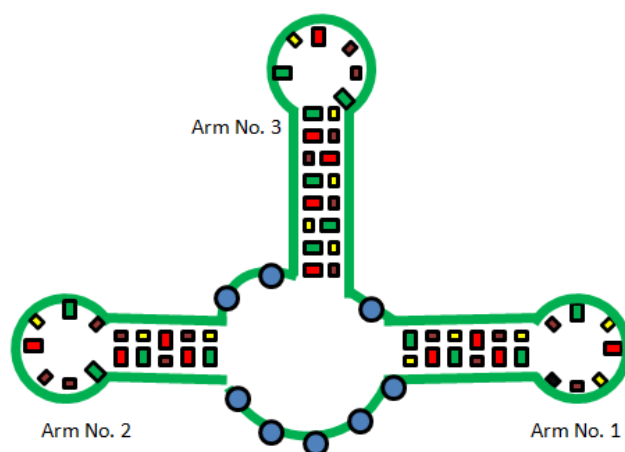


Figure (4.19). Hammerhead structure. Blue bulls refer to variable nucleotides.

Chapter Four Revision Questions

Q1- Justify the following:

1. RNA folds into multiple forms while this state is not usually taken place for DNA?
2. It's not necessary either for G to be equal to C nor for T to be equal for A in RNA?
3. There is proofreading in most DNA polymerases whereas this mechanism does not exist in RNA polymerases?
4. The region -35bps is called downstream domain in prokaryotic promoter while the region -10bps is called unwinding domain in the same promoter?
5. In prokaryotic promoter, the region -10bps is consist only of A and T exclusively?
6. Eukaryotic promoter is much more complicated than its prokaryotic counterpart?
7. Prokaryotic cells have unstable mRNA molecules which end with 1 – 3 minutes half life whereas in eukaryotic cell, mRNA is characterized with more stability?
8. Transcription is phased and spaced?
9. Enhancer sequences are promiscuous?
10. Core DNA polymerase requires to σ in transcription initiation?
11. Most of newly synthesized RNA molecules are modified with variable modes to take their functional shapes except bacterial mRNA that represent an exception for this rule?
12. The sequence poly (A) in DNA as a mean to end transcription?
13. In transcription termination, there is a tight relationship between stem loop structure and cellular need or not need to *rho* subunit?
14. The molecules of tRNA are called intermediate molecules or adaptors?
15. The summation of 30S and 50S is 70S but not 80S?

Q2- choose the correct answer:

1. RNA polymerase is called as well.
a. DNA dependent DNA polymerase b. DNA dependent RNA polymerase c. RNA dependent DNA polymerase d. RNA dependent RNA polymerase
2. RNA polymerase commits one mistake for each nucleotide.
a. 100 b. 1000 c. 10000 d. 100000
3. the closest position of transcription start site is
a. -100 bps b. -10 bps c. +1000 bps d. +9 bps
4. The eukaryotic promoter sequence TATAAA in eukaryotes is called.....
a. Pribnow box b. Hogness box c. CAAT box d. Recognition box
5. The number of RNA polymerases that can be catalyzes the synthesis of each rRNA, tRNA, and mRNA in prokaryotes
a. is only one type b. are two types c. are three types d. are four types
7. are the components of RNA polymerase in initiation of transcription.
a. $\alpha 2\beta$ b. $A2\beta\beta'$ c. $A2\beta\beta'\sigma$.Only σ
8. are the components of RNA polymerase in elongation transcription.
a. $\alpha 2\beta$ b. $A2\beta\beta'$ c. $A2\beta\beta'\sigma$.Only σ
9. The subunit In RNA polymerase has most or all active sites to form phosphodiester bond.
a. α b. β c. γ d. σ
10. the function of RNA polymerase III is the synthesis of
a. rRNA b. mRNA c. tRNA d. snRNA
11. the enzyme RNA polymerase II is considered For α -amanitin that is extracted from *Amanita phalloides* mushroom.
a. insensitive b. intermediately sensitive c. highly sensitive d. may be sensitive
12. The template strand of 5'-GATCG-3' sequences gives RNA strand with the sequence.
a. 5'-GATCG-3' b. 5'-CTAGC-3' c. 3'-CUAGC-5' d. 3'-CTAGC-5'
13. The enzyme..... is responsible on unwinding ahead and rewinding behind it during transcription.
a. helicase b. DNA polymerase c. RNA polymerase d. topoisomerase
14. The number of accessory factors that rho independent termination requires

- a. is one b. is two c. are three d. none
15. one of σ factor types is Which produced under nitrogen starvation conditions.
a. $\sigma 70$ b. $\sigma 32$ c. $\sigma 54$ e. $\sigma 45$
16. The Is involved in connection of aminoacyl tRNA molecule with the surface of ribosome during protein synthesis.
a- extra arm b- amino acid arm c- acceptor arm d- T ψ C arm
17. all tRNA molecules fold on itself in a similar way to give a structure resemble to
a- banana b- orange c- cloverleaf d- strawberry
18. stem loop structure contains Uracil residues behind it directly.
a- 5 b- 6 c- 7 d- 8
19. both of units constitute the catalytic center in RNA polymerase.
a. β and β' b. α and α' c. α and β d. β and α'
20. most of RNA molecules that have modified nucleotides are
a. mRNA b. rRNA c. tRNA d. ssRNA
21. one of tRNA arms which is Is used as a basics to classify different types of tRNA molecules.
a. T ψ C arm b. D arm c. acceptor arm d. extra arm
22. *Rho* independent termination is characterized by the existence of In addition to stem loop structure in RNA which needs low energy to break it.
a. poly A c. poly C d. poly G d. poly U
23. the most existing types of RNA in the cell is
a. mRNA b. tRNA c. rRNA d. ssRNA
24. the Is not modified after its transcription.
a. prok. tRNA b. euk. tRNA c. prok. mRNA d. euk. mRNA
25. Because, its so easy to isolate mRNA out of rRNA and tRNA that commonly exist in the cellular extract.
a. T ψ C arm. b. acceptor arm c. D arm d. polyA tail

Q3- define the following:

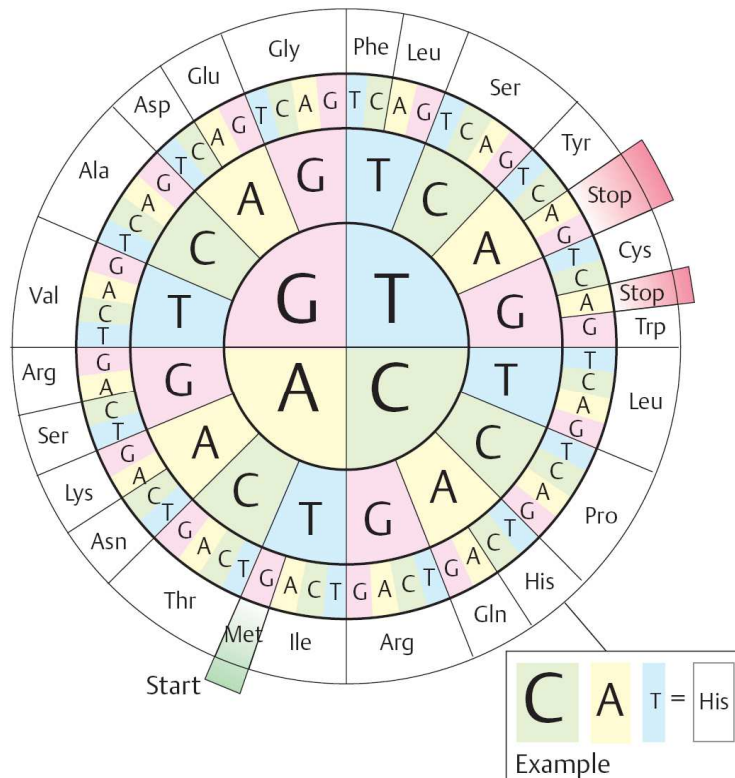
promoter, Pribnow box, spliceosome, polyadenylation, poly A polymerase, monocistronic.

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science, USA, 2008.
- Berg J.**, Tymoczko J., Stryer L. Biochemistry. Fifth edition, W. H. Freeman and company and Sumanas, Inc. , 2005.
- Bloomfield V.**, Crothers, D. M., Tinoco, I., and Hearst, J., 2000. Nucleic Acids: Structures, Properties, and Functions. University Science Books.
- Brenner S.**, Jacob F., and Meselson M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis *Nature* 190: 576-581.
- Cooper J.** and Hausman R. The cell, A Molecular Approach. Fourth edition, Sinauer Associates, 2007.
- Fiering S.** Whitelaw E., and Martin D. 2000. To be or not to be active: The stochastic nature of enhancer action *Bioessays* 22: 381-387.
- Gesteland, R. F.**, Cech, T., and Atkins, J. F. (Eds.), The RNA World (2d ed.). Cold Spring Harbor Laboratory Press, 1999.
- Gilbert H.** Basic Concepts in Biochemistry. A student survival guide. Second edition, McGraw-Hill Companies, 2000.
- Gott J.** and Emeson R. 2000. Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* 34: 499-531.
- Jacob F.** and Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* 3: 318-356.
- Hall B.**, and Spiegelman S. 1961. Sequence complementarity of T2-DNA and T2-specific RNA *Proc. Natl. Acad. Sci. U.S.A.* 47: 137-146.
- Koolman J.**, and Roehm K. Color Atlas of Biochemistry. Second edition, Thieme, 2005.
- Lewin B.** Genes. VII, Oxford University Press, 2000.
- Paolella. P.** Introduction to molecular biology. First edition, WCB/McGrawHill, 1998.
- Passarge E.** Color atlas of Genetics. Third edition, Thieme, 2007.
- Murray R.**, Bender D., Botham K., Kennelly P., Rodwell V., Weil P. Harper's Illustrated Biochemistry. 28th edition, McGrawHill Medical, 2009.
- Newman A.** 1998. RNA splicing. *Curr. Biol.* 8: R903-R905.
- Pierce B.** Genetics: A conceptual approach. 2002.
- Ro-Choi T.** 1999. Nuclear snRNA and nuclear function (discovery of 5 cap structures in RNA) *Crit. Rev. Eukaryotic Gene Expr.* 9: 107-158.
- Shatkin A.** and Manley J. 2000. The ends of the affair: Capping and polyadenylation *Nat. Struct. Biol.* 7: 838-842.
- Snyder L.**, and Champness W. Molecular genetics of bacteria. Third edition, ASM Press, Washington D C, 2007.
- Watson J.**, Baker T., Bell S., Gann A., Levine M., Losick R. Molecular biology of the gene. Fifth edition, Pearson Prentice Hall, 2004.

Chapter Five

Gene Translation



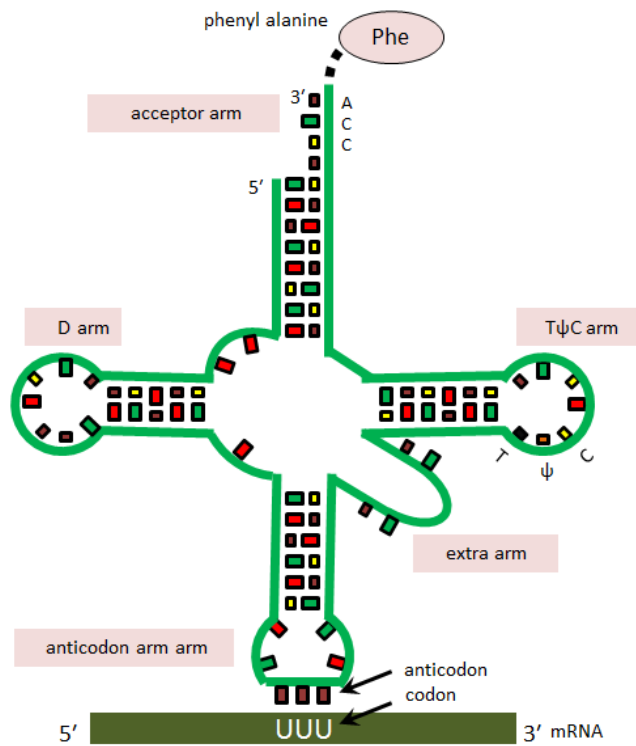
The genetic code list after its deciphering (Koolman and Roehm, 2005)

Introduction

The language of life consists of four alphabetic letters: A, G, T, and C. These letters are corresponded to nucleotides that are existed in DNA. Theses nucleotides are arranged in three letters codons. The summation of codons forms what is known the genetic code. Linearly arranged codons (genes) code for synthesis of many RNA molecules, most of them are involved in some protein synthesis activities. Protein synthesis is taken place by three main steps; initiation, elongation, and termination. This process resembles replication and transcription in its general features and in its direction (5' to 3') as well.

Cell must have the necessary machinery that translates information efficiently and precisely from nucleotide sequences in RNAs into amino acid sequences in the corresponding proteins. Some researchers have early realized that mRNA molecules themselves don't have affinity for amino acids, thus, the translation of genetic information that exists in mRNA into amino acid sequences in protein requires an intermediate molecule capable of adaptation between nucleic acids and amino acids.

These adaptor molecules must recognize the specific nucleotide sequence from one side, and its corresponding amino acid sequence from the other side (see chapter four; tRNA). With the existence of such adaptor molecules, the cell can direct the specific amino acid into its right place in protein according the sequences as it is determined in the nucleotide sequences that available in mRNA molecule (figure 5.1).



The tRNA molecule – as an adaptor molecule – utilizes anticodon arm in recognition of codons of mRNA depending on Watson – Crick base pairing rules. Each tRNA contains a certain sequence complementary to codon, which called anticodon. Since each tRNA is labeled with only one type of amino acid, each codon is specialized with only one type of amino acid.

Figure (5.1). the ability of tRNA molecule to connect between nucleic acids through anticodon arm from one side, and amino acids through acceptor arm from another side. The anticodon region consists of hepta-nucleotide which are N; the variable, Pu*; the modified purine, X,Y,Z; the triple nucleotides which are AAA as an example in this figure, and two Pyr (pyrimidine) from the direction 3' to 5'.

Codons and Protein Synthesis

The sequences that exist in mRNA are existing in each amino acid (figure 5.2). The adaptor molecules that translate codons into amino acid sequences in proteins are tRNA molecules. Ribosomes are considered as cellular components on which different functional entities are interacted to synthesize protein molecule. Many ribosomes are assembled on each other spontaneously to translate single mRNA molecule, and form what is called polyribosome (polysome) (see figure 4.9). Polysomes are attached on the surfaces of certain cytoplasmic chambers to do their task of protein synthesis. These chambers called endoplasmic reticulum and they turned rough because the existence of ribosomes on their surfaces.

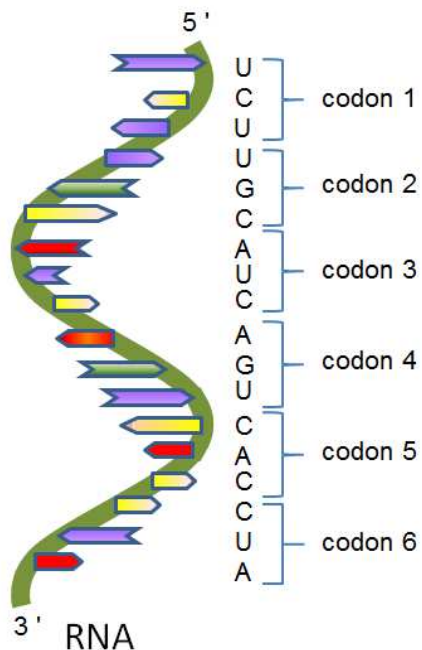
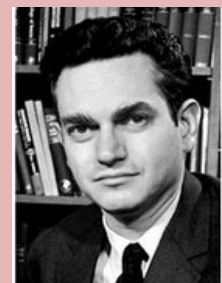


Figure (5.2). the mode of RNA sequences arrangement into different codons ready to be translated into protein. Each three nucleic acids constitute one codon which codes in turn for only one amino acid.

Cracking the Genetic Code

Scientists at the end of 1950s and beginning of 1960s were managed to solve one of the important life secrets, which is how genes acting. The problem that researchers tried to solve is how the linear sequence of four nucleotides (A, G, C, and U) is able to determine amino acid sequence in a protein consisting of 20 different amino acids. After discovery of DNA structure by James Watson, Francis Crick, and Rosalind Franklin, serious attempts were emerged to understand the nature of coding to protein. The first deciphering of the genetic code was done on the hands of Nirenberg and Mathaei at 1961.



Marshall Nirenberg

We owed to Nirenberg and Mathaei for their experiments that performed on *E. coli* extracts after removing DNA by DNase in order not to interfere with the results of these experiments. When they used a laboratory synthesized polynucleotide that is made only from Uracil (poly U) using polynucleotide phosphorylase (which randomly polymerizes without depending on template strand, hence it differs from all polymerases that never work without the presence of template strand to dictate them), and they added it into the medium. A polypeptide containing only phenylalanine (poly-phenylalanine) was obtained (figure 5.3 A). Similarly, using the same enzyme with changing of substrate, the poly (A) formed only poly-lysine, poly (C) formed poly-proline, and poly-proline formed only poly-proline. Therefore, UUU codes for phenylalanine, AAA codes for lysine, and CCC codes for proline (figure 3.5 B).

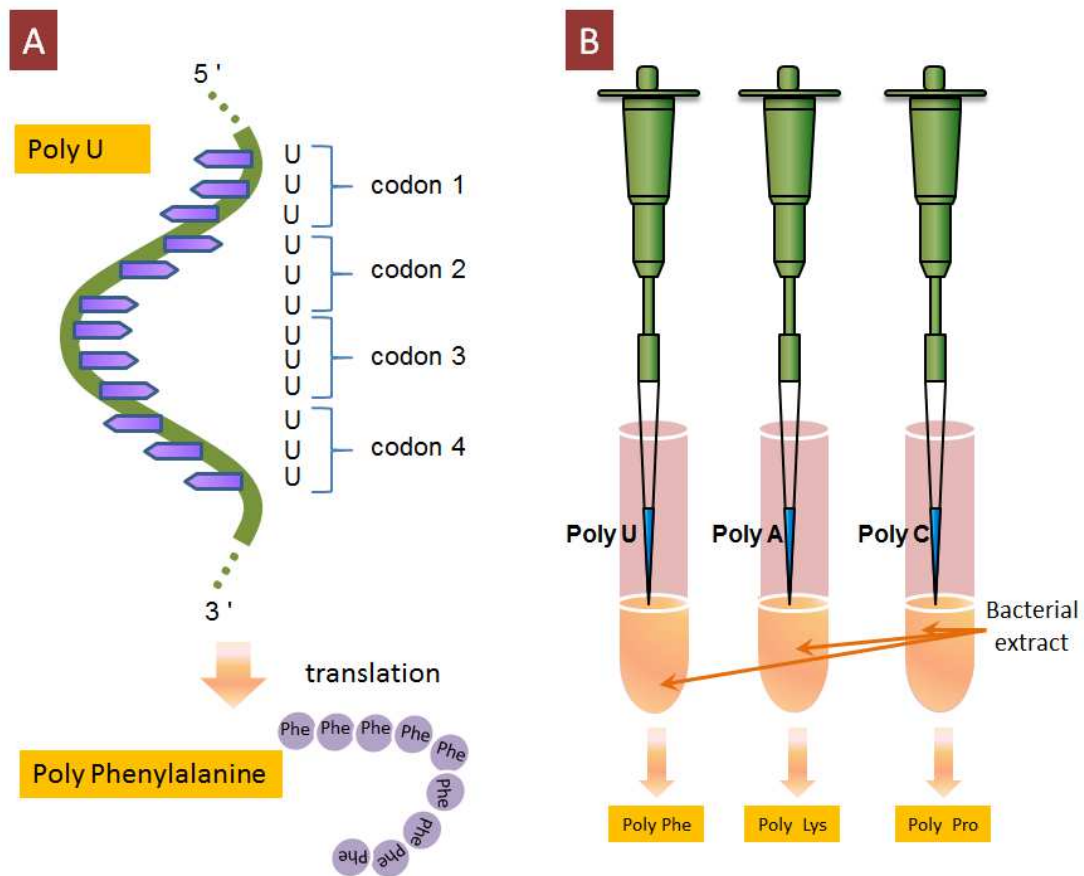


Figure (5.3). (A) Genetic code determination using laboratory synthesized mRNA through ribonucleotides of the same type which are poly (U). the addition of such synthesized mRNA into bacterial extract leads to synthesis of amino acids of the same type which is polyphenylalanine. (B) Genetic code determination using laboratory synthesized mRNA through ribonucleotides of several types, this leads to the synthesis of amino acids of different types.

Even though, Nirenberg group were unable to decode the entire genetic code, since they were unable to reach a tool by which they can synthesize a polynucleotide consists of a mixture of known sequences nucleotides. This situation remained the same until the experiments were expanded by using mixed polynucleotides and the entire genetic code was discovered (figure 5.4). This was from the share of the Indian scientist Khorana which identified the residual genetic code by his development of chemical tools that enabled him to synthesize poly-dinucleotide and poly-trinucleotides of DNA sequences can be translated into protein. For example, when Khorana and his colleges have synthesized the sequence CUCUCUC..., the polypeptide leu-ser-leu-ser... has resulted. This is not only confirms that CUC codon codes for leucine, and UCU codon codes for serine, but it confirms at the same time that the genetic code is a triple in nature. If the genetic code is a double in nature, CU CU CU CU..., or quadripartite in nature (tetra-nature), CUCU CUCU CUCU, it will produce amino acid of only one type.

Although Nirenberg and Khorana was decoded the entire genetic code, but the way by which RNA can be converted into protein was not definitely known, until Holley, after a short period of time, identified the sequence of tRNA, the adaptor molecule that facilitate translation, to get each of Nirenberg, Khorana, and Holley a Nobel prize as an appreciation for their elegant work in physiology or medicine at 1968.

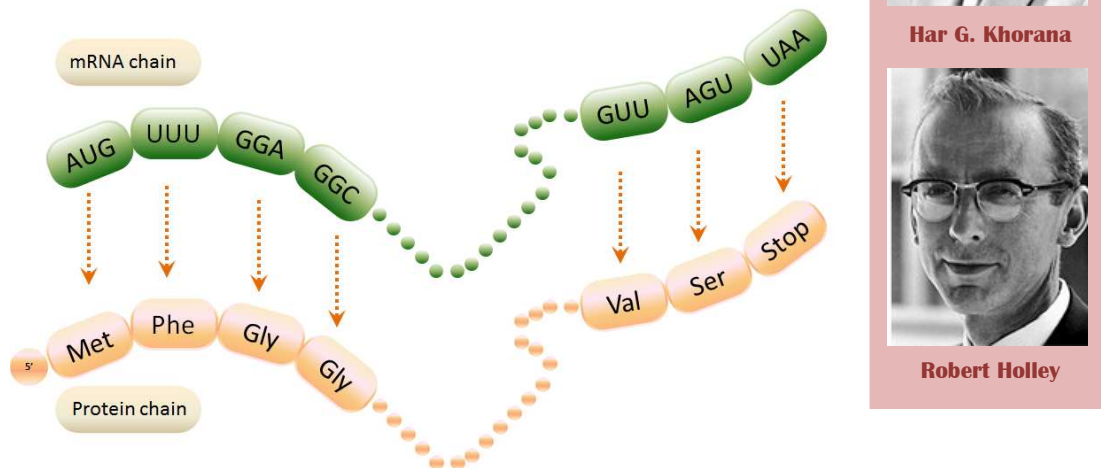


Figure (5.4). Translation of different codons in mRNA into different amino acids in protein. The word "stop" indicates stop codon.

Each Codon Consists of Three Nucleotides

Twenty types of amino acids must be provided to synthesize proteins. Thus, 20 different codons must be available to constitute the genetic code. Since only four types of nucleic acids can be found in mRNA, each codon must be consisted of more than one nucleotide. If we suppose that codons contain more than two nucleotides, they can't provide more than 4^2 specific codon, whereas if each codon contains three nucleotides it can provide 4^3 (that is 64) specific codons. When researchers found that the number of codons is 64, so, the hypothesis that says that codons are triple is the right one.

Characters of Genetic Codes

The general characters of genetic code can be summarized by the following:

1. Triple coding; as it was discussed above, the code is triple, that is it consists of three nucleotides, in which the nucleotide sequence is read as triples known as codons, but the first two nucleotides in triple codon are the most important.
2. Specificity; the genetic code is specific (unambiguous). That is, the specific code usually encodes for the same amino acid.
3. Universality; the genetic code is universal, that is, the specificity of genetic code is conserved since early stages of evolution with very slight changes in the manner of genetic code translation. The genetic code is similar in all organisms. As for now, we

realize that genetic code is almost the same in all organisms but with slight differences. Mitochondrial genome encodes from 10 to 20 proteins. Hence, some mitochondrial codons have different meanings compared with their cytoplasmic counterparts. For example, mitochondrial AUA codon encodes methionine instead of isoleucine.

4. Redundancy; the genetic code is characterized by redundancy (its described sometimes as degenerative). Most of amino acids in protein sequence are encoded by more than one codon. For example, some amino acids are represented by more than one codon, such as arginine which is encoded by six different codons, but each codon represents only one amino acid. The amino acids that encoded by one codon are not called degenerative, instead, they called synonyms which differ only in third position, the wobble position, in which the base pairing with the third position of anticodon is less stringent than first and second position of the codon. Consequently, this wobbling allows some tRNA molecules to be recognized by more than one codon.
5. Overlapping reading frames and commas; the meaning of reading frame is the nucleotide sequence from the start codon to stop codon. The spacer located that located between start and stop codon on genetic information called open reading frame (ORF). The usual reading frame is not overlapping. When we say that genetic code is non-overlapping and comma less, this means that the code reads from fixed start point as a continuous sequence of bases, three by three without any stop between codons. Actually, the reading of genetic code during protein synthesis doesn't require any overlapping in codons. Thus, the genetic code is non-overlapping. Moreover, once the reading is initiated at a specific codon, there is no punctuation between codons and the message is being decoded continuously of three nucleotides until reaching stop codon.

As it was mentioned above, the numbers of codons are 64 codons, all of these codons encode amino acids except three of them. Therefore, it was called nonsense codons (see below). At least only two of them are used as termination signals, in which the initiate signaling when polymerization of nucleic acids are about to be terminated. The rest of 61 codons encode for 20 amino acids. The list below shows that there is 64 codons regulated in 16 families, each family occupy single column between the horizontal lines (list 5.1).

For example, CCN codons, in which N can be U, C, A or G. the family located in the second column of the top second box. In some families, all four codons encode the same amino acid, as it was seen in the members of CC family that already mentioned above. These families are referred as unmixed families. These families constitute 8 out of 16 of codon families (list 5.1).

List (5.1); the mode of codons arrangement within a unified genetic code. The genetic code consists of 64 codons, 61 are encoding, and 3 are non-coding for amino acids.

First nucleotide	Second nucleotide				Third nucleotide
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U C A G
	Phe	Ser	Tyr	Cys	
	Leu	Ser	Term	Term ²	
	Leu	Ser	Term	Trp	
C	Leu	Pro	His	Arg	U C A G
	Leu	Pro	His	Arg	
	Leu	Pro	Gln	Arg	
	Leu	Pro	Gln	Arg	
A	Ile	Thr	Asn	Ser	U C A G
	Ile	Thr	Asn	Ser	
	Ile ²	Thr	Lys	Arg ²	
	Met	Thr	Lys	Arg ²	
G	Val	Ala	Asp	Gly	U C A G
	Val	Ala	Asp	Gly	
	Val	Ala	Glu	Gly	
	Val	Ala	Glu	Gly	

As for the rest two families – UG family, and AU family – they never exert any other mode and they are considered as unique. Thus, usually the third nucleotide in the codon is less important than the first and second nucleotide in determining the type of amino acid that incorporated in the protein chain. As for families that encode for more than one amino acid, they are called mixed families. In 6 of these mixed families, the pyrimidine codons, that is U or C, at third position, they encode for another amino acid, or they encode for what is called chain termination signal.

Reading Frame

As long as the sequence of mRNA molecule is decoded by groups of three nucleotides (codons) at 5' end, so this sequence can be decoded by three different reading frames depending on the identity of the nucleotide that is used first in the first codon.

Usually, one reading frame produces functional protein, while the other two frames include several stop codons. In some bacteriophages, what so called overlapping genes are exist (see chapter two) which utilize different reading frames. In addition to other viruses, bacteriophages use this style as a tool to make the most possible advantage from the work of the little viral genome. Since the genetic code is not punctual and not overlapping and triple in nature, it is theoretically possible for mRNA molecule to be translated into three reading frames. It has been appeared that some mRNA molecules have overlapping

information can be translated into different reading frames to give different polypeptides (figure 5.5).

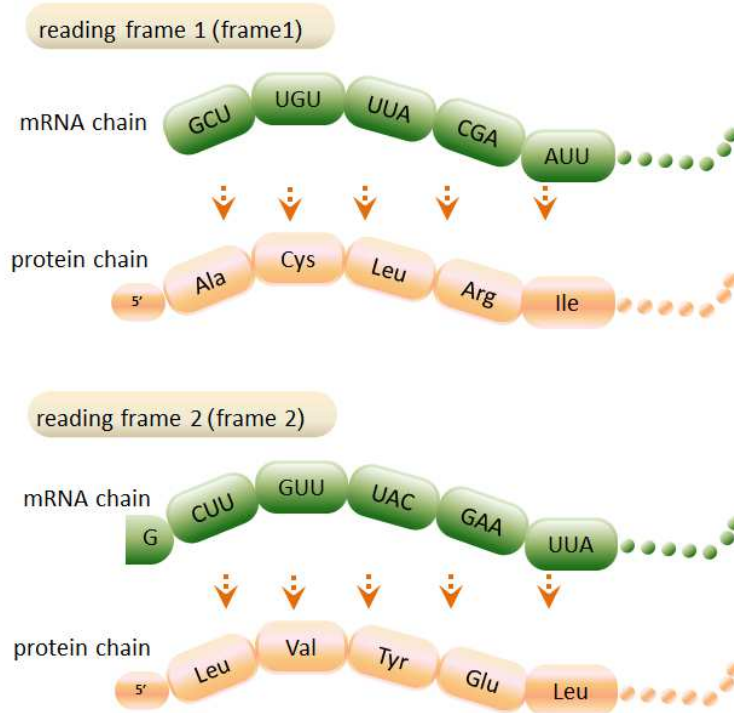


Figure (5.5); how the genetic code – the non-overlapping which is not punctual and triple – reads different reading frames. If translation of mRNA begins at different start positions (not shown in the figure), hence, it is possible for two reading frames to be exist. In this example, codons are transferred by one base to the right side of the below frame. Consequently, the same sequence encodes different amino acids during translation. Although this process is rarely happened, many overlapping cases were discovered in viral and cellular genes in prokaryotes and eukaryotes. And it is – theoretically – possible for mRNA to contain a third reading frame.

Most of mRNA molecules are decoded with one reading frame since the other two reading frames face several stop codons that terminating translation before the production of functional protein. The other unusual thing that taken place in the code arrangement caused by a change in reading frame (frame shifting). In this case, protein synthesis machinery reads four nucleotides as one amino acid and continues in reading codons in their usual triple form, or it may ignore one nucleotide base and read all sequential triple codons with a new reading frame until reaching to the end of the chain. Although this instance is not common but many examples are available which confirm its happening.

Start and Stop Codons

Translation is commenced with start codon or initiation codon. In contrast to termination codons, this codon is not capable alone to initiate protein synthesis, since it requires to some neighboring sequences and initiation factors for the translation process to begin. The sequence AUG is considered as the most common initiation codon, which encodes to methionine if it's existing elsewhere.

Stop codons are three; UAG which called *amber*, UAA which called *ochre*, and UGA which called *opal*. The UAG codon is called *amber* because of its discoverer Harris Bernstein, since his last name means "*amber*" in Germany. The second two

codons are called *ochre* and *opal* to keep the same context of nomenclature. All these three codons are called termination codons as well which give a signal to release the nascent polypeptide from ribosome. This is taken place because of the binding of releasing factors in the absence of cognate tRNA that carries an anticodon complementary to stop codons.



Protein Synthesis

Before we get into the stages of translation, we have to clarify that sequences located after transcription start point which are the sequences that give RNA copy or RNA transcript are available on three regions. The RNA copy region at which transcription begins and finish at start codon (AUG) is called "leader" or 5' untranslated region. Whereas the third region which also untranslated and extended from stop codons (UAA, UAG, and UGA) to the last nucleotide is called "trailer" or 3' untranslated region. These sequences play a role in regulation of gene expression (see chapter six). Thus, the only translatable region (in case of protein encoding) is the second region which is located between the first and third region, that is between start and stop codon. This means that no all synthesized RNA in transcription and even after processing as in eukaryotes will be translated from end to end (figure 5.6).

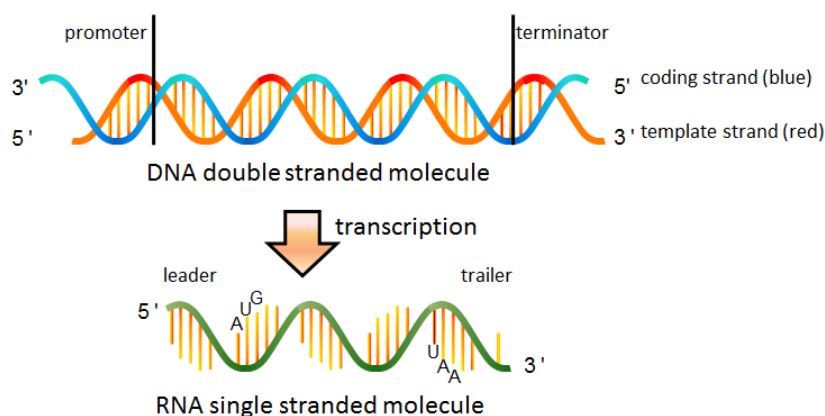


Figure (5.6); segment of prokaryotic RNA which is copied from its own DNA template region. Note that regions of promoter (primer) and terminator are on DNA and regions of leader and trailer are on RNA. Ribosomes undertakes reading on start and stop codons of protein synthesis that are shown in this figure.

In previous chapter we have presented the main molecules that participating in protein synthesis – which are mRNAs, tRNAs, and ribosomes that containing large and small rRNAs. Here, we will take a detailed look on how these components are interacted with each other to create a biochemical portrait which eventually leads to the synthesis of proteins (figure 5.7).

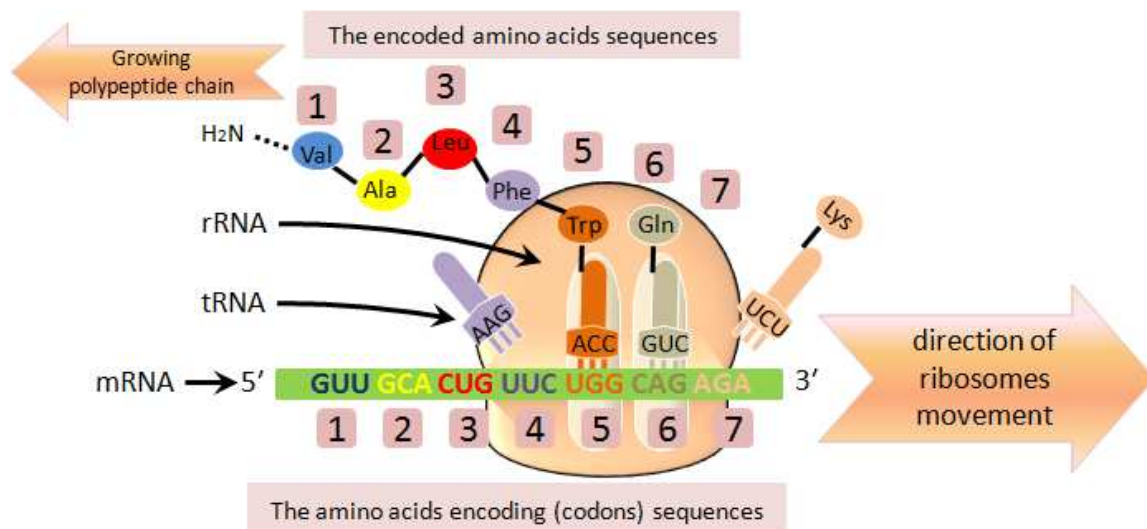
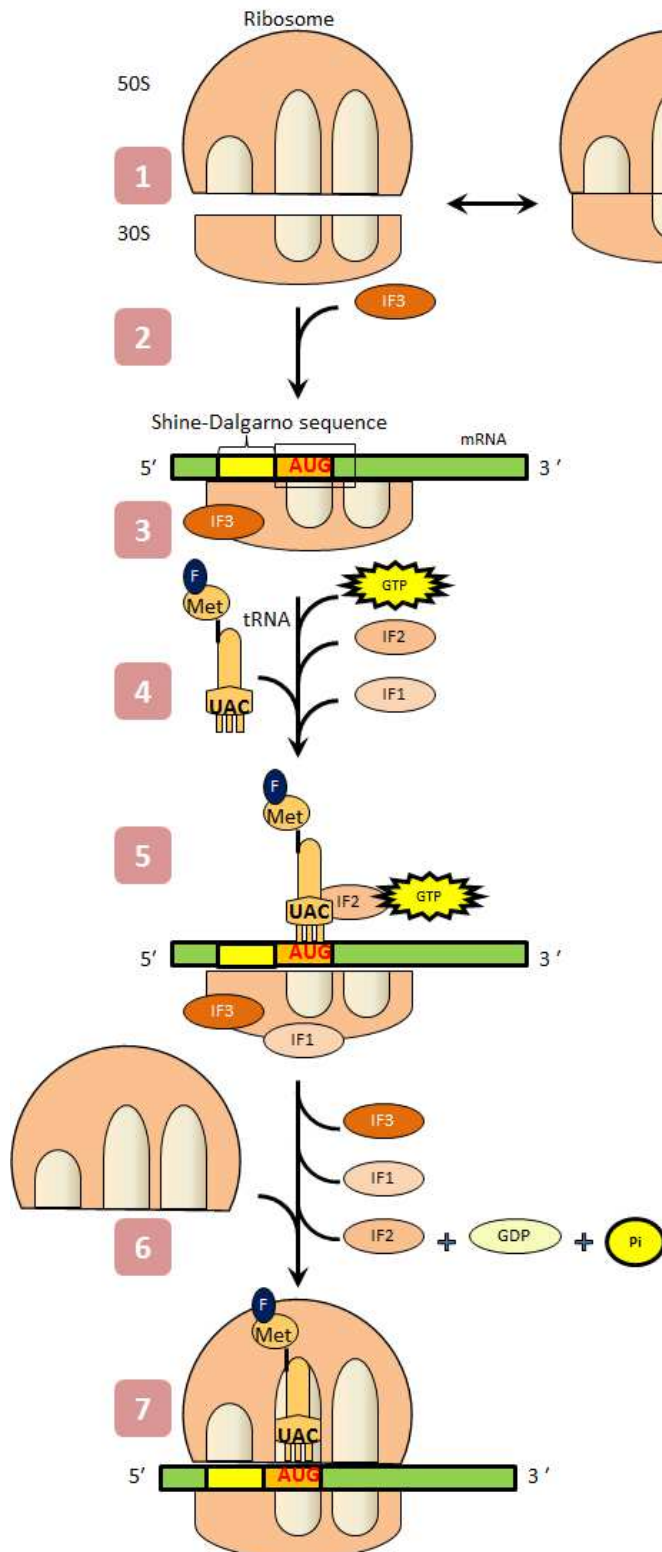


Figure (5.7); three roles achieved by RNA through protein synthesis. The mRNA is translated into protein by the help of both tRNA and ribosome, which is composed of multiple proteins and two types of rRNA molecules. Note that rRNA molecules constitute with ribosomal proteins the ribosome structure.

Ribosomes are acting like machines on which nucleic acids sequences of mRNA are translated into amino acid sequences in the corresponding protein for this mRNA as it was discussed above. Translation of mRNA is begun near 5' end, in which the message is decoded from 5' to 3'. This means the formation of N-terminus of polypeptide chain is started. This terminus is synthesized once the first amino acid is synthesized, and by which this end represents the 5' end of the protein. As for the C-terminus, it represent the end that is synthesized after synthesis of the last amino acid in polypeptide chain, and by which this end represents the 3' end of the protein. Again, the concept of polarity is appeared as well. So, protein synthesis resembles DNA replication and transcription not only in polarity but in its division into three stages which are initiation elongation, and termination as its discussed in previous two chapters.

Initiation Stage

The initiation stage of translation includes all necessary components for protein synthesis, which are composed of; (1) mRNA, (2) small and large ribosomal subunits, (3) a group of three proteins called initiation factors, (4) N-formylmethionine bounded initial tRNA which called fMet-tRNA, and (5) guanosine triphosphate (GTP). Initiation stage consists of three steps. First, mRNA bind alone with the small ribosomal subunit. Second, initial tRNA binds with mRNA through base pairing between codon and anticodon. Third, large ribosomal subunit binds with initiation complex. Let us see what is happening in each individual step with much more details.



complex.

The functional ribosome exists in the form of two subunits (see chapter four) which are 30S and 50S in bacterial cells. When both of subunits are inactive in translation, both subunits are existing in a form of dynamic equilibrium, which are binding and separating from each other (figure 5.8 a). It is impossible for mRNA to bind with small ribosomal subunit unless this subunit is being separated from the large ribosomal subunit.

Figure (5.8); initiation of protein synthesis in bacterial cells. (1) The ribosomal subunits exist in dynamic equilibrium, constantly joining and separating, (2) IF-3 binds to the small subunit of the ribosome, preventing the large subunit from binding, (3) which allows the small subunit of the ribosome to attach to mRNA, (4) A tRNA charged with *N*-formylmethionine forms a complex with IF-2 and GTP, (5) and joins the small subunit of the ribosome and the mRNA, (6) IF-1, IF-2, and IF-3 dissociate from the complex, GTP is hydrolyzed to GDP, (7) and the large subunit joins to create a 70S initiation

The initiation factor 3 (IF-3) binds with small ribosomal subunit and prevent the large subunit from binding during initiation stage (figure 8.5 b). The ribosome binds with Shine-Dalgarno sequence as it shown above that is complementary to 16S rRNA which constitutes a part of ribosomal subunit. This permit the small subunit to bind with mRNA in which ribosome is localized directly above initiation codon. After this, fMet-tRNA binds with initiation codon (figure 8.5 c). This step requires to initiation factor 2, which form a complex with GTP.

The third factor which is IF-1 enhances the dissociation of large and small ribosomal subunits from each other. At this point, initiation complex consist of (1) small ribosomal subunit, (2) mRNA, (3) initial tRNA with its methionine (fMet-tRNA), (4) GTP, and initiation factors 1, 2 and 3. These components are known altogether as 30S initiation complex (figure 5.8 c).

In the last step of initiation, IF-3 is dissociated from ribosomal subunit and this allows large ribosomal subunit to join to initiation complex. The source of energy; GTP (which is supplied from IF-2) is hydrolyzed into GDP, then both IF-1 and IF-2 are departed (figure 8.5 d). When the large ribosomal subunit binds with initiation complex, the later is converted from "30S" into "70S" initiation complex.

The differences between Initiation of Translation in Eukaryotes and Prokaryotes

Although the general context of initiation of translation in prokaryotes is similar to that found in eukaryotes, there is important differences must be noticed between both of them.

The most important differences are the absence of Shine Dalgarno sequence in prokaryotes and absence of its corresponding sequence in eukaryotes. Many prokaryotic mRNA are polycistronic, that is, mRNA molecule encodes to more than one polypeptide chain (as it was discussed in chapter two). The polycistronic mRNA should contain several start codons. However, the polypeptide must contain more than one methionine not exist in the beginning (non-initiating methionine). In typical genetic code, the codon of initiating methionine and non-initiating methionine is AUG. To recognize between them, prokaryotes use a specialized sequence located 5 to 10 bps upstream of the start codon AUG, in an untranslated region (UTR) that located 5' of translatable region of mRNA. This specialized sequence called Shine-Dalgarno according to its discoverer. This purine rich sequence is complementary to the core sequence UCCU of 3' end of 16S rRNA (that located in 30S ribosomal subunit). The activity of ribosome binding site (RBS) which is belong to one of the two portions of this sequence is affected with the length and nucleotide sequence of the other portion, which is the spacer portion that separate between ribosome binding site and the start codon AUG.

It is so obvious that Shine-Dalgarno sequence is acting on mRNA arrangement on ribosome in the initiation of translation using base pairing with the complementary sequence near 3' of rRNA (figure 5.9). This interaction by base pairing enables bacterial ribosomes to initiate protein synthesis near 5' end of mRNA. Conversely, in eukaryotes, ribosomes recognize mRNA by the binding with its cap (which is composed of 7-methylguanosine) at 5' end of mRNA (figure 5.9). Then, ribosome start precise scanning of mRNA starting from 5' end until it reaches to AUG sequence which here represents the initiation codon.

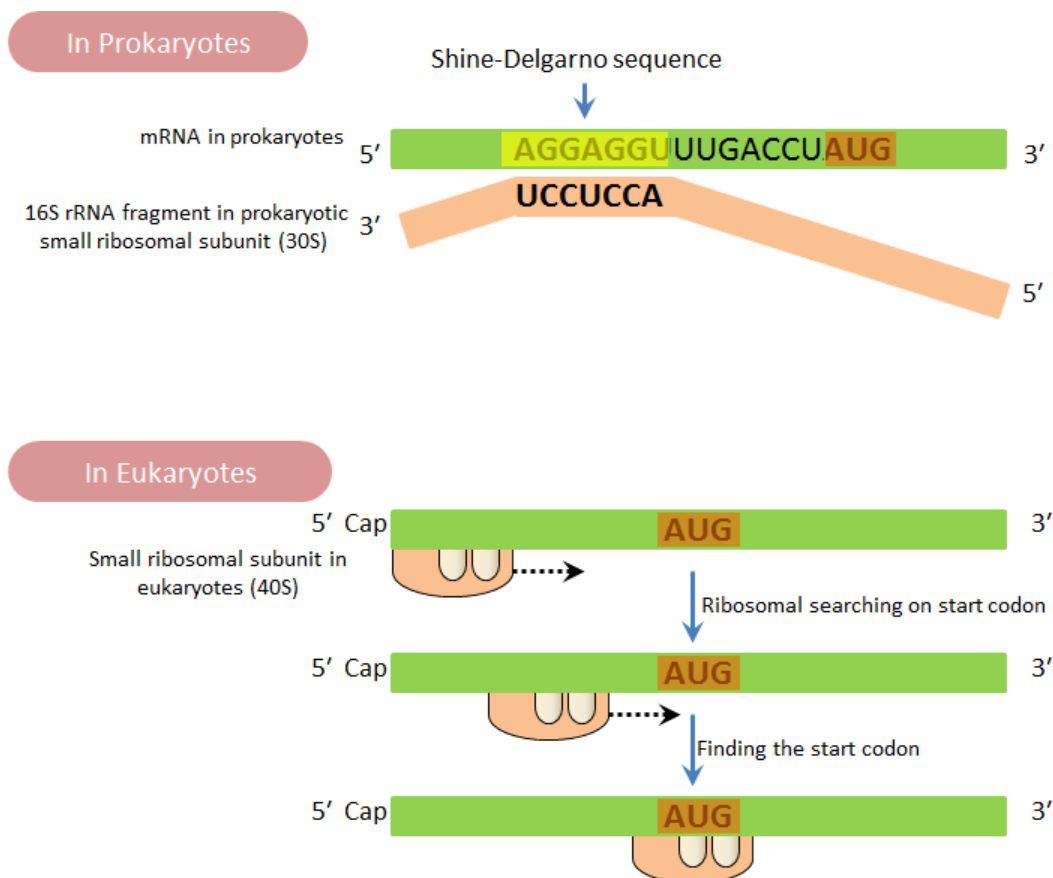
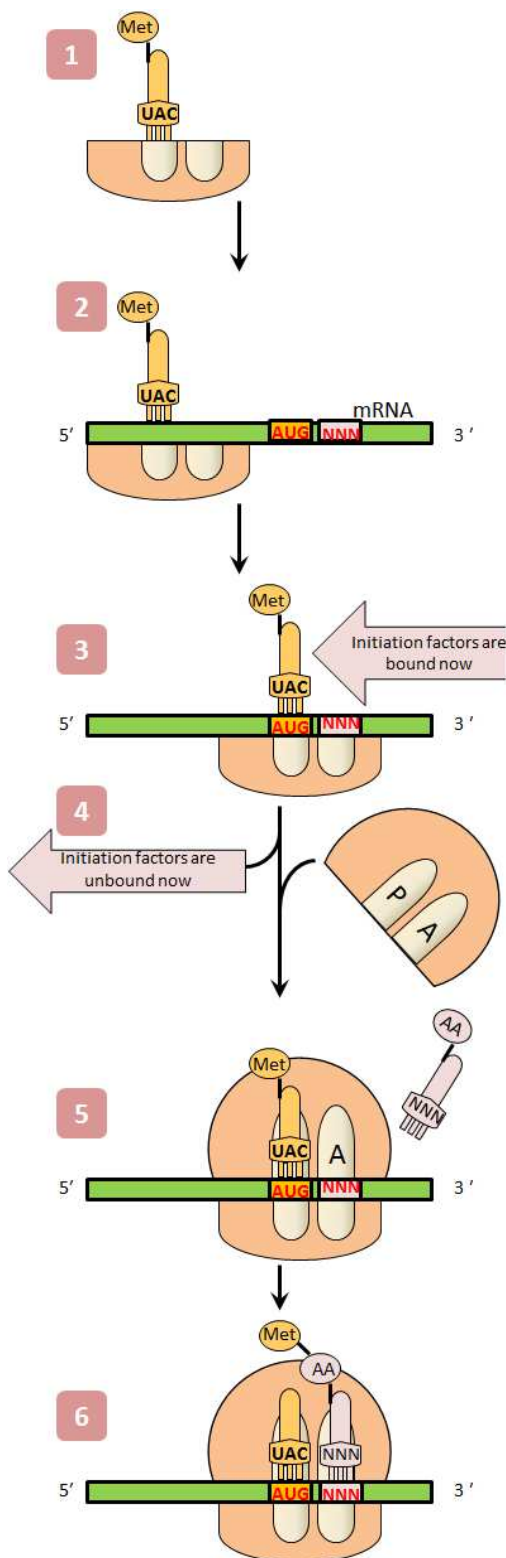


Figure (5.9); the mode of initiation codon recognition in prokaryotic and eukaryotic ribosomes.

The mechanism used by eukaryotic organisms to recognize the start codon AUG is not clear entirely, but some researchers suggest that eukaryotic ribosomes simply scan mRNA molecules that bounded to them from the 5' cap and recognize the first AUG codon once they find it and consider it as a site for initiation of translation. This mechanism is described as possible, since almost all eukaryotic mRNA are monocistronic (which encode for single polypeptide). Initiation codon recognition is facilitated by a consensus sequence called Kozak sequence which confine initiation codon, and it's composed of about 13 base pairs (GCCGCCACCAUGG), that most of it is localized within the 5' untranslated region to bind and initiate translation.



Other important difference is represented by the requirements of translation initiation of eukaryotes into more initiation factors which are at least 10 different factors each of which plays a distinguishing role in facilitating initiation of translation compared with only 3 initiation factors that exist in bacterial cells. While the third difference is correlated with the addition of formyl group to the first methionine. In most bacteria, translation is beginning with formyl-modified methionine, while in mammals (figure 5.10); methionine is used in its unmodified form (except mitochondria and chloroplast, in which their ribosomes resemble that found in bacteria).

Figure (5.10); initiation of translation in eukaryotes. (1) binding of the small ribosomal subunit with initiation factors, (2) binding of mRNA, (3) the scanning permits codon recognition by binding with the initial tRNA, (4) dissociation of initiation factors and the binding of large ribosomal subunit with its small partner – the letter P indicates to peptide site and the letter A refers to amino site, (5) binding of aminoacyl tRNA that is indicated as AA, (6) formation of the first peptide bond.

In prokaryotes, the addition of formyl to methionine in tRNA appears to mislead peptide site (P-site) in ribosome since the addition of formyl to methionine make it appears like a peptide bond. Since the peptide bond accepts peptide bonds, so, N-formylmethionyl – tRNA stabilizes in the peptide bond. This process is taken place off course for only one instance during protein synthesis.

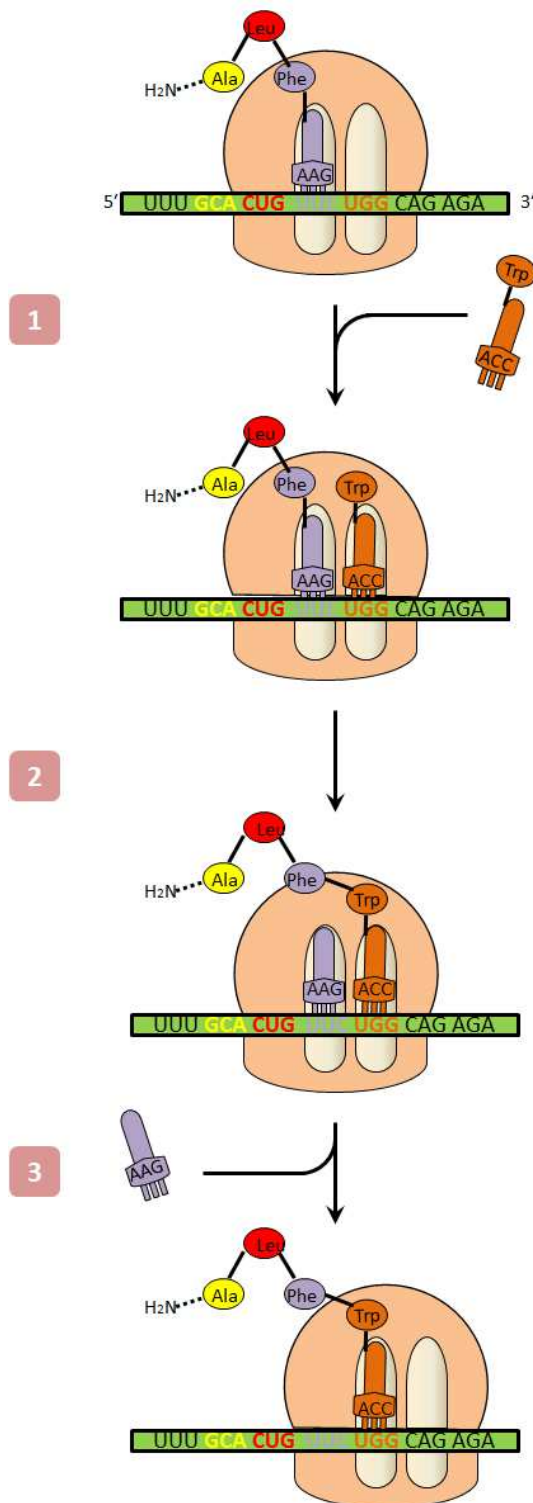
The sequences that surrounding AUG codon affect on the efficiency of this codon in this methionine encoding location. Therefore, and in many instances, the first AUG codon in mRNA is ignored, in which this codon is a codon gives permission to initiation stage only without expressing on any amino acid, while if this codon is located elsewhere it will be decoded as the rest 61 codons. Therefore, it will encode for methionine.

Elongation Step

Before initiation complex is formed, translation is preceded by elongating polypeptide chain. The mechanism of elongation in prokaryotic and eukaryotic cells is very similar. Ribosome has three sites for binding with tRNA, which are peptidyl site, abbreviated P-site, aminoacyl site, abbreviated A-site, and exit site, abbreviated E-site. The initiator methionyl tRNA binds with P-site. Remember that initiator tRNA that carries methionine is bound at this time with P-site and base paired with the start codon AUG. If the anticodon of incoming (second) aminoacyl-tRNA corresponds that codon in mRNA correctly, a strong binding at A-site is initiated. If the incoming aminoacyl-tRNA doesn't correspond this codon in mRNA, it will be dissociated. With the existence of the first methioninyl-tRNA at P-site and the incoming aminoacyl-tRNA that bound firmly with A-site, α amino group of the second amino acid will react with the activated methionine in the first tRNA, forming a peptide bond. The revolution that taken place in biochemistry has affirmed that this reaction is activated by an enzyme called peptidyl transferase in eukaryotes, which affect on the transfer of the growing peptide bond in tRNA that located in P-site into the activated amino acid in aminoacyl-tRNA that located in A-site. The elongation of polypeptide bond on ribosome is a cycle composed of three characterized steps (figure 5.11).

Step No. 1; aminoacyl-tRNA binds in the empty A-site (adjacent to occupied P-site) by formation of base pairs with the three mRNA nucleotides (codon) that exposed at that A-site.

Step No. 2; after the binding of carboxylic end of polypeptide chain of tRNA in P-site is finished, it will be attached by peptide bond of amino acid bound with tRNA in A-site. Since α amino group in the new aminoacyl-tRNA attack (nucleophilic attack) carboxyl group of peptidyl-tRNA that occupy P-site. This essential reaction of protein synthesis is catalyzed by peptidyl transferase, which is a component of a specialized region of rRNA in the small ribosomal subunit. The enzymatic activity of this region is one of live examples of ribozyme catalytic activity that some RNA molecules enjoy (see appendix of chapter four). The reaction result in the binding of the growing polypeptide chain with the tRNA at A-site.



Step No. 3; the new peptidyl tRNA is transferred from A-site to P-site by movement of ribosome exactly three nucleotides along mRNA. This step needs energy which leads to inaugurating a series of conformational changes in one of the ribosomal subunits through GTP hydrolysis. Add to that, the role of elongation factor should not be ignored in this process. After removing peptidyl group from tRNA in P-site, the removed tRNA is dissociated quickly from P-site. In bacteria, the empty (uncharged) tRNA leave ribosome through another site (E-site), while in eukaryotes it will be expelled directly into cytoplasm. After finishing step no. 3, the empty A-site will be free to accept new aminoacyl-tRNA, and then the cycle starts again. In bacteria, each cycle requires about to 20 parts of second under the optimum conditions. Therefore, the entire synthesis of a protein made of 400 amino acids will be completed in about 20 seconds. Ribosomes moves along mRNA in 5' to 3' direction, which is the same direction of protein synthesis as it was discussed above (review figure 5.6). the newly synthesized peptidyl-tRNA is transported with its corresponding codon into P-site in order to made A-site ready to accept new cycle of aminoacyl-tRNA molecules. From this discussion, it was obvious that the flowing of tRNA is done through A-site, through P-site, into exit through E-site (in prokaryotes). The figure (5.12) compares between the movement of tRNA and mRNA even the clear differences in their behavior but they have the same direction of movement.

Figure (5.11); a sum up of elongation of protein synthesis in eukaryotes

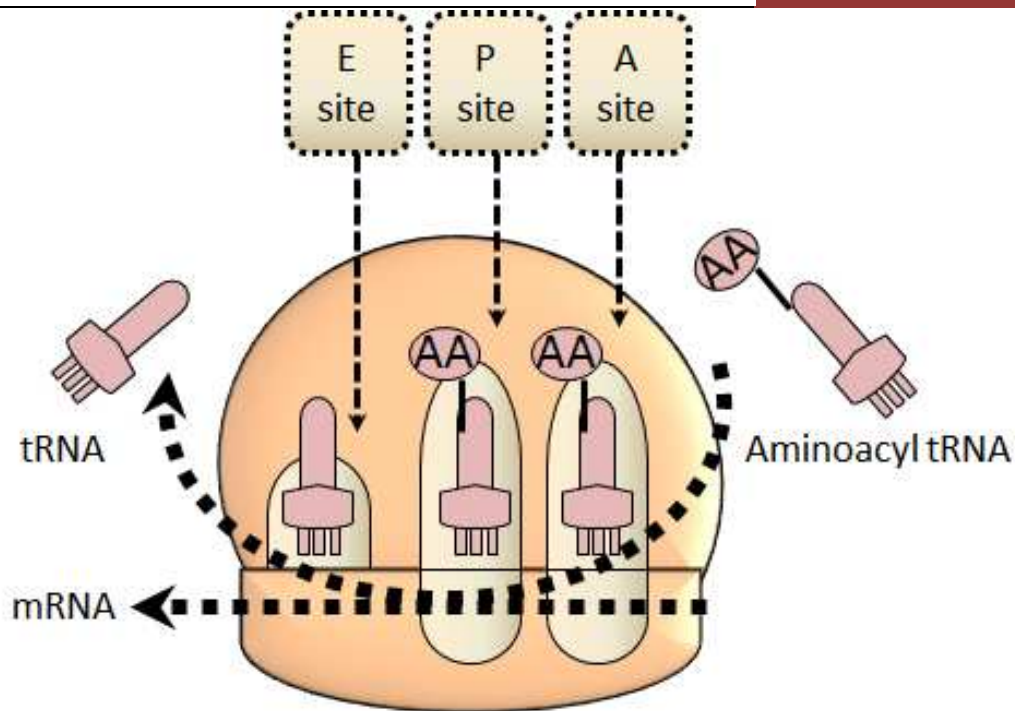


Figure (5.12); movement of tRNA and mRNA in the same direction through ribosome.

Termination Stage

When the polypeptide chain gets the stop codon of mRNA, it falls off and releases the created protein. This is called termination, and it's important to know that during this process many enzymes either help or facilitate the entire process. After several elongation cycles that polymerizes amino acids into proteins, stop codon appears at A-site. Usually, there is no tRNA has anticodon capable to recognize termination signal. But the so called releasing factors are capable to recognize termination signal that located in A-site (figure 5.13). The releasing factor, with the binding of GTP and peptidyl transferase, stimulates hydrolysis of the bond between the polypeptide and tRNA that still occupies P-site. This hydrolysis releases the polypeptide and tRNA from this location.

After hydrolysis and releasing, the ribosome dissociated into its separated large and small subunits, which will be reused again in later protein synthesis events. Thus, releasing factors are proteins that hydrolyze peptidyl – tRNA bond when stop codon occupies A-site.

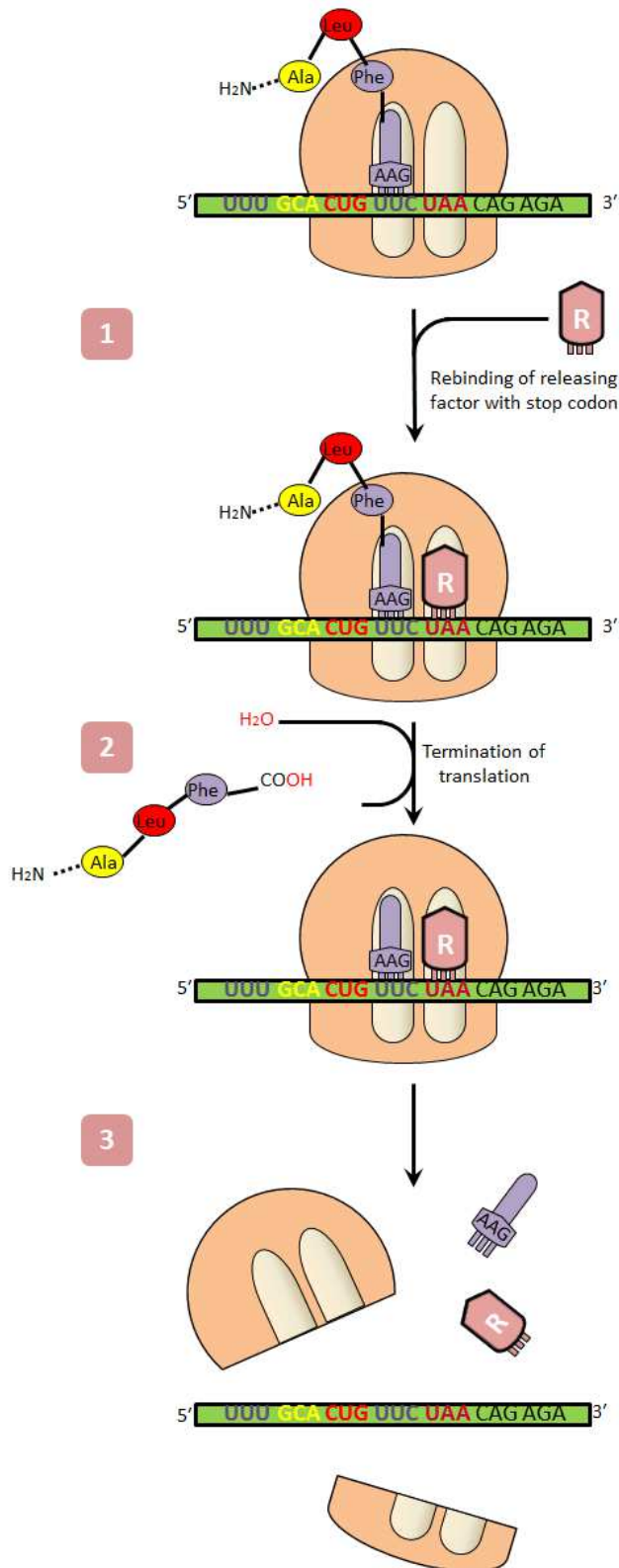


Figure (5.13); termination stage of protein synthesis in prokaryotes and eukaryotes. (1) The binding of releasing factor with stop codon stops translation, (2) the peptide that has been synthesized is released, (3) ribosomes dissociated into its separated units.

Ribosomes Roles in Protein Synthesis

Many ribosomes can translate the same mRNA simultaneously. But, because of their relatively large size, ribosome molecules are unable to bind with mRNA in a distance less than 80 nucleotides apart (that is, the smallest distance between two ribosomes that acting on the same mRNA is 80 nucleotides). When ribosomes are working together on the same mRNA, they create what is called polyribosomes, or polysomes, which their numbers are increased proportionally with the increasing of mRNA length (figure 5.14).

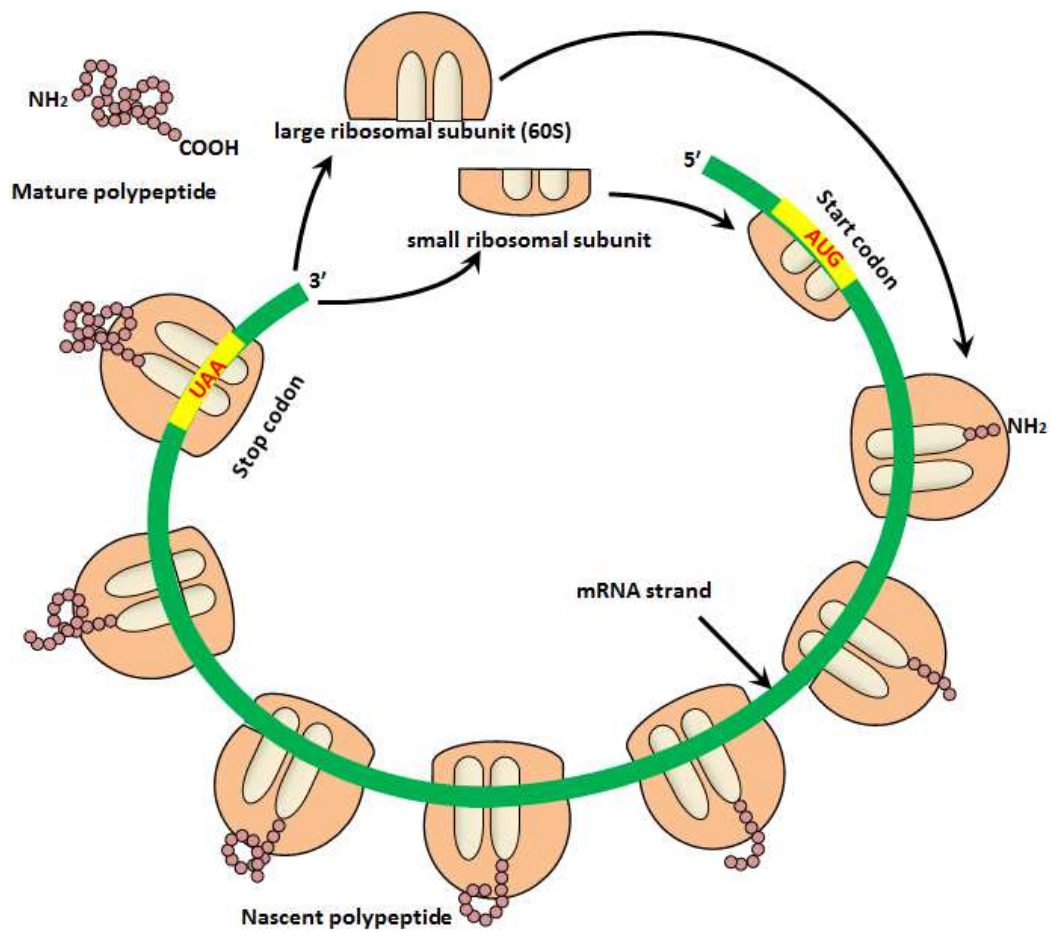


Figure (5.14); polyribosome complex. This illustration shows how for a chain like ribosomes to stimulate translation of the same mRNA simultaneously.

The single ribosome in mammals is able to synthesize about 100 peptide bonds each minute. Polysomes, which actively synthesize proteins, can be located as free particles in the cytoplasm, or they may bind with cytoplasmic membranous material folds called endoplasmic reticulum. The binding of ribosomes with endoplasmic reticulum is responsible for the "rough" appearance that some regions of this reticulum characterize under electron microscopy. The manufactured protein from polysomes that bound on endoplasmic reticulum is emerged into an open space called cisternae space, which is located between endoplasmic reticulum and Golgi apparatus, then, into zymogen particles in order to be eventually exported. The word "zymogen" is used to describe protein precursors, the primary proteins that are converted into mature proteins by several means.

Protein Synthesis has Proofreading

It was noticed the presence of error rates in protein synthesis approximate to one amino acid per 10000 polymerized amino acids, this means that one protein molecule of 400 amino acid average must contain this error. Therefore, cells have developed proofreading mechanism to reduce number of mistakes in these basic steps in protein synthesis.

Cell used two different mechanisms of repair; both of them consume energy, since it's a cost that should be paid for any increasing that taken place in protein activity of the cell. A relatively simple mechanism is used to enhance the efficiency of amino acid binding to tRNA. This is clarified from the presence of two active sites for aminoacyl tRNA synthetase instead of only one site. While the first site loads the amino acid, the other site recognizes the mistakenly bound amino acid to tRNA and displays it through hydrolysis. This is a costly step in respect to energy as in the proofreading that used in DNA replication (see chapter three). The other mechanism used to enhance codon – anticodon pairing is represented by the formation of an available protein in the cell called elongation factor (EF) once tRNA molecules acquire amino acid. This protein binds tightly with both amino acid and tRNA that pairs with the suitable codon in mRNA.

The bound elongation factor allows codon – anticodon pairing but it prevents amino acid incorporation in the growing polypeptide chain. The recognition of primary codon, however, leads to hydrolysis of GTP (into GDP and inorganic phosphate), whereas elongation factor is dissociated from ribosome without its bound tRNA allowing protein synthesis to proceed. Thus, elongation factor causes a short delay in base pairing between codon and anticodon and in polypeptide elongation. This leads to a creation of an opportunity for tRNA to get out of ribosome. This step might be responsible about the delay of translation (about 40 amino acids per second) compared with replication (1000 nucleotides or more per second).

The incorrect tRNA makes a lower number of hydrogen bonds between codon and anticodon compared with the correct tRNA; therefore it has weaker binding with ribosome and it's possible to be dissociated during this stage. Because of the delay that caused by elongation factor tRNA molecules departure their incorrectly bound ribosomes without being used by protein synthesis, in which this factor increases of correct amino acids ratio to their incorrect incorporated counterparts in protein (figure 5.15).

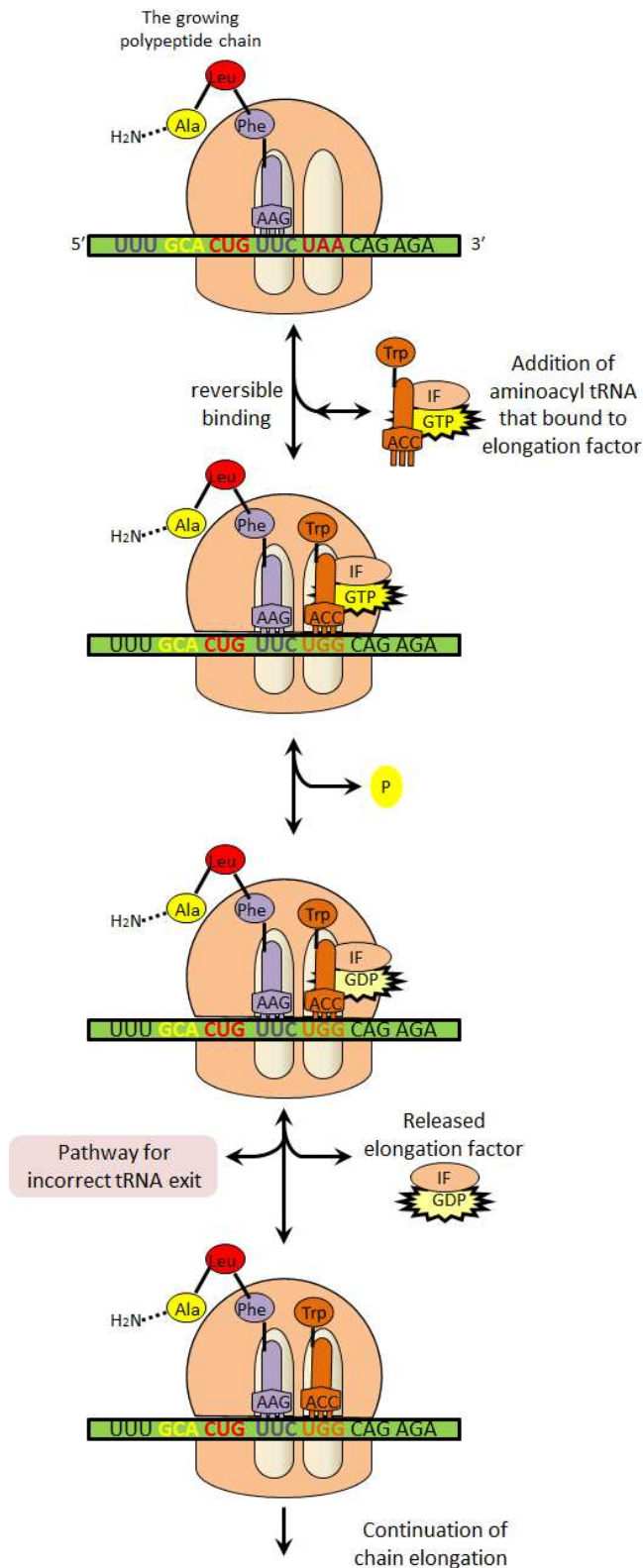
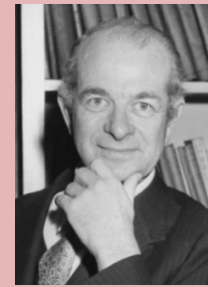


Figure (5.15); the proofreading selects for the correct RNA on ribosome during protein synthesis. This more detailed view of step 1 of the elongation phase of protein synthesis shows how, in the initial binding event, an aminoacyl-tRNA molecule that is tightly bound to an elongation factor pairs transiently with the codon at the A-site. This pairing triggers GTP hydrolysis by the elongation factor, enabling the factor to dissociate from the aminoacyl-tRNA molecule, which can now participate in chain elongation (see Figure 6-22). A delay between aminoacyl tRNA binding and its availability for protein synthesis is thereby inserted into the protein synthesis mechanism. As a result, only those tRNAs with the correct anticodon are likely to remain paired to the mRNA long enough to be added to the growing polypeptide chain. The elongation factor, which is an abundant protein, is called EF-Tu in prokaryotes and EF-1 in eukaryotes.

Protein Structure

Proteins are considered three-dimensional molecules which achieve most of the cellular functions. While DNA determines the genotype, the phenotype of the cell is the result of protein action. Each amino acid contains a unique functional group (which also called R group) which establishes its biochemical activity. With the proceeding of translation, the active groups of polypeptide amino acids start to interact with each others. Once this happens, the protein starts to fold into its three dimensional conformation. There are four levels of protein structure (figure 5.16).

The primary level of structure in a protein is the linear sequence of amino acids as joined together by peptide bonds. This sequence is determined by the sequence of nucleotide bases in mRNA. This is the information included within the gene exon (review chapter two). The single gene may produce polypeptides with different primary structures. Most proteins are unable to stay in their primary structure since the functional groups of their polypeptides chain that formed during translation start to be interacted forming secondary and tertiary structures. The secondary structure refers to the stable arrangements particularly of different amino acids which give repeated structural patterns. Thus, the second level of series of folding within polypeptide chain. These patterns are a consequence of hydrogen bonds between active groups of certain amino acids. Specific patterns of amino acids result in formation of the helical structure that called α -helix that discovered by Linus Pauling. This structure don't resemble DNA double helix, instead, it characterizes with a unique helical pattern differ from that noticed in the DNA double helix.



Linus Pauling

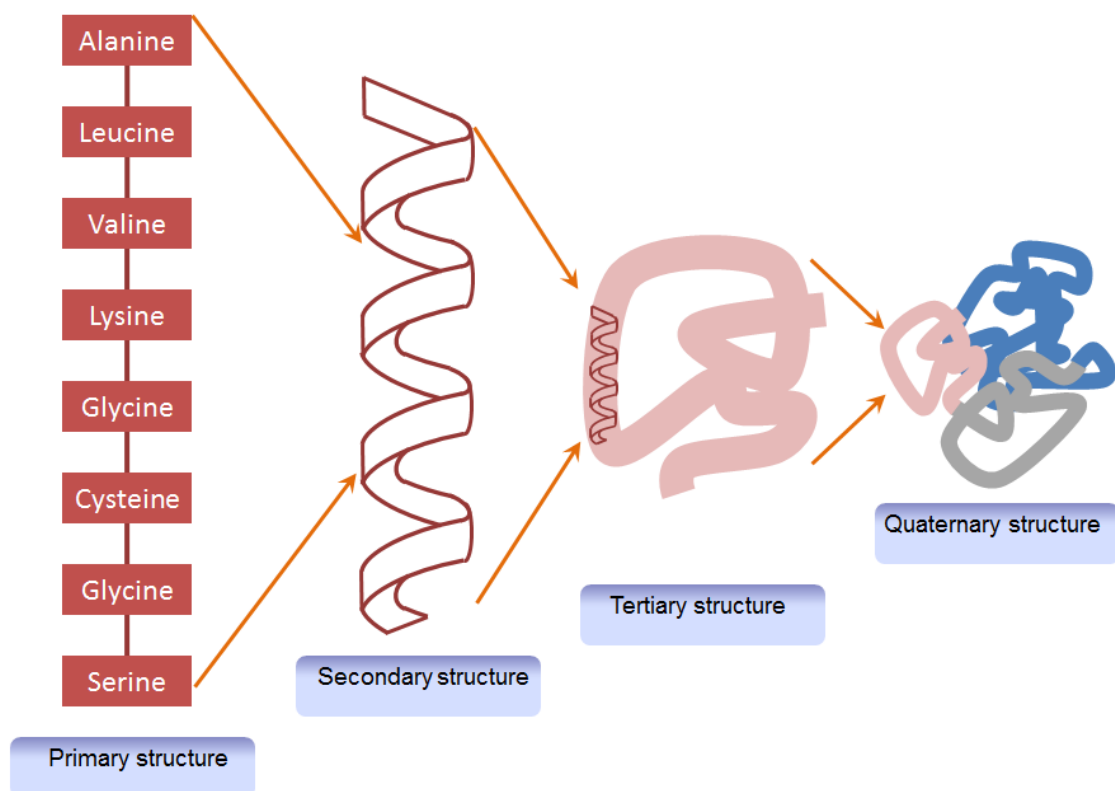


Figure (5.16); four levels of structures in protein, arranged from left (primary structure) to right (quaternary structure), in the left of this figure the only linking bond in primary structure is restricted to peptide bonds between amino acids, then these bonds are varied and complicated until they reach to quaternary structure.

There are other groups of amino acids can be resemble the zigzag plate in the polypeptide. This form called β -sheet. It's possible for polypeptide to make both appearances; α -helix and β -sheet in the same protein molecule, and not only once, but multiple times for each appearance. The tertiary structure describes all appearances of three dimensional folding in polypeptide. It's possible for secondary and tertiary structure to be formed simultaneously in polypeptide. By forming of tertiary structure, polypeptide makes a three dimensional complex which determine protein function in the cell. Many proteins perform their functions at this level. One the protein contain more than one polypeptide subunit, there arrangement in space is called quaternary structure (polypeptide chains in quaternary protein are called subunits). The binding of these multiple polypeptides or subunits is called the quaternary level of protein structure (figure 5.16). Not all proteins exploit the quaternary structure, but when some of these proteins do so, they create very big structures. However, proteins after their synthesis and folding are exposed into certain modifications to add the final touch on them in order to achieve their scheduled role inside or outside the cell. These modifications called post-translational modifications.

Post-Translational Modifications

The meaning of post-translational control can be identified as the mechanisms by which it's possible to process protein after its translation. This processing may be done by modification of the three dimensional structure of protein or by modification of amino acids side chains or by modifying protein backbone itself (see below). Because the protein is consist of a series of amino acids which their number are twenty, so, both arrangement and identity of these amino acids are important in determining the role played by protein in the cell. These changes can change of structure and protein function, or they can determine the protein age itself.

First; Three-Dimensional Structure Modifications

Translation ends the flow of genetic information within the cell. The sequence of nucleotides in DNA is converted into sequence of amino acids in protein. The synthesis of polypeptides, however, is not enough to produce active protein. To make the translation resulting polypeptide useful, it's necessary for to be folded into a distinct three dimensional conformation. Therefore, many proteins suffer from other modifications, which are necessary for function and position of these proteins in the cell. The classical example of proteins folding is that all necessary information for protein t take its correct three dimensional structure is provided from its amino acids sequence. This has been proved by Christian Anfinsen experiments which were taken place in RNase and got Nobel Prize in chemistry at 1972 as a result of this discovery. He discovered that denatured RNase can be renatured spontaneously *in vitro* to return its active conformation (figure 5.17).



Christian Anfinsen

The discovery of RNase ability to return its natural folding state (spontaneous renaturation) without presence of additional cellular factors, made scientists thought that what was happened to RNase could be applied to other proteins. But most of modern experiments have shown that the correct folding of proteins within the cells require mediation of other protein activities. In bacteria, certain proteins (see below) called chaperons are participate in the folding of 85% of proteins, whereas in prokaryotes, a higher percent of these proteins depend largely on chaperons in their folding.

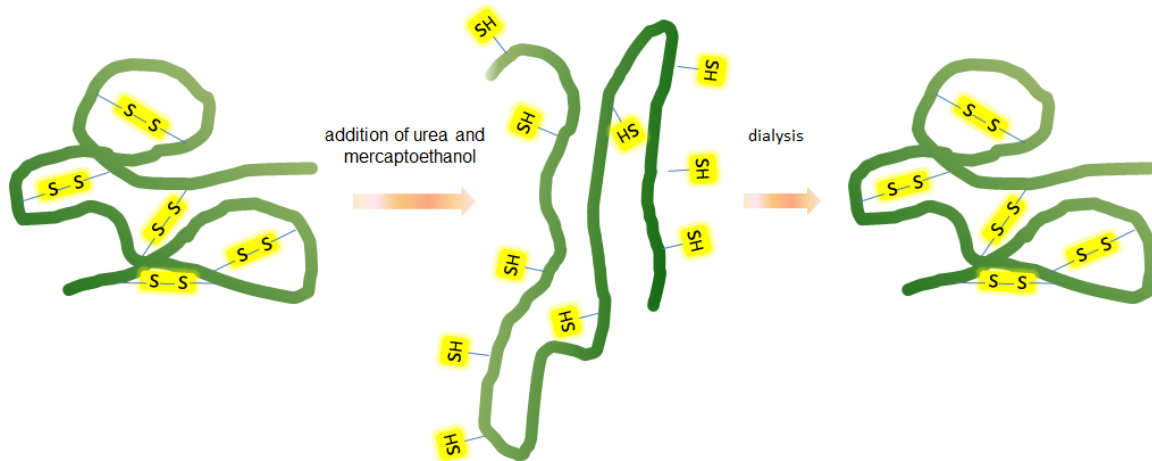


Figure (5.17); spontaneous renaturation in RNase after its denaturation.

Proteins that facilitate folding of other proteins are called molecular chaperons. The idiom "chaperons" is used for the first time to describe the nucleoplasmin that is necessary for nucleosomes assembly from histone and DNA (see chapter two). Nucleoplasmin binds histones and help its assembly into nucleosomes, but the structure of nucleosomes itself is not incorporated in the final nucleosomes. Thus, molecular chaperons facilitate the assembly without being part of the assembled complex. Later experiments showed that molecular chaperons have a fundamental role in facilitating folding of proteins. Chaperons are located in each cellular chamber, and bind with a vast range of proteins, which act of the general folding mechanism of proteins. Researchers have distinguished two fundamental families for chaperons;

1. Molecular chaperones, which bind and stabilize the unfolded or partially folded proteins, thus, they prevent the accumulation and destruction of these proteins.
2. Chaperonins, which directly facilitate the folding of proteins.

Molecular chaperones composed of heat shock proteins (Hsp) and their homologs such as Hsp70 in cytosol and mitochondrial matrix, BiP in endoplasmic reticulum, and DnaK in bacteria. These proteins have been identified for the first time because rapid appearance in cells stressed by elevated temperatures (hence, the name *heat shock*

proteins). The Hsp70 and his homologs of other proteins are the main chaperones in all organisms.

When bound to ATP, Hsp70-like proteins assume an open shape in which an exposed hydrophobic pocket transiently binds to exposed hydrophobic regions of the unfolded target protein. Hydrolysis of the bound ATP causes molecular chaperones to assume a closed form in which a target protein can undergo folding. The mode of molecular chaperones action is to prevent the formation of incorrect structures instead of stimulating the formation of correct structures. For example, prokaryotic molecular chaperones bind with nascent polypeptide chain that still being translated by ribosomes, thus, they prevent incorrect folding of polypeptide chain from the beginning of their synthesis. The exchange of ATP for ADP plays an important role in releasing of target protein. This cycle is speeded by certain protein known as co-chaperone Hsp40 in eukaryotes which accompanying to Hsp70 work, and co-chaperone DnaJ in bacteria which accompanying to DnaK work, since it was thought that molecular chaperones bind with all nascent polypeptide chains that still being synthesized in ribosomes.

The proper folding of a large variety of newly synthesized or translocated proteins also requires the assistance of chaperonins. These assemblies are formed from two rings of oligomers. The eukaryotic chaperonin *TriC* consists of eight subunits per ring. In the bacterial, mitochondrial, and chloroplast chaperonin, known as *GroEL*, each ring contains seven identical subunits. The GroEL folding mechanism, which is better understood than TriC-mediated folding, serves as a general model. In bacteria, a partly folded or misfolded polypeptide is inserted into the cavity of GroEL, where it binds to the inner wall and folds into its native conformation. In an ATP-dependent step, GroEL undergoes a conformational change and releases the folded protein, a process assisted by a co-chaperonin, GroES, which caps the ends of GroEL. The GroES forms dome like structure which bind with only one surface of the paired ring, thus it covers the ring opening. We can differentiate (as in figure 5.18) between the GroEL two rings, where we call the first the proximal ring (GroES bound) and second ring the distal ring (GroES unbound).

The reason behind calling GroEL with chaperonin, and GroES with co-chaperonin, because GroEL plays an important role in leading of folding process, as for GroES, it only helps GroEL in its action. The role of Hsp70 is to Hsp70 may assist a nascent protein in folding, but then passes it on to GroEL for the process to be completed. Then, the substrate that represented by the nascent protein binds with the distal ring of GroEL causing increasing of the volume of the central cavity. When GroES close the opening of GroEL, it causes change in GroEL proximal ring conformation, increases its central cavity, while, in the same time, the volume of the central cavity is reduced in the distal ring. The energy that drives this reaction is ATP hydrolysis. Therefore, after completing of protein processing inside GroEL/GroES complex, the affinity of GroEL to GroES is reduced because of ATP hydrolysis, this in turn leads to the releasing of the processed protein. Moreover, ATP hydrolysis of the distal ring is necessary for expelling of the processed protein by this complex. This step is taken place after completing protein processing. But if the protein is not enough processed, this complex is stimulated, and bound with it again until this protein get into its mature conformation. Hence, GroEL/GroES complex don't find a substrate to

act on, to release protein from the domination of GroEL/GroES complex, then it will go to do its determined function.

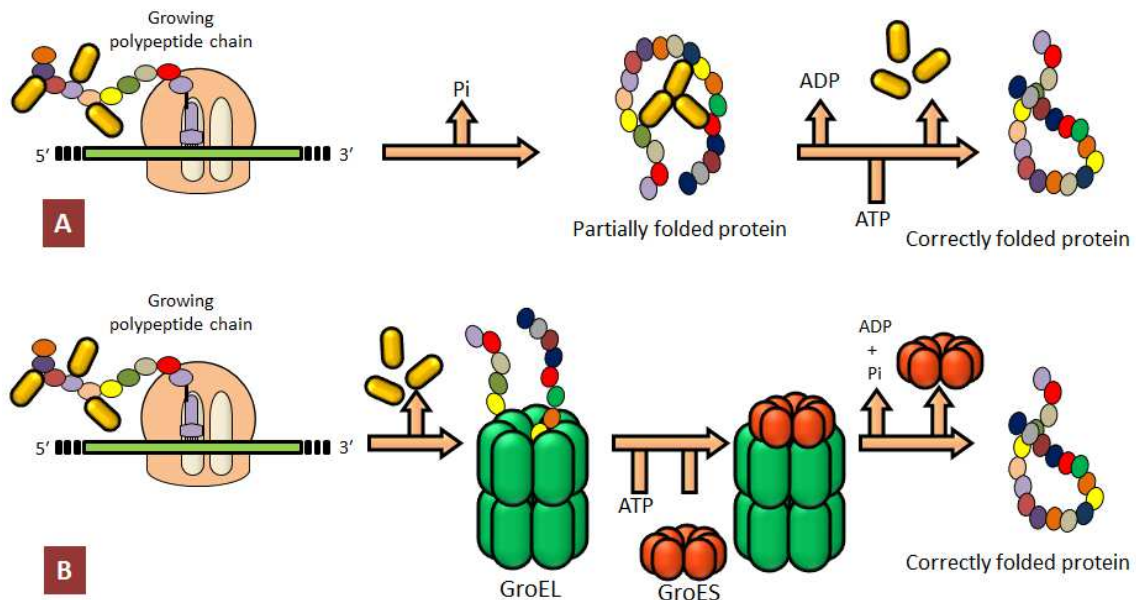


Figure (5.18); chaperone- and chaperonin-mediated protein folding. (a) Many proteins fold into their proper three dimensional structures with the assistance of Hsp70-like proteins (*top*). These molecular chaperones transiently bind to a nascent polypeptide as it emerges from a ribosome. Proper folding of other proteins (*bottom*) depends on chaperonins such as the prokaryotic GroEL in prokaryotes, a hollow, barrel-shaped complex of 14 identical subunits arranged in two stacked rings. One end of GroEL is transiently blocked by the co-chaperonin GroES.

Second; Amino Acids Side Chains Modifications

Many proteins contain, such as RNase and chemotrypsin, on amino acids only and don't contain on any other chemical compounds, so, these proteins called simple proteins. But some proteins contain permanently on chemical compounds in addition to amino acids, therefore they called conjugated proteins. The non-amino acid chemical component is called prosthetic group. The conjugated proteins are classified on the basis of prosthetic group chemical nature into many classes, such as lipoproteins; proteins having lipids, glycoproteins; proteins having sugar group, and metallo-proteins; proteins having certain metal. Moreover, many proteins contain more than one prosthetic group. Actually, these groups play an important role in the function of biological proteins. However, post-translational control of protein function or structure can be achieved through many methods such as the mentioned chemical modification of amino acids side chains. Although many chemical modifications of amino acids side chains are exist as it shown above, we satisfy only with mention of two examples of these modifications (figure 5.19).

a) Phosphorylation

Phosphorylation involves the addition of phosphate to an amino acid side chain, usually to the side chain hydroxyl (-OH) of serine, or tyrosine. This modification results from the action of a protein known as a kinase and uses ATP as the source of phosphate. The phosphate can be removed by another enzyme, a phosphatase. Phosphorylation can alter protein function and is relevant in cellular signaling pathways. Aberrant phosphorylation can lead to disruption of the cell cycle and the induction of cancer.

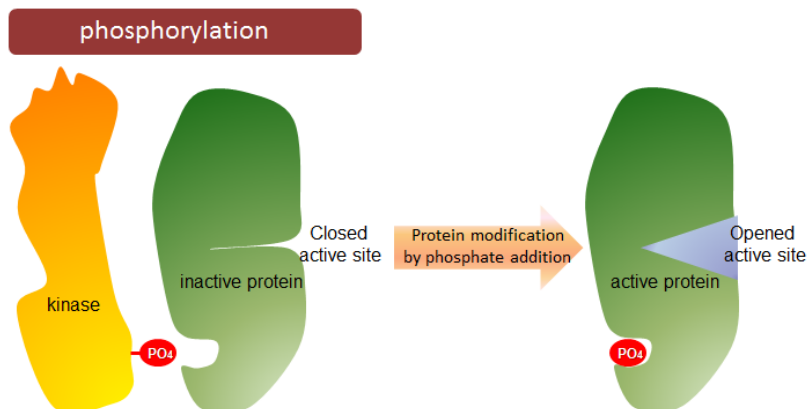


Figure (5.19);
posttranslational
control by protein
phosphorylation.

b) Glycosylation

Glycosylation involves the addition of one or more sugar monomers to the side chains of amino acids, either during or after translation, to make a glycoprotein. Sugars are attached either to the side-chain nitrogen of the amino acid asparagines or to the hydroxyl of the amino acids serine or threonine. The structure of these carbohydrates can be complex and variable and often does not affect the protein's stability, it's targeting to a particular part of the cell, it's folding into a three-dimensional structure, its lifetime before it is degraded, and its interaction with other proteins. The proteins that considered as typical example of this type of modifications are glycoproteins that found on the outer surface of cellular plasma membrane which function as specific receptors for different molecules (figure 5.20).

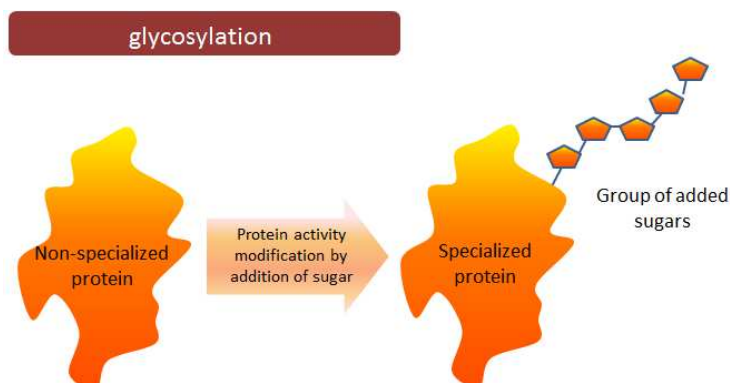


Figure (5.20);
posttranslational control by
addition of sugar groups.

Third; Polypeptide Backbone modifications

The control of protein function may also be occurring by the altering the order of amino acids arrangement in the protein backbone. These modifications may be promoted by other proteins or they may be self-directed (figure 5.21).

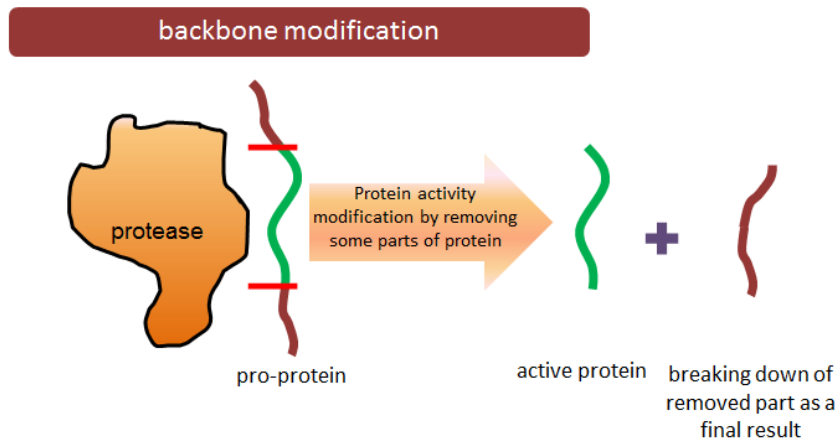


Figure (5.21); backbone modification of the newly synthesized polypeptide (primary protein) by removing some nucleic acids which – in some cases – by their removing the active protein is formed.

Many of large protein precursors, called zymogen or proproteins, are synthesized with an N-terminal signal sequence that instructs the cell to export the protein. (Proteins have an amino group at one end and a carboxyl group at the other. The amino-group end is called the N-terminal, the carboxyl end is called the C-terminus.) The N-terminal signal sequences are then cleaved, but the exported protein may still be inactive until cleaved again by another protease. Proteins activated by this mechanism include digestive proteases such as trypsin, the activity of which must be controlled before export by the cell. Serum albumin is also processed by this way, and the case is the same for insulin, vasopressin and oxytocin.

Proteins Breakdown

The living cells don't only synthesize proteins, but they break them as well. Although the process of protein breakdown is not as complex as its synthesis, this process is under strict control. Proteases are responsible about breakdown of proteins. So, they are – for this reason – dangerous on the organisms that made them, and they must be controlled carefully. Proteases are not the only pathway that responsible about protein breakdown, since there are much more sophisticated structures called proteasomes, These are cylindrical, with the protease active sites inside. The top and bottom of the cylinder are covered by protein complexes that recognize and bind damaged or unwanted proteins. Proteins destined for degradation are recognized because they are tagged with ubiquitin. This is a small protein that is fixed to damaged or mis-folded proteins and also to certain proteins that are needed only for a brief period (Fig. 5.22). The ubiquitin tagged proteins

are unfolded and then fed into the barrel of the proteasome where they are degraded into short peptides. The ubiquitin tags themselves are cleaved off and recycled.

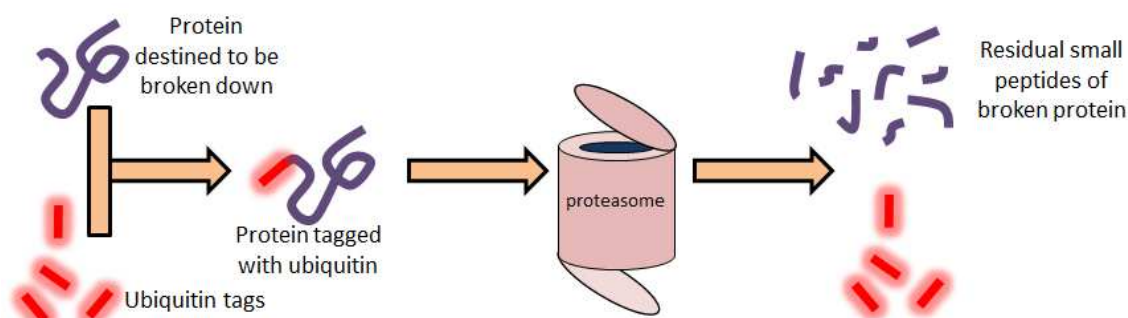


Figure (5.22); the mode of proteasome action. The broken down proteins are recognized by ubiquitin tags, then these proteins are converted into small polypeptides.

Chapter Five Appendix

The Inhibitors of Protein Synthesis.... A Tool for Treatment!

Bacterial ribosome differs from its eukaryotic counterpart, since bacterial ribosome is smaller (70S instead of 80S). This difference has been exploited for clinical purposes, since many active antibiotics are actively interacting with the proteins on which only prokaryotic ribosomes are acting on, thus, they inhibit their protein synthesis. Many antibiotics exert their effect through their targeting to translation in bacteria. They utilize the differences in translation mechanism between eukaryotes and prokaryotes, in order to selectively inhibit protein synthesis in bacteria without affecting on the host. Examples include; streptomycin, which causes an error in the reading of the genetic code in bacteria in relatively low concentrations, and inhibit translation initiation by binding with 30S ribosomal subunit (figure 5.23).

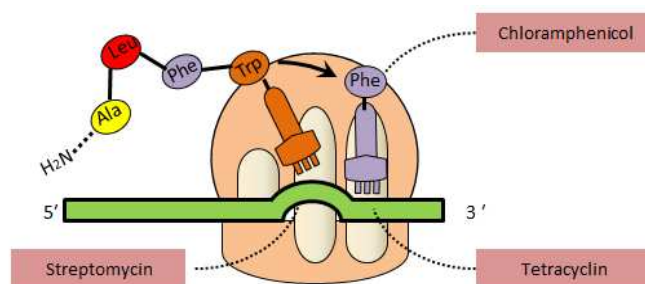


Figure (5.23). Inhibition of protein synthesis in prokaryotes by antibiotics. Different points of effects are exerted by chloramphenicol, tetracycline, and streptomycin respectively on protein synthesis. Chloramphenicol binds to 50S portion and inhibits formation of peptide bond. Tetracycline interferes with attachment of tRNA to mRNA-ribosome complex. Streptomycin changes shape of 30S portion, causing code on mRNA to be read incorrectly.

While kanamycin prevents other bindings at the end of initiation, and causes error in the reading of the genetic code. The antibiotic puromycin, which has a structure similar to tyrosinyl aminoacyl tRNA, binds to A-site and participates in the formation of peptide bond, producing the peptidyl-puromycin. However, it doesn't involve in translocation and it dissociates quickly from ribosome causing early termination for polypeptide synthesis. Puromycin is efficiently inhibits the protein synthesis in prokaryotes and eukaryotes. Diphtheria toxin, an exotoxin for *Corynebacterium diphtherium* that infected by a specialized lysogenic phage, triggers ADP ribosylation of EF-2 in mammalian cells. This modification inhibit EF-2, thus, it specifically inhibits the protein synthesis. Many animals (such as mice) are resistant to *diphtherium* exotoxin. This resistance belongs to the inability of this exotoxin to penetrate cell membrane in addition to insensitivity

of mice EF-2 to *diphtherium* exotoxin. The antibiotic tetracycline blocks A-site in ribosome, preventing the binding of aminoacyl tRNAs with it. The antibiotic chloramphenicol blocks the peptidyl transfer step that taken place in elongation stage of the large bacterial ribosomal subunit (figure 5.23).

Chapter Five Revision Questions

Q1- Justify the following:

1. The translation of genetic information that present in mRNA sequence into amino acid sequence in protein requires intermediate molecules
2. The hypothesis saying that codons are triple is the right one
3. The codon AUG don't always encodes methionine
4. The antibiotic tetracyclin is an efficient inhibitor of protein synthesis in bacteria
5. The genetic code is described with universality
6. Its possible for some aminoacyl-tRNA molecules to recognize more than one codon
7. Most of mRNA molecules read in only one frame
8. The translation of the polypeptide chain in mRNA ends once UAA sequence is reached

Q2- choose the correct answer:

1. The first one who synthesized the sequence CUCUCUC is the scientist
a- Nirenberg b- Khorana c- Watson d- Crick e- Holley
2. Translation in most bacteria is started with modified methionine, while in mammals, the unmodified methionine is used (except in, in which their ribosomes resembles these found in bacteria).
a- Golgi apparatus b- endoplasmic reticulum c- nucleotides d- mitochondria and chloroplast e- nuclear envelope
3. The antibiotic puromycin inhibits protein synthesis through
a- Its similarity to tyrosinyl aminoacyl tRNA b- EF-2 modification c- error in the reading of the genetic code d- blockage of peptidyl transfer site
4. the ribosomes that exist in bacteria are called
a- 40 S ribosomes b- 50 S ribosomes c- 60 S ribosomes d- 70 S ribosomes e- 80 S ribosomes
5. The sequence is working on the alignment of mRNA on ribosome in the beginning of translation by using the base pairing with the complementary sequence near the 3' end of rRNA.
a. Enhancer sequence b. 5' Cap sequence c. Shine-Dalgarno sequence d. promoter sequence e. operon sequence
6. The tRNA molecule that initiates translation is called in prokaryotes.
a. Met-tRNA b. uncharged tRNA c. fMet-tRNA d. (a) & (c) e. none
7. The factor prevents the binding of large ribosomal subunit with the small subunit in prokaryotes.
a. IF1 b. IF2 c. IF3 d. IF4 e. IF5
8. the first step of translation initiation is the binding of mRNA with in prokaryotes.
a. large ribosomal subunit b. 5' cap of mRNA c. met-tRNA d. initiation factors e. none
9. the 400 amino acids length protein requires to 20 seconds to be synthesized under the optimum conditions. Therefore, the time required to synthesize one amino acid is estimated part of second.
a. ten b. twenty c. thirty d. forty e. fifty
10. the direction of tRNA flow during translation in ribosome is
a. (E-A-P) sites b. (E-P-A) sites c. (A-E-P) sites d. (P-A-E) sites e. (A-P-E) sites
11. polyribosome is unable to bind with mRNA with a distance of less than nucleotides apart.
a. 60 b. 70 c. 80 d. 90 e. 100
12. the resulting protein from endoplasmic reticulum bound protein is flow out to an open space called
a. Golgi apparatus b. cisterna space c. zymogen particle d. cytosol e. Peroxisomes
13. when we say that genetic code is specific, that's mean its
a. ubiquitous b. non-ubiquitous c. ambiguous d. unambiguous e. promiscuous

14. The enzyme randomly polymerizes nucleotides without depending on the template strand.
a. primase b. topoisomerase c. DNase d. polynucleotide phosphorylase e. DNA dependent DNA polymerase

15. the factor is responsible on the slowness of translation.
a. IF1 b. EF1 c. IF2 d. EF2 e. RF

Q3- define the following:

polysomes, nonsense codons, chaperons, heat shock proteins, wobble hypothesis, reading frame, Shine-Dalgarno sequence, Kozak sequence, zymogen, chaperonins

Q4- connect terms exist in right with its suitable counterparts in left

RNA polymerase	Polymerase want template and primer
DNA polymerase	Polymerase want template and don't want primer
None	Polymerase don't want template and want primer
Polynucleotide phosphorylase	Polymerase don't want template and don't want primer

Further Readings

- Alberts** B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science, USA, 2008.
- Anfinsen**, C. 1973. Principles that govern the folding of protein chains. *Science* 181: 223-230.
- Berk** A., Zipursky S., Baltimore D., Darnell J., and Lodish H. Molecular Biology. Fourth edition. Paul Matuataira, USA, 1998.
- Clark** D. Molecular biology: understanding the genetics revolution. Elsevier, 2005.
- Garret**, R., Douthwaite, S. R., Liljas, A., Matheson, A. T, Moore, P. B., and Noller, H. F., The Ribosome: Structure, Function, Antibiotics and Cellular Interactions. The American Society for Microbiology. 2000.
- Gesteland**, R. F., Cech, T., and Atkins, J. F. (Eds.) *The RNA World*. Cold Spring Harbor Laboratory Press, 1999.
- Glazer** A., Derange, R., and Sigman, D. *Chemical Modification of Proteins*. North-Holland, 1975.
- Hames** B., and Hooper N. Instant notes in biochemistry. Third edition, BIOS Scientific Publishers Limited, 2005.
- Ibba** M., Becker H., Stathopoulos C., Tumbula D., and Söll D. 2000. The adaptor hypothesis revisited *Trends Biochem. Sci.* 25: 311-316.
- Koolman** J., and Roehm K. Color Atlas of Biochemistry. Second edition, Thieme, 2005.
- Lodish**, H., Berk, A., Zipursky, L., and Matsudaira, P., Molecular Cell Biology (fifth edition). W. H. Freeman and Company, 2000.
- Murray** R., Bender D., Botham K., Kennelly P., Rodwell V., Weil P. Harper's Illustrated Biochemistry. 28th edition, McGrawHill Medical, 2009.
- Passarge** E. Color atlas of Genetics. Second edition, Thieme, 2001.
- Pierce** B. Genetics: A conceptual approach, 2002.
- Robinson** R. Genetics. Volume 3. Thomson Gale, 2003.
- Schultz**, G. E., and Schirmer, R. H., 1979. Principles of Protein Structure. Springer-Verlag.
- Tamarin**. Principles of genetics. Seventh edition. The McGraw-Hill. 2001.
- Watson** J., Baker T., Bell S., Gann A., Levine M., Losick R. Molecular biology of the gene. Fifth edition, Pearson Prentice Hall, 2004.

Chapter Six

Regulation of Gene Expression



Excessive increasing in the expression of growth hormone of the biggest man in the world besides excessive decreasing in the expression of the same hormone for the shortest man in the world!

Introduction

Many researchers have discussed of how genes exert their effect, the most famous two in this aspect might be George Beadle and Edward Tatum. They studied the microorganism *Neurospora* by growing it in a medium of different chemical composition. Both Beadle and Tatum designed a system to detect specific changes in the biochemical abilities of the organism of interest. The precise study of *Neurospora* made researchers to postulate "one gene one enzyme" theory, which connects the genes with specific enzymes and with the chemical reaction they catalyze it.

Later, researchers discovered that many proteins are not enzymes, but instead, they may act as signaling molecules or receptors, or may play structural roles. Thus, "one gene one enzyme theory" had been modified into "one gene one protein" theory. Add to that, researchers have discovered that many functional proteins are composed of many of characterized amino acid chains (polypeptides) that correspond to DNA which not necessarily found adjacently to each other or even on the same chromosome. This led to the formation of "one gene one polypeptide" in gene definition. Therefore, the definition of the gene is a fragment of DNA contains the necessary information to make a specific polypeptide (or specific subunit) of specific function (instead of specific protein).



George Beadle



Edward Tatum

Regulation of Gene Expression in Prokaryotes

The biological importance of studying gene expression is held on many reasons; the most important of them is attributed to our understanding to the mechanisms of gene expression leads us to develop the inhibiting or stopping factors of pathogenic organisms. Consequently, the regulation of gene expression mechanisms was studied first in bacterial cells. And because of the abundance of mutants and the feasibility of their practical manipulation, it was rational to unmask their mechanisms. When these mechanisms were started to be studied in eukaryotes, it was so obvious that regulation of gene expression in eukaryotes widely differs from its prokaryotic counterpart.

Since more than half a decade, with the understanding of the flow of genetic information from the gene into a specific protein molecule. Since that time, complex concepts about the mode of regulation of gene expression in bacteria had been developed. And before we deal with the details of regulation of gene expression, we have to know some genetic terms related with this topic. The most important of them is may be the cistron; the genetic unit that encodes the protein subunit, acting as the smallest unit in gene expression. Thus, be more precise, the idea of "one gene one enzyme" must be called "one gene one cistron", or "one subunit concept". Moreover, we have to know the fact that not all genes are inducible, since some of them are far away from the domination of gene expression. However, the inducible gene is identified as the gene that its expression is increased as a response to a specific regulating signal known as inducer. As for these genes that are far away from the negative or positive gene effect are known as genes of constitutive expression, this means that they are expressing with a steady rate and it wasn't known that they are subjected to any regulation.

First; Transcriptional Control on Gene Expression

E. coli have been separately studied through many years to understand how regulation of gene expression is performed. *E. coli* contain single chromosome, its length 4.5×10^6 bp, this amount of DNA is sufficient to encode about 3000 genes. Under is the active growth condition, only 5% of chromosome is actively transcribed at any time of these conditions. As for the rest, its either being silent or its being transcribed with very slow rate. When growth conditions are changed, some active genes are stopped while the non-active genes didn't. it's so important to notify that the cell must respond to the changes that taken place in the environment by changing gene production, therefore, within a short time, in most cases seconds to minutes, it's possible for any gene to stop or start working entirely.

In *E. coli*, most of RNA and protein production is regulated at transcriptional level. The rapid response of conditions changes are partially supported by a short life cycle of mRNA (in which it will be for 1 to 3 minutes for most of mRNA molecules). Some mRNA molecules have longer life cycles (10 minutes or more), this mean bigger production rates of proteins for each mRNA molecule. These non-typical molecules of mRNA might be exposed to control at translational level.

The Discovery of Operon

Both Jacob and Monod have discussed their model that known as operon in a classical paper at 1961. Their hypothesis was depended on large extent on the lactose metabolism regulation by *E. coli*. The molecular mechanisms that are responsible on genes regulation of lactose metabolism are the most understood mechanisms in any organism. The enzyme β galactosidase hydrolysis lactose into its components; galactose and glucose. For *E. coli* to utilize glucose as a source of energy, it breaks lactose down into glucose and galactose in a reaction catalyzed by β galactosidase. It's possible for this enzyme to convert lactose into allolactose, a compound that plays an important role in the regulation of lactose metabolism. When allolactose generated (which is produced from a minor reaction performed by β -galactosidase), it reduces the affinity of repressor binding with the operator site for about thousand folds. Consequently, the repressor dissociated from operator and the transcription begins (figure 6.1)



Francis Jacob



Jacques Monod

The structural gene of β galactosidase (known as *lac Z* gene) binds with the structural gene of β galactoside permease (*lac Y* gene) that is responsible on penetration of lactose inside the cell, and with the structural gene of thiogalactoside transacetylase (*lac A* gene) which its function sum up is to remove lactose like compounds that β galactosidase cannot break them. The structural genes of these three enzymes are physically bound to form the lactose (*lac*) operon as it shown in the figure (6.1).

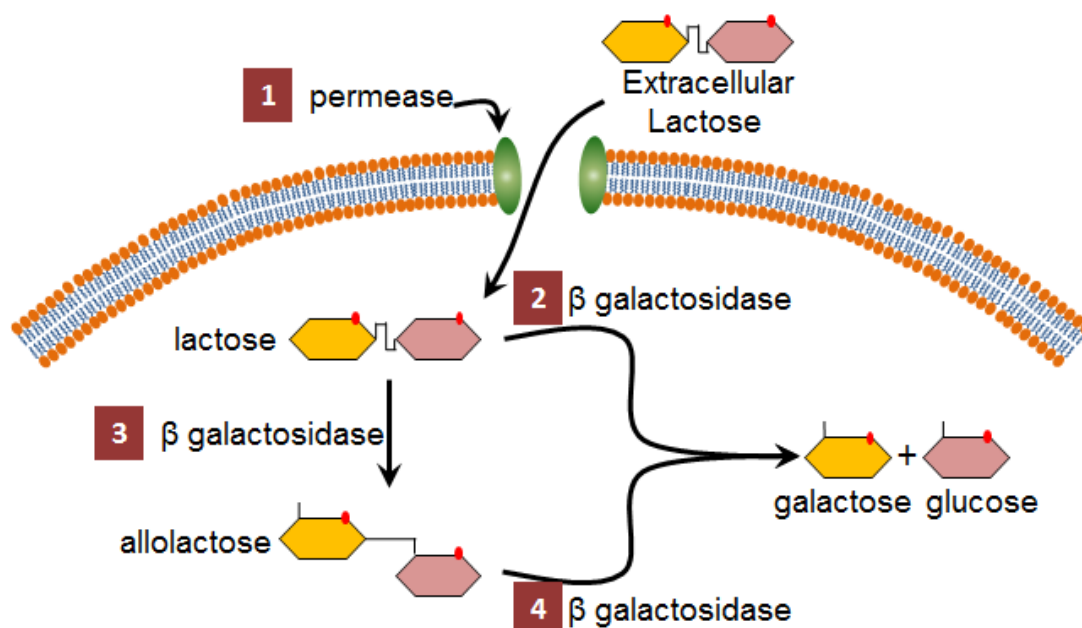


Figure (6.1); the mode of lactose breakdown by *E. coli* enzymes in order to be utilized as a source of energy through several steps; 1) Permease actively transports lactose into the cell, 2) where the enzyme β -galactosidase breaks it into galactose and glucose, 3) β -Galactosidase also converts lactose into the related compound allolactose, 4) and converts allolactose into galactose and glucose.

This genetic structure of the structural genes and their regulating genes allows the coordinate expression of the three enzymes that are involved in lactose metabolism, in such way all these connecting genes are transcribed into one large mRNA molecule that contains several translation start sites (AUG) and stop codons (UAA) for each cistron. This sort of mRNA is called polycistronic mRNA which are available permanently in prokaryotic organisms. Jacob and Monod researches were led to what is called now operon hypothesis. The operon is a group of genes under the control of one operator site in the same DNA strand (*in cis*).

Lac Operon

Lac operon consists of two types of genes; structural genes, (genes that simply codes for proteins or RNA which are necessary for the normal cellular functions), which are three genes in the case of lac operon; lacZ gene; codes for β galactosidase, lacY gene codes for β galactoside permease, and lacA gene; codes for thiogalactoside transacetylase. These three structural genes are regulated as a unit which is lacZ-lacY-lacA which is expressing as on unit from lacZ into lacA (figure 6.2).

Regulatory genes are genes that codes for proteins or RNA which their function is to regulate the expression of other genes. It's possible for this type of genes to be produced in one region and act in another region (*in trans*). Some of the regulatory proteins that are resulted from these genes are called repressors. The DNA sites on

which repressors bind are called operators. In the absence of repressor, it's possible for RNA polymerase to bind with promoter and start in transcription of operon. In the presence of repressor, RNA polymerase is unable to transcribe operon.

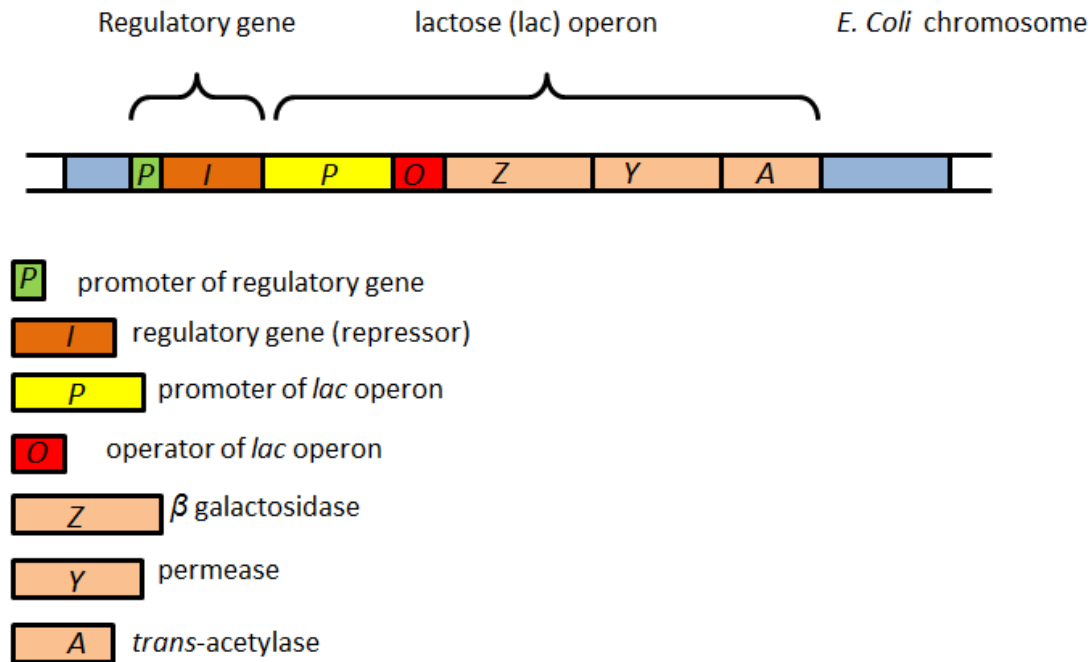


Figure (6.2); the gene structure of *lac* operon in *E. coli*.

The regulatory gene *lacI* is located nearby and upstream the structural genes *lacZ-lacY-lacA*. It's not necessary of this regulatory gene to be located in very close region of the genes that regulates them, and it can be affect them in some cases even when its location is far away from them, therefore it's called regulatory gene. It differs in this feature from these structural genes that are never involved in regulation of other genes action, even when it's located near them. The normal expression of *I* gene in *lac* repressor is described as constitutional expression, in which it will be expressed in a steady rate, this result in the formation of *lac* repressor subunits. The repressor protein molecule, *I* gene product, has high affinity to operator site with which it binds. The operator site is located between promoter region with which RNA polymerase binds to start transcription, and transcription start site of *Z* gene, the structural gene of β galactosidase (figure 6.3). When the repressor molecule binds with the operator site, it prevents the transcription of operator site in addition to the structural genes that represented by *A*, *Y* and *Z* genes. Thus, the repressor molecule is a negative regulator, when it's present (and in the absence of inducer, see below) it prevents the expression of *A*, *Y*, and *Z* genes.

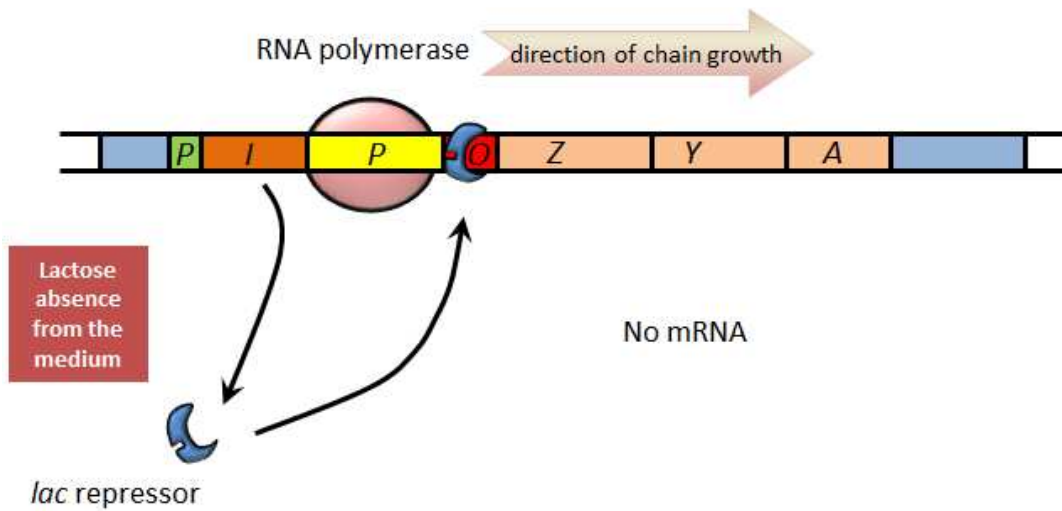


Figure (6.3); the role of lac repressor in hampering transcription of *lac Z*, *Y*, and *A* genes in the absence of the inducer molecule (lactose). Note; this picture talks about lactose irrespective of glucose

When *I* gene is mutated in such way it protein product (lac repressor) is unable to bind with operator DNA, hence the organism exert a constitutive expression of lac operator. In contrast, the organism that containing *I* gene mutation that prevent the binding of inducer it will stay inhibited even in the presence of the inducer molecule, since the inducer is unable to bind with operator on the operator site to remove operon repression. When small amounts of lactose are added, these amounts are able to get into the cell even in the absence of permease. Both of the repressor molecules, these that connected with the operator site, and these that not connected with the operator site have great affinity to the inducer. The binding of inducer with the repressor molecule that is bound with the operator site triggers conformational changes in the repressor appearance, and leads to its separation form DNA (figure 6.4). if RNA polymerase is attached with the coding strand at promoter region, transcription begins. RNA polymerase initiates what is known polycistronic mRNA. In this way, the inducer molecule (lactose here) removes the repression of lac operator and allows transcription of the structural genes of β galactosidase, galactoside permease, and galactoside transacetylase. Thus, removing of *lac* operator gives the cell the ability to synthesize the necessary enzymes to breakdown lactose as a source of energy.

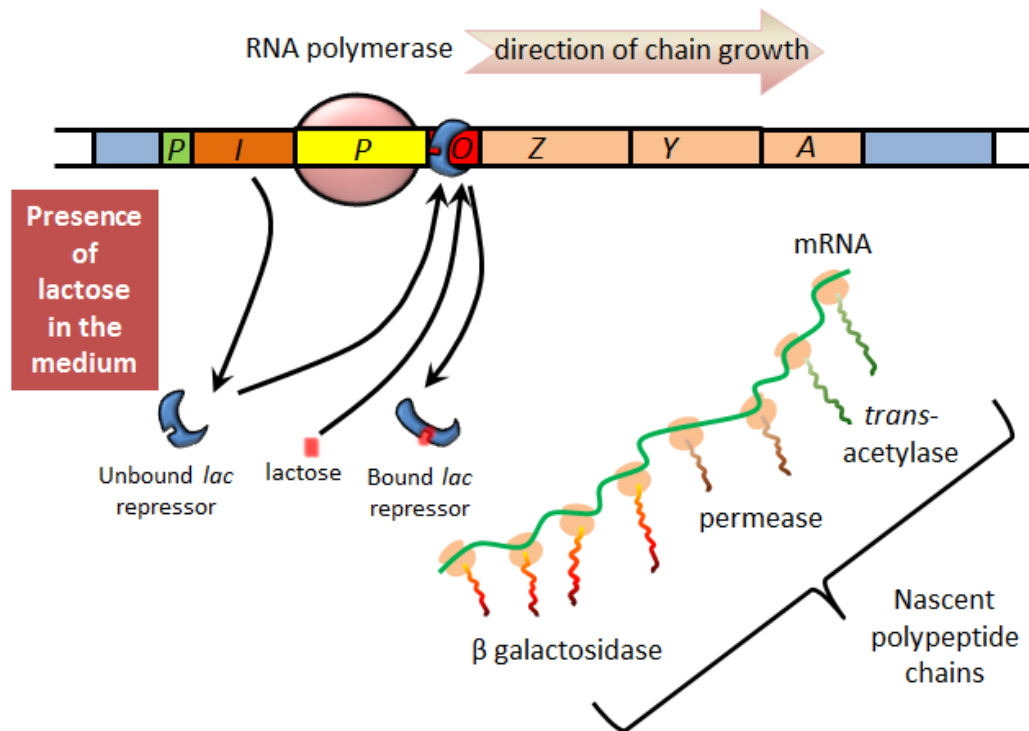


Figure (6.4); the inhibition of lac repressor role in hampering the transcription of lac operator genes in the presence of the inducer (lactose). Note; this picture talks only about the effect of lactose on lac operon irrespective of glucose.

There is another regulatory protein of other sort is called CAP protein, its role is appeared when *E. coli* exposed to both lactose and glucose as a source of carbon. Organisms start metabolizing glucose, and then stop growth temporarily until *lac* operon genes become induced to give the ability to metabolize glucose. Although of lactose presence since the start of bacterial growth phase, the cell doesn't induce these enzymes to catabolize until glucose is depleted. This phenomenon is attributed to *lac* operon repression by some glucose catabolizers. Hence, this process is called catabolite repression, in which catabolite gene activator protein (CAP) with cyclic AMP (cAMP) are mediated in this process (figure 6.5).

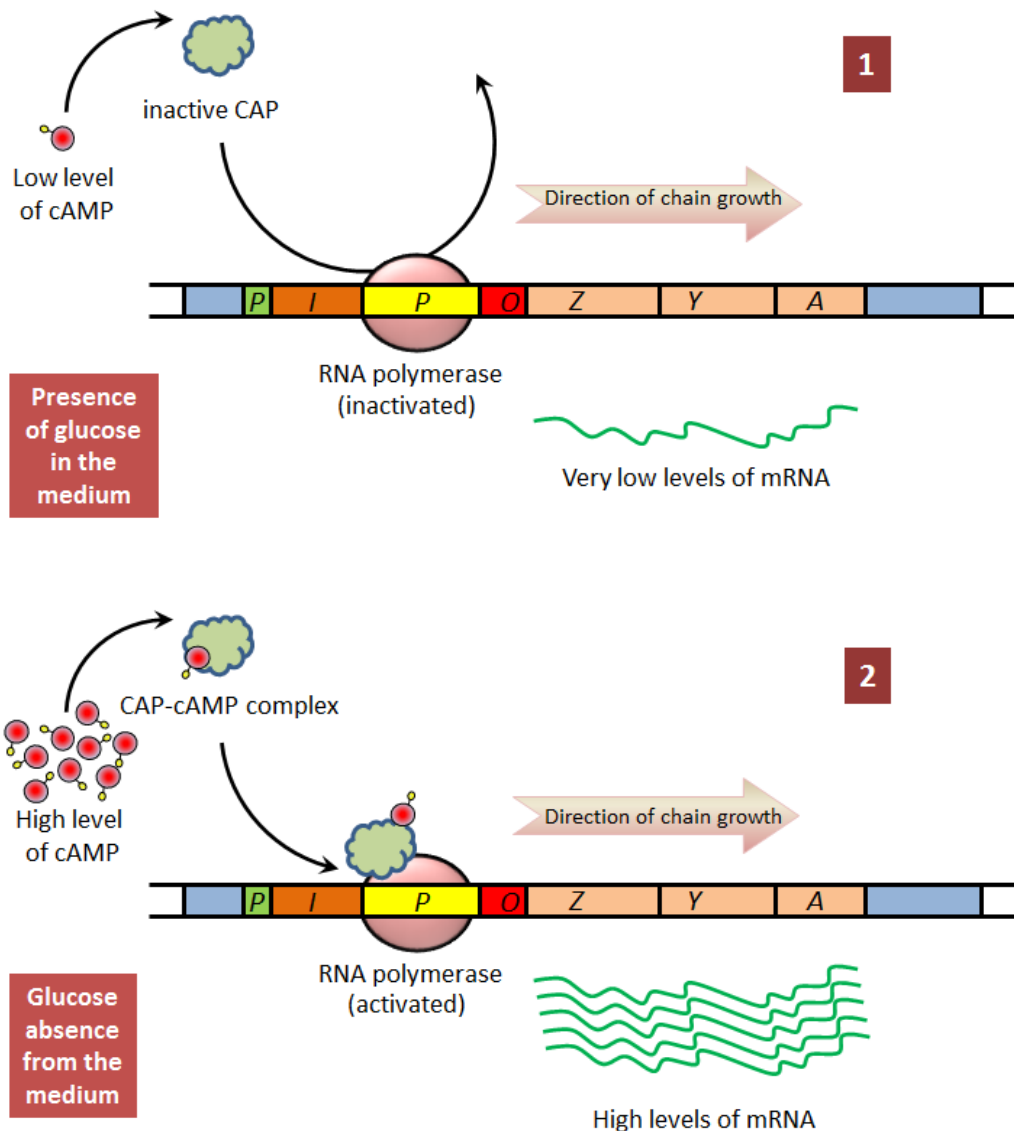


Figure (6.5); the role of CAP protein and cAMP in *lac* operon regulation. Illustration No. refers to the events that taken place in the presence of glucose in the culture medium. Illustration No. 2 refers to the events that taken place in the absence of glucose from the culture medium. Notice; this picture talks about the effect of the presence or absence of glucose only on *lac* operon irrespective of lactose.

In order to attach RNA polymerase with promoter region, CAP protein must be present. By an independent mechanism, bacteria accumulate cAMP only when they are starved from glucose. When glucose present in sufficient concentration for growth, bacteria lose the binding of cAMP with CAP. Thus, when glucose present, cAMPs level decreases acutely in such away CAP doesn't find any CAP to activate it, thus, CAP is loses its ability to bind with the activator binding site. Consequently, RNA polymerase is unable to start *lac* operon transcription. When glucose absent, the level of cAMP concentration is highly increased. Hence, CAP activated to form CAP-cAMP complex,

which binds with DNA slightly upstream of promoter region, hence, transcription can then be taken place (figure 6.5). Thus, CAP-cAMP complex acts as positive regulator, since its presence is necessary to induce gene expression. As for CAP, it's an activator stimulates the binding with the same promoter. But, when we want to be precise, we have to refer to CAP as apoactivator instead of being activator, since it induces transcription only when it forms a complex with a small molecule called coactivator 3-5 cyclic AMP (cAMP). So, *lac* operon is exposed to both negative and positive types of control. But both controls doesn't act separately from each other, instead, *lac* operon is exposed to both of them simultaneously. By this way, the effect of each of them is couldn't be taken into account alone.

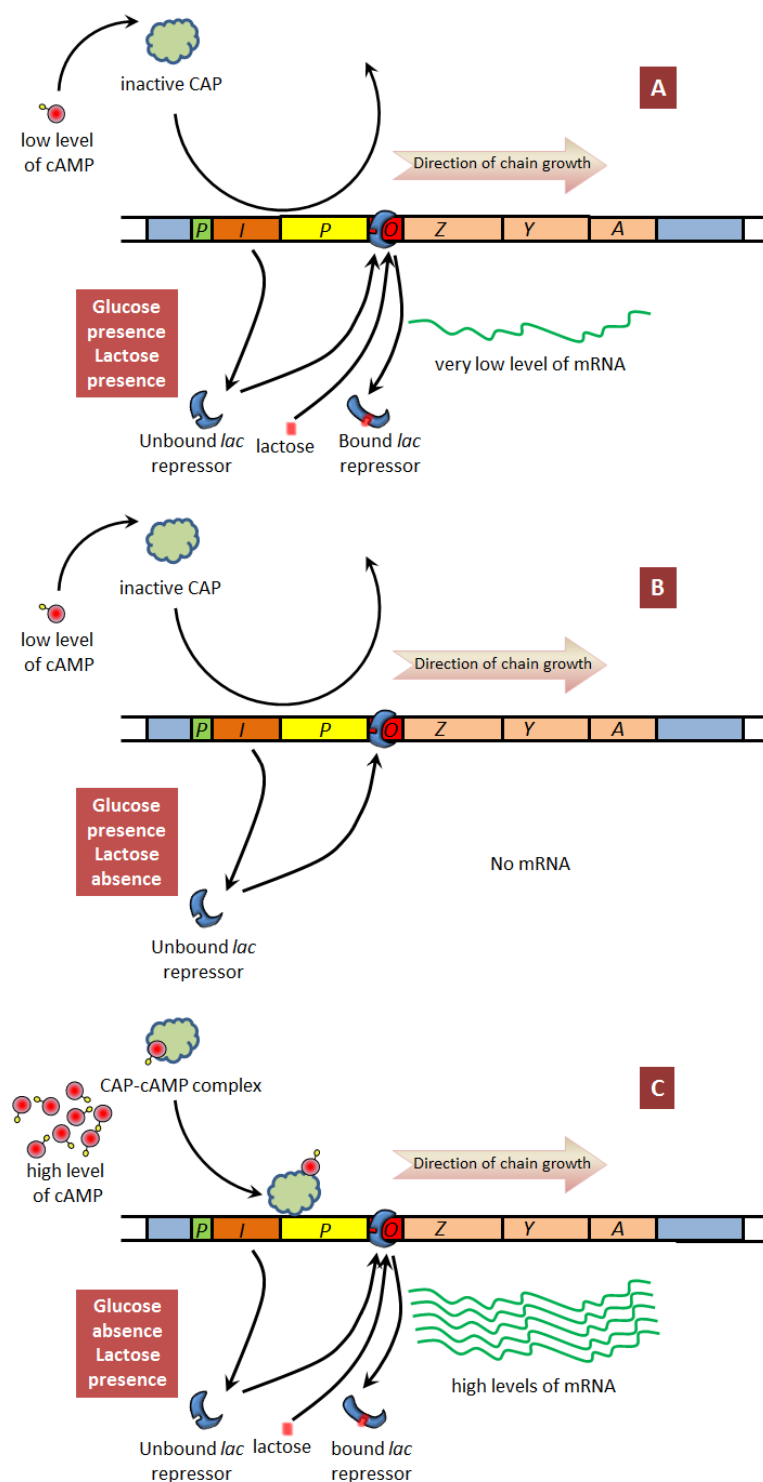
***Lac* Operon; The Complete Picture**

Many bacterial types grow on glucose before they utilize from other compounds such as lactose as growth substance, in case of both sources (glucose and lactose) are present in the culture.

The *lac* operon is transcribed by two types of regulatory proteins, CAP and *lac* repressor, which act together to ensure the production of the three necessary proteins of lactose growth in the absence of glucose (the efficient substrate). It's possible for CAP to bind with a DNA region near promoter called activator binding site and facilitates transcription. And it's possible for *lac* repressor to bind with the operator site and stops the transcription.

As it was discussed above, it's possible for CAP to bind with activator binding site only if cAMP is available in high concentration, and the later is doesn't available in high concentration unless the glucose is absent. When CAP binds with cAMP, the resulting complex binds with the activator binding site and allows RNA polymerase to bind with the promoter. As it was discussed also, if lactose is not available in the culture medium, *lac* repressor binds with the operator site, and this stops transcription. So, how each of the positive regulator that is represented by cAMP-CAP complex and the negative regulator that is represented by *lac* repressor to give a complete picture for *lac* operon induction? To answer this question, we have to review four main cases of glucose interaction with lactose and the effects of these interactions on the induction of *lac* operon.

The first case; when both glucose and lactose are available, there is no sufficient concentration for cAMP to bind with CAP. Therefore, it's impossible for it to bind with activator binding site. Without the binding of CAP, RNA polymerase is unable to undertake noticeable transcription, in such away very little amounts of mRNA are produced in this case even when lactose binds with its repressor (*lac* repressor) and ban it from binding with the operator site (figure 6.6 a).



The second case; when glucose is available and lactose is absent from the culture, there is no sufficient concentration of cAMP to bind with CAP because the presence of glucose. Therefore, it's impossible for CAP to bind with the activator. Consequently, RNA polymerase is unable to start transcription of *lac* operon because the absence of the positive regulator. Since lactose is absent, *lac* repressor remains attached tightly with the operator of *lac* operon. And this is another obstacle against *lac* operon induction (figure 6.6 b).

Figure (6.6); the couple effect for both glucose and lactose on the regulation of *lac* operon which includes (a) the presence of both glucose and lactose, (b) the presence of glucose and absence of lactose, and (c) the absence of glucose and presence of lactose.

The third case; when glucose is absent and lactose is available in the culture, hence, cAMP binds with CAP, which in turn binds with the activator binding site and facilitates transcription. Lactose binds with the repressor to turn it into a conformation that could not bind with *lac* promoter. By this way, it prevents it from the binding with the operator site. And it's now possible for RNA polymerase to bind with the promoter and achieve transcription (figure 6.6 c).

The fourth case; when both glucose and lactose are not available, the concentration of cAMP is high and CAP binds with the activator binding site. Hence, RNA polymerase is unable to bind with promoter but its stopped by *lac* repressor that bound with the operator site.

It's obvious from the above discussion that *lac* operon induction is an interactive process between both the positive and negative regulators altogether which is affected differently in the presence or absence of glucose and lactose in *lac* operon induction. Therefore, it's impossible to look at *lac* operon induction separately.

Tryptophan Repressor

The tryptophan (*trp*) operon consists of five structural genes (*trpE*, *trpD*, *trpC*, *trpB*, and *trpA*). These five genes are contributed in the production of the structural components of three enzymes. These enzymes convert chorismate into tryptophan. The first structural gene (*trpE*) contains of 5'-untranslated region which is abbreviated 5'-UTR that is transcribed but does not encode any of these enzymes (figure 6.7). Instead, this 5'-UTR plays an important role in another regulatory mechanism (see below).

Upstream of the structural genes is the *trp* promoter. When tryptophan levels are low, RNA polymerase binds to the promoter and transcribes the five structural genes into a single mRNA, which is then translated into enzymes that convert chorismate into tryptophan. Some distance from the *trp* operon is a regulator gene, *trpR*, which encodes a repressor that alone cannot bind DNA (see Figure 6.8).

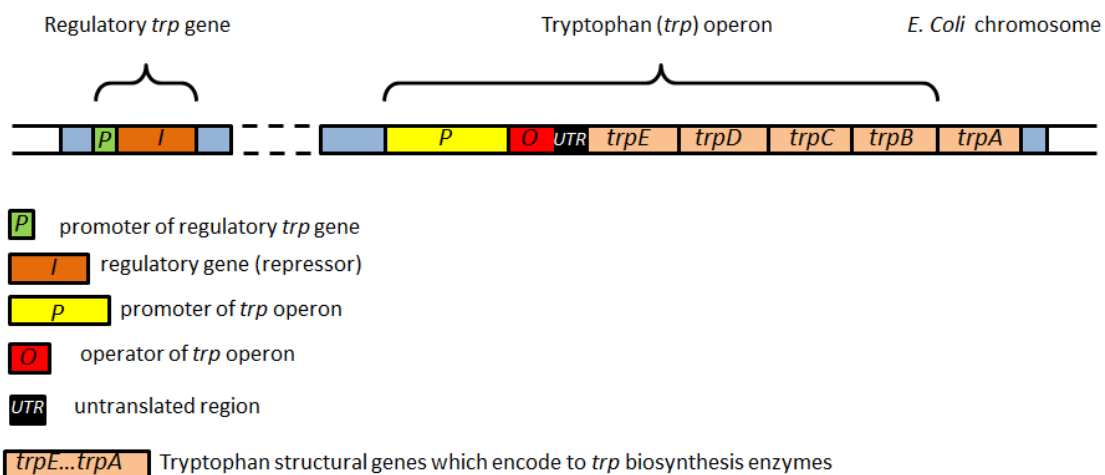


Figure (6.7); the genetic structure of *trp* operon in *E. coli*.

Like the *lac* repressor, the tryptophan repressor has two binding sites, one that binds to DNA at the operator site and another that binds to tryptophan (the activator). Binding with tryptophan causes a conformational change in the repressor that makes it capable of binding to DNA at the operator site, which overlaps with the promoter to repress the transcription of *trp* operon genes.

When the operator is occupied by the tryptophan repressor, RNA polymerase cannot bind to the promoter and the structural genes cannot be transcribed. Thus, when cellular levels of tryptophan are low, transcription of the *trp* operon takes place and more tryptophan is synthesized; when cellular levels of tryptophan are high, transcription of the *trp* operon is inhibited and the synthesis of more tryptophan does not take place.

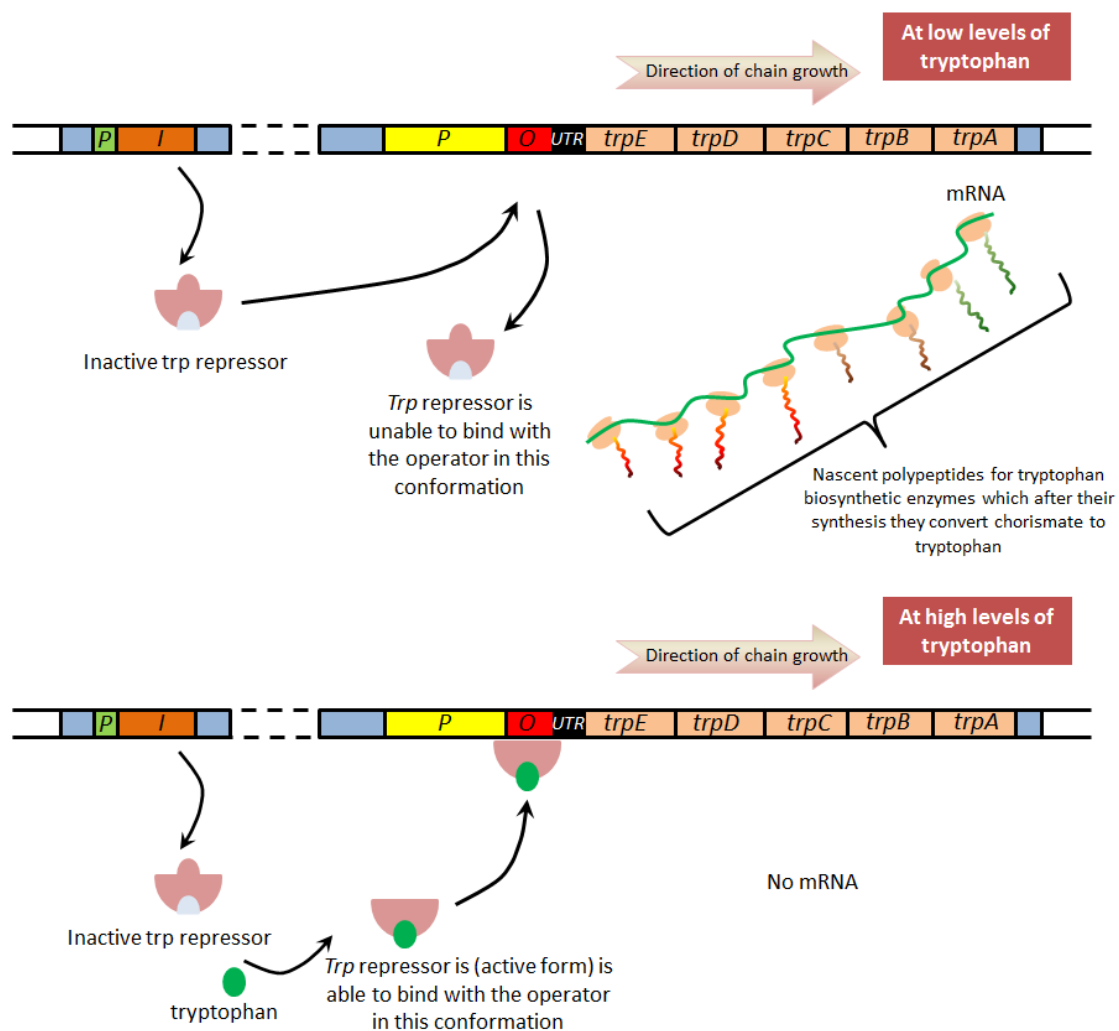
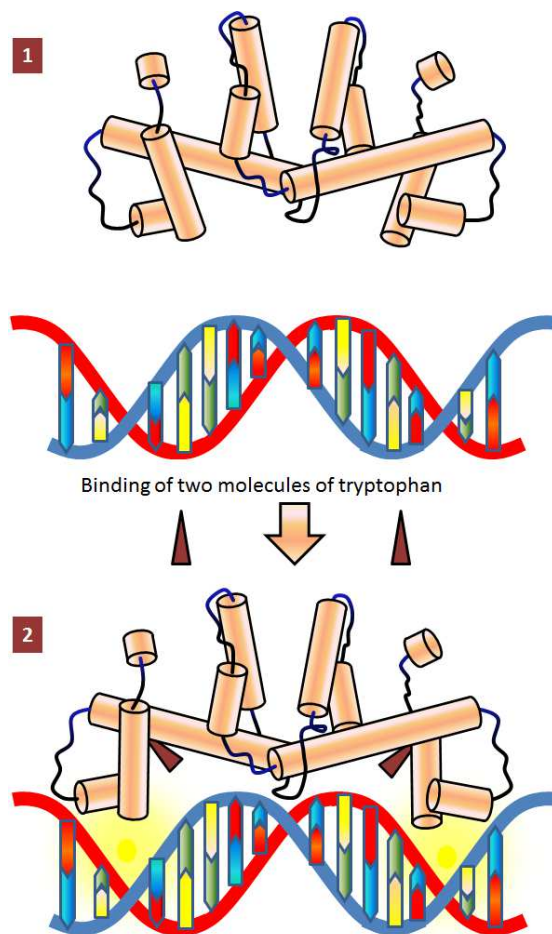


Figure (6.8); control of the amino acid tryptophan in *E. coli* by the control of *trp* operon in both high and low levels of tryptophan.



domains of the two helices of this protein. This allows the repressor to coordinate comfortably with the operator and this allows the action of repressed genes (in case No. 1).

The tryptophan repressor is a regulatory protein that takes a conformation called a helix-turn-helix. This conformation has a fundamental role in the regulation of repressor action. When tryptophan is absent from the environment, the repressor takes an inactive conformation, and it is unable to bind with DNA to prevent the transcription. When tryptophan is available, two of these molecules bind with the repressor. This changes the direction of the helix-turn-helix motif in such a way that it makes it competent to bind with the neighboring DNA major groove. Thus, tryptophan synthesis is taken place in need, but it is inhibited when tryptophan is available (figure 6.9).

Figure (6.9); binding of tryptophan with *trp* repressor protein changes the latter conformation. These structural changes enable a tight binding of the gene regulator protein with the operator sequences. In the case No. 1, the repressor prevents the transcription of the coding genes of the necessary enzymes for *trp* operon production by its three-dimensional structure of helix-turn-helix type that assists it to do so. In the case No. 2, the binding with tryptophan increases the distance existing between the recognition

Like *lac* repressor, *trp* repressor binds with a site by which it's possible for it to interact with the RNA polymerase binding site. The main difference between *lac* repressor and *trp* repressor's action is correlated with the function of the small effector molecule. In the case of lactose, the effector molecule represented by lactose acts as an anti-repressor, causing the releasing of the repressor from the operator. In the case of tryptophan, the effector molecule represented by tryptophan acts as a corepressor, inducing the repressor binding with the operator.

Attenuation: the premature termination of transcription

In addition to the presence of a regulation between the positive and negative control on transcription initiation in the operon system, some operon types have an additional level of control which affects the continuity of transcription instead of affecting its

initiation. In attenuation, transcription starts at the start site, but termination is prematurely taken place and even before RNA polymerase reaches into the structural genes. Attenuation is taken place on several types of operons which codes for the contributing enzymes in the biosynthesis of amino acids.

We can understand the process of attenuation easily by looking at one of the best-studied examples, which is found in the *trp* operon of *E. coli*. Several observations by Charles Yanofsky and his colleagues in the early 1970s indicated that repression at the operator site is not the only method of regulation in the *trp* operon. They isolated a series of mutants that possessed deletions in the transcribed region of the operon. Some of these mutants exhibited increased levels of transcription, yet control at the operator site was unaffected. Furthermore, they observed that two mRNAs of different sizes were transcribed from the *trp* operon: a long mRNA containing sequences for the structural genes and a much shorter mRNA of only 140 nucleotides. These observations led Yanofsky to propose that another mechanism - one that caused premature termination of transcription - also regulates transcription in the *trp* operon.

Close examination of the *trp* operon reveals a region of 162 nucleotides that corresponds to the long 5'-UTR of the mRNA (mentioned earlier) transcribed from the *trp* operon (figure 6.10 a). The 5'-UTR (also called a leader) contains four regions: region 1 is complementary to region 2, region 2 is complementary to region 3, and region 3 is complementary to region 4. These complementarities allow the 5'-UTR to fold into two different secondary structures (figure 6.10 b). Which secondary structure is assumed determines whether attenuation will occur.

One of the secondary structures contains one hairpin produced by the base pairing of regions 1 and 2 and another hairpin produced by the base pairing of regions 3 and 4. Notice that a string of uracil nucleotides follows the 3-4 hairpin. Not coincidentally, the structure of a bacterial intrinsic terminator includes a hairpin followed by a string of uracil nucleotides; this secondary structure in the 5'-UTR of the *trp* operon is indeed a terminator and is called an attenuator. When cellular levels of tryptophan are high, regions 3 and 4 of the 5'-UTR base pair, producing the attenuator structure; this base pairing causes transcription to be terminated before the *trp* structural genes can be transcribed. The alternative secondary structure of the 5'-UTR is produced by the base pairing of regions 2 and 3 (figure 6.10 b). This base pairing also produces a hairpin, but this hairpin is not followed by a string of uracil nucleotides; so this structure does not function as a terminator. When cellular levels of tryptophan are low, regions 2 and 3 base pair, and transcription of the *trp* structural genes is not terminated. RNA polymerase continues past the 5'-UTR into the coding section of the structural genes, and the enzymes that synthesize tryptophan are produced. Because it prevents the termination of transcription, the 2-3 structure is called an antiterminator. This means the ability of 5'-UTR to fold on itself on one of two structures. When tryptophan is high, the 3-4 structure forms, transcription is terminated within the 5'-UTR, and no additional tryptophan is synthesized. When tryptophan is low, the 2-3 structure forms, transcription continues through the structural genes, and tryptophan is synthesized. The critical question, then, is, why does the 3-4 structure arise when tryptophan is high and the 2-3 structure when tryptophan is low?

To answer this question, we must take a closer look at the nucleotide sequence of the 5'-UTR. At the 5'-end, upstream of region 1, is a ribosome-binding site. Region 1 actually encodes a small protein (see Figure 16.15b). Within the coding sequence for this protein are two UGG codons, which specify the amino acid tryptophan; so tryptophan is required for the translation of this 5'-UTR sequence. The protein encoded by the 5'-UTR has not been isolated and is presumed to be unstable; its only apparent function is to control attenuation. Although it was stated in Chapter 14 that a 5'-UTR is not translated into a protein, the 5'-UTR of operons subject to attenuation are exceptions to this rule.

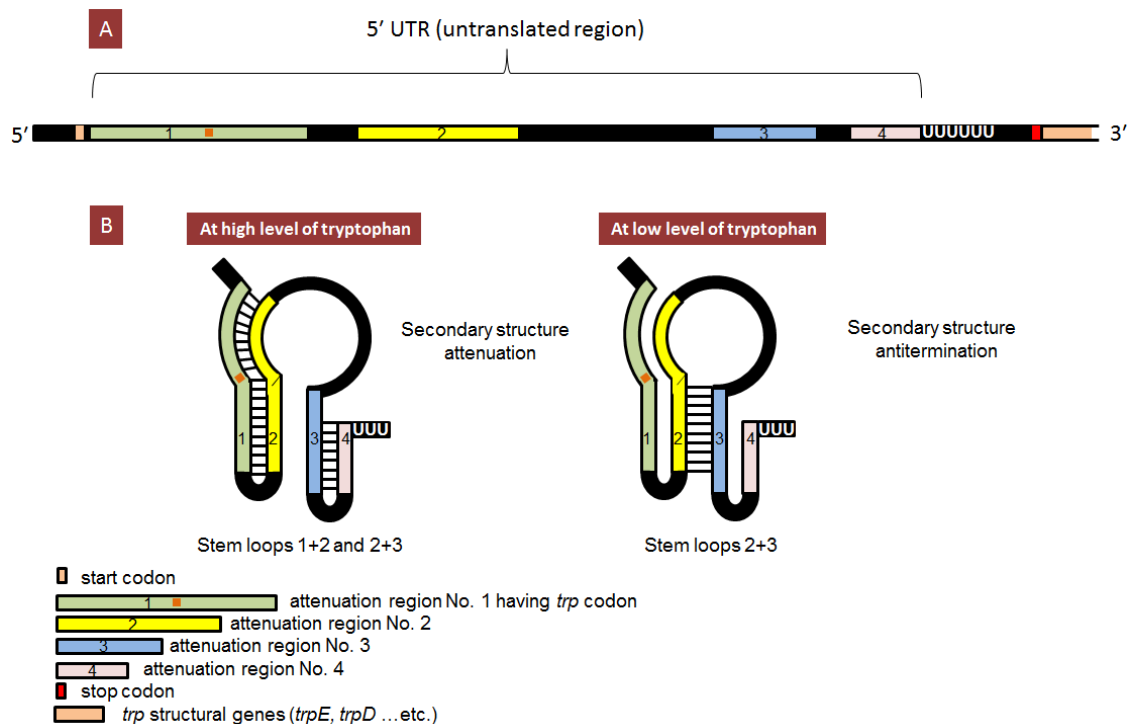


Figure (6.10); the structure of attenuator in *trp* operon. (a) Structure of UTR that located upstream of structural genes of *trp* operon. (b) Formation of two of secondary structures by 5'-UTR region of mRNA that copied from *trp* operon. When the level of tryptophan increases, the region 3 pairs with the region 4, this structure terminates transcription, while when the level of tryptophan is low region 2 pairs with region, this structure don't terminates transcription.

The formation of hairpins in the 5'-UTR of the *trp* operon is controlled by the interplay of transcription and translation that takes place near the 5'-end of the mRNA. Recall that, in prokaryotic cells, transcription and translation are coupled: while transcription is taking place at the 3'-end of the mRNA, translation is initiated at the 5'-end. The precise timing and interaction of these two processes in the 5'-UTR determine whether attenuation occurs.

Trp Operon; The Complete Picture

Transcription When Tryptophan Levels are High

Let's first consider what happens when intracellular levels of tryptophan are high. RNA polymerase begins transcribing the DNA, producing region 1 of the 5'-UTR (figure 6.11, step 1). Following RNA polymerase closely, a ribosome binds to the 5'-UTR (at the Shine-Dalgarno sequence, see chapter five) and begins to translate the coding region. Meanwhile, RNA polymerase is transcribing region 2 (figure 6.11, step 2). Region 2 is complementary to region 1 but, because the ribosome is translating region 1, the nucleotides in regions 1 and 2 cannot base pair. As RNA polymerase begins to transcribe region 3, the ribosome is continuing to translate region 1 (figure 6.11, step 2). When the ribosome reaches the two UGG tryptophan codons, it doesn't slow or stall, because tryptophan is abundant and tRNAs charged with tryptophan are readily available. This point is critical to note: because tryptophan is abundant, translation can keep up with transcription.

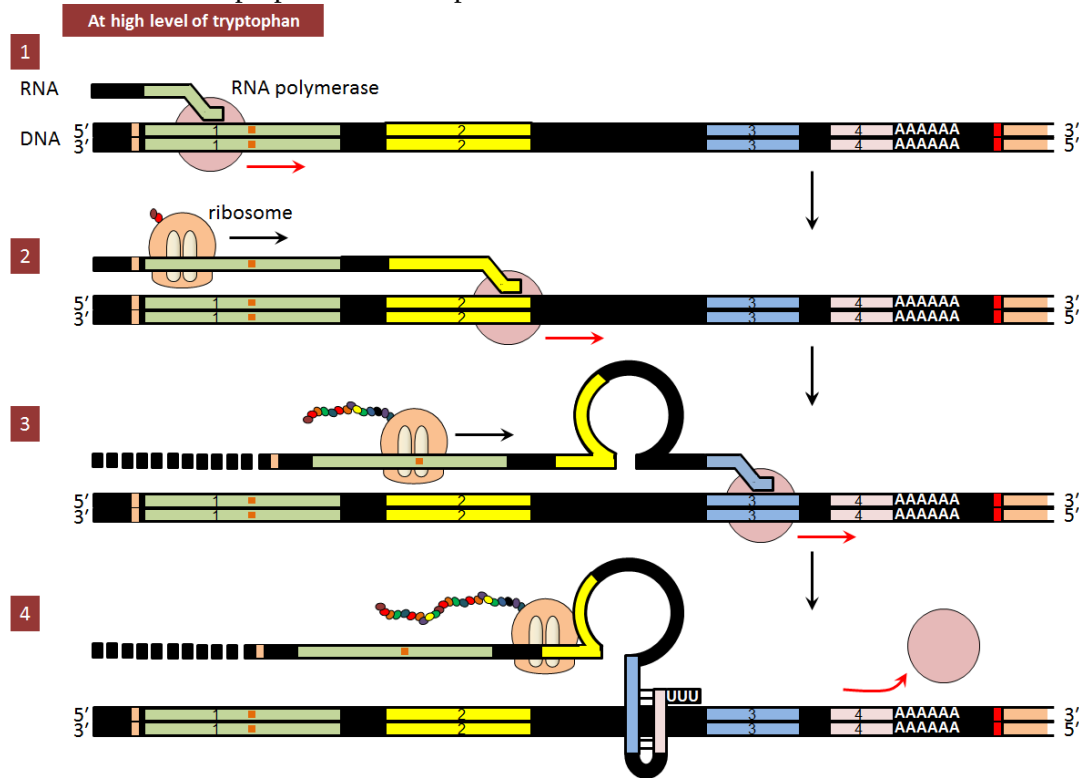


Figure (6.11); attenuation in *trp* operon when the cellular level of tryptophan is high which is summarized in four steps; (1) RNA polymerase begins transcribing DNA, producing region 1 of the 5'-UTR, (2) a ribosome binds to the 5' end of the 5'-UTR and begins to translate region 1, while region 2 is being transcribed, (3) the ribosome translates region 1 while RNA polymerase transcribes region 3, (4) the ribosome does not stall at the *trp* codons, because tryptophan is abundant, (5) the leading edge of the ribosome covers part of region 2, preventing it from pairing with region 3, (6) region 4 is transcribed and pairs with region 3. The pairing of regions 3 and 4 produces the attenuator that terminates transcription.

As it moves past region 1 to the stop codon, the ribosome partly covers region 2; (figure 6.11, step 4); meanwhile, RNA polymerase completes the transcription of region 3. Although regions 2 and 3 are complementary, region 2 is partly covered by the ribosome; so it can't base pair with 3. RNA polymerase continues to move along the DNA, eventually transcribing regions 4 of the 5'-UTR. Region 4 is complementary to region 3, and, because region 3 cannot base pair with region 2, it pairs with region 4. The pairing of regions 3 and 4 produces the attenuator - a hairpin followed by a string of uracil nucleotides - and transcription terminates just beyond region 4. The structural genes are not transcribed, no tryptophan producing enzymes are translated, and no additional tryptophan is synthesized.

Transcription When Tryptophan Levels are Low

What happens when tryptophan levels are low? Once again, RNA polymerase begins transcribing region 1 of the 5'-UTR (figure 6.12, step 1), and the ribosome binds to the 5' end of the 5'-UTR and begins to translate region 1 while RNA polymerase continues transcribing region 2 (figure 6.12, step 2). When the ribosome reaches the UGG tryptophan codons, it stalls (figure 6.12, step 3) because the level of tryptophan is low, and tRNAs charged with tryptophan are scarce or even unavailable. The ribosome sits at the tryptophan codons, awaiting the arrival of a tRNA charged with tryptophan. Stalling of the ribosome does not, however, hinder transcription; RNA polymerase continues to move along the DNA, and transcription gets ahead of translation. Because the ribosome is stalled at the tryptophan codons in region 1, region 2 is not covered by the ribosome when region 3 has been transcribed. Therefore, nucleotides in region 2 and region 3 base pair, forming the 2-3 hairpin (figure 6.12, step 4). This hairpin does not cause termination, and so transcription continues. Because region 3 is already paired with region 2, the 3-4 hairpin (the attenuator) never forms, and so attenuation does not occur. RNA polymerase continues along the DNA, past the 5'-UTR, transcribing all the structural genes into mRNA, which is translated into the enzymes encoded by the *trp* operon. These enzymes then synthesize more tryptophan.

The key factor controlling attenuation is the number of tRNA molecules charged with tryptophan, because their availability is what determines whether the ribosome stalls at the tryptophan codons. A second point concerns the synchronization of transcription and translation, which is critical to attenuation. Synchronization is achieved through a pause site located in region 1 of the 5'-UTR. After initiating transcription, RNA polymerase stops temporarily at this site, which allows time for a ribosome to bind to the 5' end of the mRNA so that translation can closely follow transcription. A third point is that ribosomes do not traverse the convoluted hairpins of the 5'-UTR to translate the structural genes. Ribosomes that attach to the ribosome binding site at the 5' end of the mRNA encounter a stop codon at the end of region 1. Ribosomes translating the structural genes attach to a different ribosome-binding site located near the beginning of the *trpE* gene.

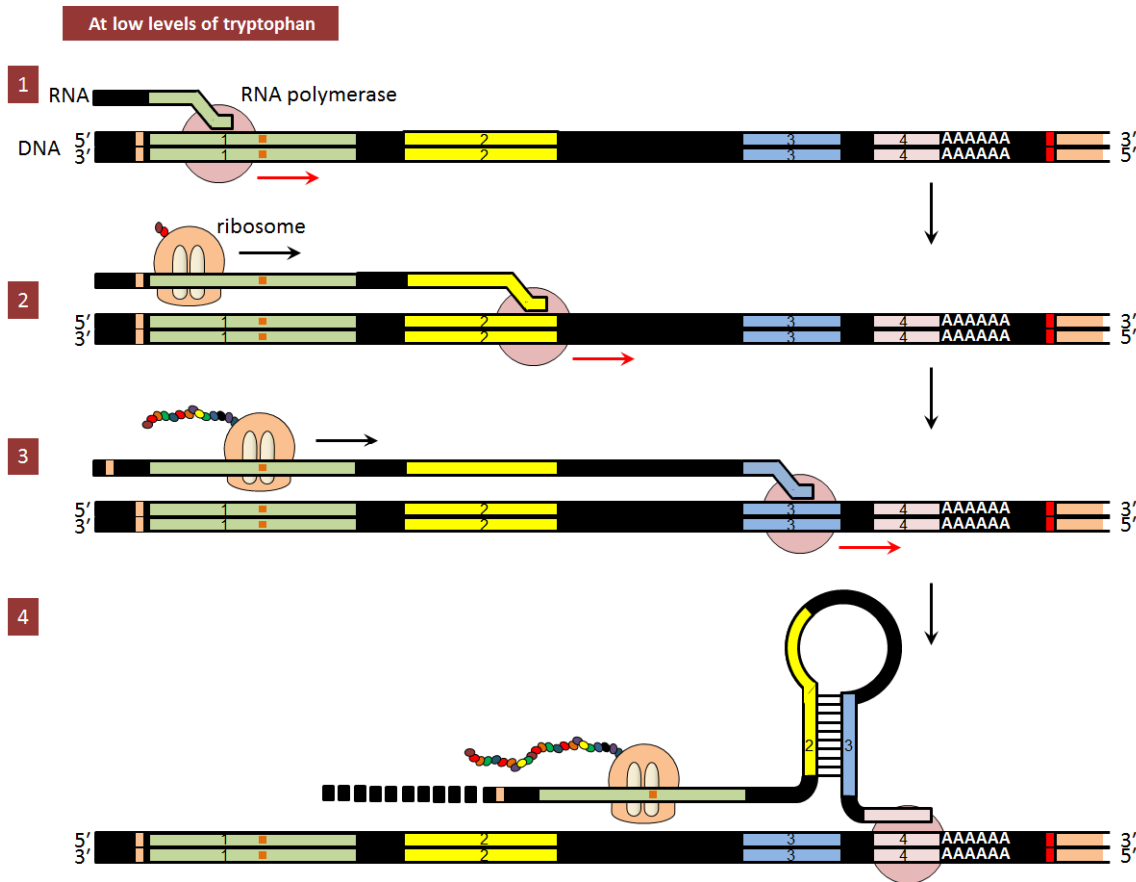


Figure (6.12); attenuation in *trp* operon when the cellular level of tryptophan is low which is summarized in four steps; (1) RNA polymerase begins transcribing the DNA, producing region 1 of the 5'-UTR, (2) a ribosome attaches to the 5' end of the 5'-UTR and begins to translate region 1 while region 2 is being transcribed, (3) the ribosome stalls at the *trp* codons in region 1 because tryptophan is low, (4) because the ribosome is stalled, region 2 is not covered by the ribosome when region 3 is transcribed, (5) when region 3 is transcribed, it cannot pair with region 3, because region 3 is already paired with region 2, (6) when region 4 is transcribed, it cannot pair with region 3, because region 3 is already paired with region 2 the attenuator never forms, and transcription continues.

We have to ask ourselves the question, why does attenuation occur? Why do bacteria need attenuation in the *trp* operon? Shouldn't repression at the operator site prevent transcription from taking place when tryptophan levels in the cell are high? Why does the cell have two types of control? Part of the answer is that repression is never complete; some transcription is initiated even when the *trp* repressor is active; repression reduces transcription only as much as 70-fold. Attenuation can further reduce transcription another 8- to 10-fold; so together the two processes are capable of reducing transcription of the *trp* operon more than 600-fold. Both mechanisms provide *E. coli* with a much finer degree of control over tryptophan synthesis than either could achieve alone. Another reason for the dual control is that attenuation and repression respond to different signals: repression responds to the cellular levels of tryptophan,

whereas attenuation responds to the number of tRNA molecules charged with tryptophan. There may be times when it is advantageous for the cell to be able to respond to these different signals. Finally, the *trp* repressor affects several operons other than the *trp* operon. It's possible that at an earlier stage in the evolution of *E. coli*, the *trp* operon was controlled only by attenuation. The *trp* repressor may have evolved primarily to control the other operons and only incidentally affects the *trp* operon. Attenuation is a complex process to grasp because you must simultaneously visualize how two dynamic processes - transcription and translation - interact, and it's easy to get the two processes confused. Remember that attenuation entails the early termination of transcription, not translation (although events in translation bring about the termination of transcription). Attenuation often causes confusion because we know that transcription must precede translation. We're comfortable with the idea that transcription might affect translation, but it's harder to imagine that the effects of translation could influence transcription, as it does in attenuation. The reality is that transcription and translation are closely coupled in prokaryotic cells, and events in one process can easily affect the other.

Other Means to Regulate Gene Expression in Bacteria at Transcriptional Level

Most of known means of regulation of gene expression in bacteria is the controlling the rate of this process initiation. We have described transcription mechanisms in chapter four, and when recall some information mentioned in that chapter, we found the RNA polymerase holoenzyme is attached to a segment of DNA called promoter. These sequences control the affinity between DNA and RNA polymerase. And there are two recognized regions for attachment, one of them is localized at -35 of promoter, in the sequence TTGACA, and the other is localized at -10 that have the sequence TATAAT. These two hexanucleotides regions are referred as polymerase binding site 1 (BS1), and polymerase binding site 2 (BS2) respectively. The rate of transcription initiation can be changed by the changes that taken place in RNA polymerase (see chapter four). These changes include; subunit replacement, and subunit covalent modification.

Subunit Replacement; When temperature increased from 30 C° to 42 C°, the σ^{70} is temporally replaced with σ^{32} , and since σ^{32} recognizes a promoter differ form that found in the normal conditions. Therefore, when the polymerase binds with this subunit it's misled and consequently it's directed to other promoter to produce other protein (see chapter four).

Subunit Covalent Modification; When T₄ phage infects *E. coli*, RNA polymerase subunits became modified, adenine groups are added to these subunits at its ribose residues (ribose adenylated), reducing the affinity of this enzyme to bacterial promoter (of *E. coli* that exposed to infection by phage), increasing the affinity of this enzyme to the phage promoter. For example, when bacteriophages infect bacteria, some bacteriophages genes are immediately expressed after their entry into bacteria, whereas other group of genes, which called delayed genes, are expressed after 5 to 10 minutes.

The third group of genes, which called late genes, is expressed during the packaging of phage particles and remains through the cell lysis. Early genes have a type of promoter sequences can be recognized by the bacterial sigma factor. Within these genes, the new sigma factor gene, which replaces the bacterial sigma factor on RNA polymerase and makes the enzyme able to recognize the promoters of delayed genes. Within delayed genes, a third factor is synthesized resembles the bacterial sigma factor as well, which replaces the delayed sigma factor and allows the transcription of the late genes.

Eventually, the initiation of RNA synthesis can be regulated by auxiliary regulatory proteins which affect the rate of open promoter complex formation either with a positive or negative manner. These activator proteins are known as activator or repressor proteins respectively. Each promoter has natural affinity for RNA polymerase, and this is determined by the sequence of the promoter itself. The activator proteins increase the affinity of the enzyme, and by this way they increase the rate of transcription. As for the repressor proteins, they have a reversible action. For example, *lac* repressor that was discussed above.

Second; Translational Control on Gene Expression

The gene expression at translational level may be depended on the genetic determination of mRNA half life. The enzymatic degradation of mRNA is from 5' to 3'. This means, the RNA end that synthesized first is the end that degraded first. The average of many mRNA molecules of *E. coli* is only two minutes at 37°C. And the specialized nucleotide sequence at the 5' end may affect on its sensitivity for enzymatic degradation. Moreover, catabolic enzymes are never allowed to approach mRNA molecules when they are being covered at their 5' ends (as in polysomes). Hence, the half life of mRNA molecules may be related with the number of ribosomes that are available in any given moment to translate mRNA molecules. The rate of protein synthesis is varied according to variation of the number of bound ribosomes with mRNA molecules that being translated instead of the variation of the translation rate. For example, in the lactose system of *E. coli*, there are three genes under the control of an operator site, so, the production of these enzymes that located under its control is determined as well. These proteins are produced with different rates indicated to their distance from the 5' end (operator site) of polycistronic mRNA from which these enzymes are produced (these differences are examples to the regulations that taken place in translation). Thus, there is a variation in the polarity within polycistronic mRNA in which the rate of cistron translation is reduced as long as it's located away from the 5' end. It's supposed that ribosomes are bound with different starting points (ribosome binding sites) along polycistronic mRNA at different rates. This is reflected on the level of the synthesized proteins.

There is an exciting type of the translational control related to the control of releasing factor 2 (RF2) expression in *E. coli*. This regulation is achieved by what is known translational frameshifting. The mRNA of RF2 contains the stop codon UGA after the 25th amino acid (Leu). Thus, when enough proteins of RF2 are available, the translation of RF2 mRNA is terminated prematurely. However, if there is no enough

RF2 protein inside the cell, a translation frameshift is taken place, in which U residues in the stop codon UGA are skipped to turn the codon reading as GAC (Asp). This frameshift establishes an open reading frame of the complete translation of RF2 protein.

Third; Post-translational Control (Feedback or End Product Inhibition)

The gene expression is regulated also after the protein synthesis. This is called post-translational control of the gene action. The feedback control is a regulatory mechanism which doesn't affect on the enzyme synthesis. Instead, it inhibits the enzyme activity. In this case, the end product of the biosynthetic pathway (if it is in high concentration) is bound with the first enzyme of the pathway. This binding doesn't occur in the catalytic site, but it modifies the tertiary structure of the enzyme. Hence, it inhibits the catalytic site. This allosteric transition inhibits the enzymatic activity of the protein molecules and prevents the overproduction of the end products and their intermediate metabolic products. There is an obvious example describes this phenomenon, which is related with the synthesis of the amino acid isoleucine (the end product of the five steps of threonine conversion) into the medium that bacteria grow on which results in a direct termination of threonine – isoleucine pathway. When the added isoleucine is available, the cell prefers to use the end product of the extracellular amino acid isoleucine and the synthesis of its own isoleucine is terminated. Moreover, the production of the five enzymes doesn't interfere, but the action of the enzyme that is responsible on the removing of amine from threonine into α -ketobutyrate is inhibited by the end product which is isoleucine (figure 6.13).

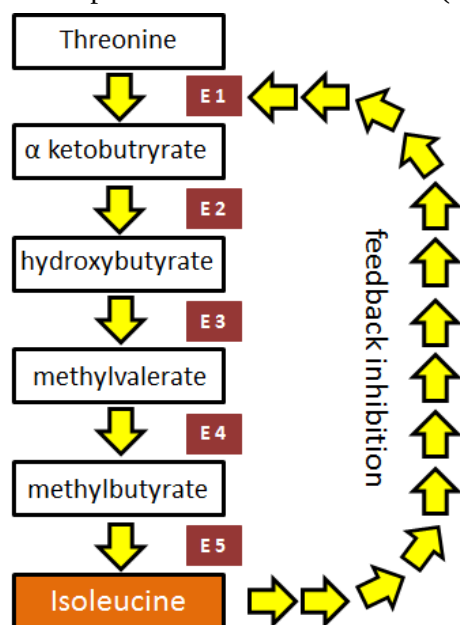


Figure (6.13); feedback inhibition in *E. coli*. The letter E1, E2 Etc. refers to the enzymes that convert the amino acids from one form to another form. When the concentration of the amino acid isoleucine is increased, the first enzyme (E1) is inhibited.

Increasing the concentration of the last amino acid leads to feedback inhibition

Regulation of Gene Expression in Bacteriophages

Some bacteria harbor viruses that can remain in a dormant state within the bacterial chromosome or they can replicate within bacteria to lead consequently to the lysis and killing of the bacterial host. Some *E. coli* types harbor temperate viruses, such as lambda (λ) phage. When this phage infects its sensitive *E. coli*, it injects its own double stranded linear chromosome that containing of 45000 bps (figure 6.14). and depending on the nutritional state of the cell, λ phage either integrates into the bacterial chromosome (lysogenic pathway) and remains dormant until its activated (see below), or it synthesizes a large number of copies that are possibly reach 100 copy of complete viruses that packaged within special protein structures. At this point, λ phage dissolves the host (lytic pathway). And it can be for the newly synthesized viral particles to infect other sensitive hosts. When this phage enters in its dormant state, it will stay in this state until it's activated by the exposure of its bacterial host to DNA degrading factors. And as a response for such damaging factor, the dormant phages become "induced" and start to transcript and translate these genes that localized on their own chromosome that are necessary to excise them out of the host chromosome, in addition to other genes that have a major role in DNA replication, the genes that encode to proteins that synthesize the viral protein envelopes, and genes that encode for hydrolytic enzymes. This transition that is taken place between dormancy state (or prophage state) into a lytic state is clearly understood on the molecular level, and it will be discussed with some details here. After this phage infects bacteria, a decision will be made whether it replicates and produces new phages to lead to the lysis of the cell, or it integrates into the bacterial chromosome and become genetically repressed (the lysogenic state). A decision is made between these two states of λ phage by two repressors, which are *ci* and *cro*, which compete between each other on the operator region of λ phage DNA (figure 6.14).

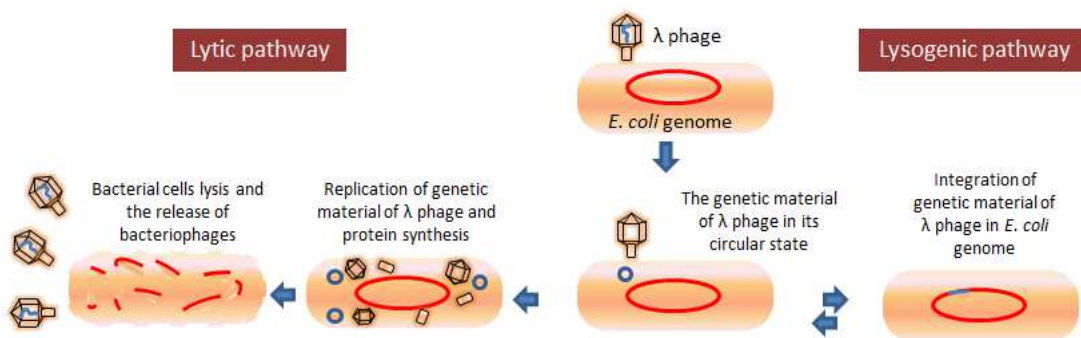


Figure (6.14); the genetic switch of λ phage. The two opposite arrows refer to the reversal nature of the lysogenic pathway, or the ability of λ phage DNA releasing after its integration into *E. coli* genome once the bacterium are exposed to a suitable environmental conditions for this to happen such as UV light.

The transition activity of λ phage is localized on a region that contains 80 bps in DNA double helix which called the right operator (OR) (figure 6.15). The right operator is flanked from its left side with a structural gene of λ phage repressor, and

from its right side with a structural gene of other regulatory protein called *cro*. When λ phage in its prophage state, *i.e.*, when its integrated into the host chromosome, the repressor gene is expressed, which is the only λ phage gene that expressed on the dormant state of the phage, which inhibits the expression of all other genes except it.

When λ phage is exposed to the lytic growth, the repressor gene is never expressed on, but the *cro* gene is, in addition to many other genes in λ phage. This means, when the repressor gene is turned on, the *cro* gene is turned off and vice versa. This decision between the transcription of repressor gene and the transcription of *cro* gene is an example of the molecular switch between two different approaches.

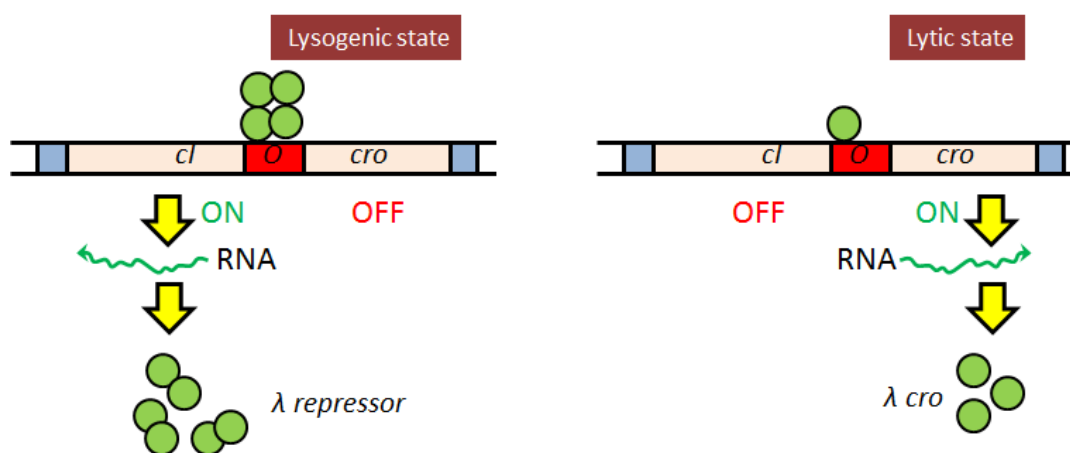


Figure (6.15); the molecular explanation of the genetic switch of λ phage. In the first state (the lytic state) the repressor protein λ *cro* is synthesized, in the second state (lysogenic state) the λ repressor protein is synthesized. The letter *O* refers to the operator.

There are crucial differences in the gene regulation between prokaryotes and eukaryotes. In addition to the transcription, the eukaryotic cells employ various mechanisms to regulate gene expression. The nuclear envelope of the eukaryotic cells separates the transcription of the gene from its translation physically, as long as the ribosomes exist only in the cytoplasm. In eukaryotic cells, many additional steps are involved, especially in the processing of RNA, in the expression of the eukaryotic genes compared with these genes that exist in prokaryotes. These steps provide additional locations for the regulatory effects that cannot be occurred in the prokaryotes. The processing steps of RNA in eukaryotes involves the capping of 5' end of the primary transcript, the addition of polyadenylated tail to the 3' end of the same molecule, in addition to the excision of the intron regions to generate exons ligated with each other in the mature RNA (mRNA) molecule.

Regulation of Gene Expression in Eukaryotes

The study of gene expression in the eukaryotic organisms is considered so complicated and less understood on the molecular level compared with prokaryotes because of many reasons such as the enormous number of the eukaryotic genes

compared with their prokaryotic counterparts. Therefore, we will give few examples about what is taking place in the regulation of gene expression in eukaryotes. But despite of their few numbers, they give us a realistic picture about the pattern that eukaryotic cells pursue in the regulation of gene expression.

Mammals have a genetic material more than 8000 folds compared with these found in *E. coli*. Many of these genetic materials are involved in the regulation of gene expression during tissues differentiation and during the variability of biological activities in the multicellular organism to ensure the possibility of organism's response to the complicated environmental challenges. Simply, there are two types of gene regulation; the positive regulation and negative regulation. When the expression of the genetic information is increased quantitatively when certain regulatory factor is exist, it can be said that the regulation is positive, whereas when the gene expression of the genetic material is decreased when certain regulatory factor is exist, it can be said that the regulation is negative. The factor or molecule that intermediate the negative regulation is called the negative regulator, whereas these that intermediate the positive regulation are called the positive regulators. However, the double negative regulation has the effect of the positive regulator in its action. Consequently, the regulator that inhibits the function of the negative regulator is considered as positive regulation.

The Levels of Gene Regulation in Eukaryotes

Many genes are regulated at several points along the pathway of the genetic information flow from the genotype (that is usually represented by DNA) into the phenotype (that is usually represented by protein) (figure 6.16).

First, regulation may be through the alteration of gene structure. Modifications to DNA or its packaging may influence which sequences are available for transcription or the rate at which sequences are transcribed. DNA methylation and changes in chromatin are two processes that play a pivotal role in gene regulation. A second point at which a gene can be regulated is at the level of transcription. For the sake of cellular economy, it makes sense to limit protein production early in the transfer of information from DNA to protein, and transcription is an important point of gene regulation in both bacterial and eukaryotic cells. A third potential point of gene regulation is mRNA processing. Eukaryotic mRNA is extensively modified before it is translated; a 5' cap is added, the 3' end is cleaved and polyadenylated, and introns are removed (see chapter four). These modifications determine the stability of the mRNA, whether mRNA can be translated, the rate of translation, and the amino acid sequence of the protein produced. There is growing evidence that a number of regulatory mechanisms in eukaryotic cells operate at the level of mRNA processing. In addition to the processing of RNA, sometimes the regulation of RNA stability plays an important role in the regulation of gene expression. The quantity of synthesized protein doesn't depend on the quantity of synthesized mRNA, but it depends on the rate at which the mRNA is degraded. Thus, the stability of RNA plays an important role in gene expression. A fourth point for the control of gene expression is the regulation of RNA stability. The amount of protein produced depends not only on the amount of mRNA synthesized, but also on the rate at

which the mRNA is degraded; so RNA stability plays an important role in gene expression.

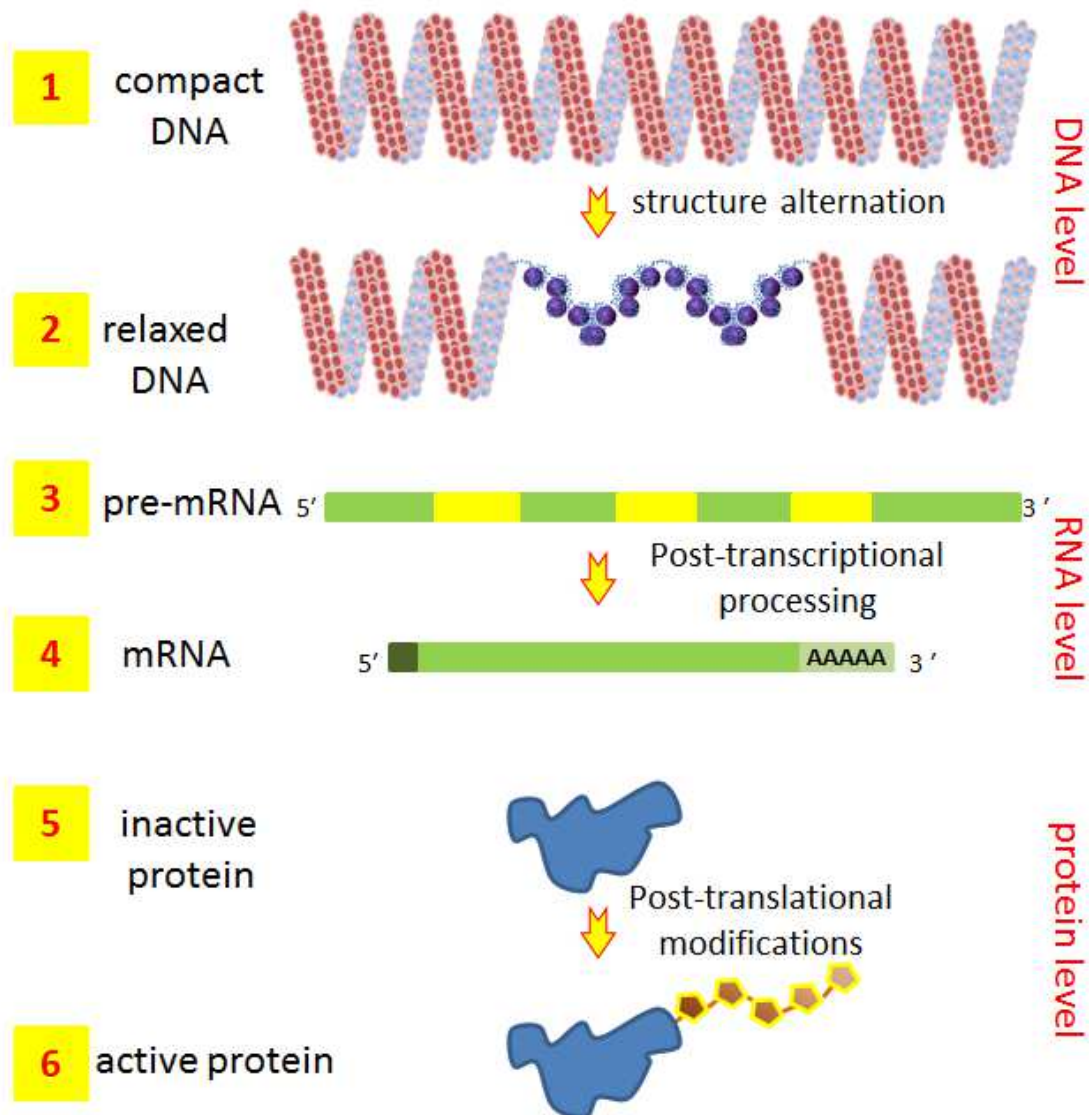


Figure (6.16); the regulation of gene expression at several levels.

The fifth point of gene regulation is at the level of translation, a complex process requiring a large number of enzymes, protein factors, and RNA molecules (see chapter five). All of these factors, as well as the availability of amino acids and sequences in mRNA, influence the rate at which proteins are produced and therefore provide points at which gene expression may be controlled. Finally, in the fifth point, many proteins are modified after translation (see chapter five), and these modifications affect whether

the proteins become active; so genes can be regulated through processes that affect posttranscriptional modification. Gene expression may be affected by regulatory activities at any or all of these points.

First; regulation of Gene Expression at Genomic Level

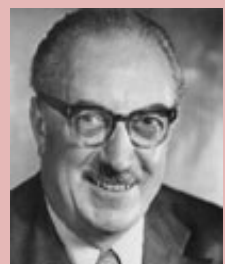
In eukaryotes, there are apparently certain types of genes that are copied in more or less continuity on the level of genome, and their activities are repressed in a rare case. The pattern of the differential expression that present on the level of transcription or translation may be existed on the genomic level (such as DNA). Before we explore some examples that are related with this level of regulation we have to know very important terms which are Euchromatin and heterochromatin, the first one is transcriptionally active and contains genes in an active state, whereas, the heterochromatin is in the opposite. Usually, euchromatin replicates before heterochromatin. Moreover, heterochromatin stains non-homogeneously under microscope, hence its name "hetero", while euchromatin stains homogeneously. There are two types of heterochromatin: constitutive and facultative. Constitutive heterochromatin is always condensed and thus inactive. It is found in the regions near the chromosomal centromere and at chromosomal ends (telomeres). Facultative heterochromatin is at times condensed, but at other times it is actively transcribed and, thus, uncondensed and appears as euchromatin. Of the two members of the X chromosome pair in mammalian females, one X chromosome is almost completely inactive transcriptionally and is heterochromatic. However, the heterochromatic X chromosome decondenses during gametogenesis and becomes transcriptionally active during early embryogenesis— thus, it is facultative heterochromatin.

The intensity of chromatin condensation is varied during the cell cycle. In interphase (non-dividing cells), most of chromatin is called euchromatin, which spreads throughout the nucleus. During this stage of the cell cycle, most genes are transcribed and DNA is replicated in preparation for cell division. It appears that most of euchromatin in interphase cells assembles 30-nm fibers, thus, chromatin connected with gene expression. However, mammalian chromosome has bifacial symmetry in metaphase, with homologous sister chromatids connected at centromere. Each sister chromatid consists of single stranded DNA molecule, during interphase, the DNA packaging is less condensed than the condensed chromosome in interphase. Therefore, metaphase chromosomes are transcriptionally inactive. Hence, when we review some examples about the control of gene expression at genome level, it will be clarified to us how both euchromatin and heterochromatin alter the expression of the gene.

- (a) Genetic shutdown expression pattern. There are many examples to clarify this pattern; the most important one may be;
 - (1) During mitotic phase of the cell cycle, chromatin is highly condensed to form chromosomes, and transcriptional activity of all genes is suspended.
 - (2) Sperm cells clearly contain a complete genetic endowment but no transcription occurs until the sperm nucleus is activated within the egg cytoplasm.

- (3) X- Inactivation; this is correlated with the condensation and complete shutting down of gene expression of one of the two X-chromosomes in cells of female mammals. Females possess two X chromosomes, whereas males have one X chromosome plus a much shorter Y chromosome. Consequently, females have two copies of most genes carried on the X chromosomes, whereas males only have one. In mammals, one of X chromosomes is turned off in each female cell. The mechanism of X-chromosome inactivation in mammals involves the interaction of noncoding RNA molecule which inhibits only one X-chromosome. In mammals, X-inactivation is controlled by methylation of the *Xist* gene, which is itself located on the X-chromosome. The *Xist* gene of the X-chromosome that remains active is inactivated by methylation.

Once established, this methylation pattern is inherited at cell division; thus, the same one of the pair of X-chromosomes will remain active in the daughter cells. The expression of the *Xist* gene is in turn regulated by the antisense RNA, *Tsix*, which is transcribed from the *Xist* locus, but in the reverse direction. However, the timing of expression of *Tsix* RNA varies among different mammals and although *Tsix* appears to be involved in choosing which X chromosome to inactivate, its role is presently unclear. Expression of the *Xist* gene causes the inactivation of the X chromosome that carries it. A long non-translated RNA is transcribed from the *Xist* gene. This *Xist* RNA coats the inactive X-chromosome. Starting from the *Xist* gene and proceeding along the X chromosome in both directions (figure 6.17), the DNA is converted into heterochromatin, a condensed form of DNA that cannot be transcribed (see chapter two).



Murray Barr

Highly condensed X chromosomes are visible in the cells of female mammals and are called Barr bodies, after Murray Barr, who discovered them in 1948. The presence or absence of Barr bodies has sometimes been used to check whether female Olympic athletes are genetic females.

- (b) Constitutive expression pattern; we restricted in clarification of this pattern with only one example about this pattern of gene expression. This example correlated with "house-keeping genes", in which the chromatin locations that have such genes are characterized with their continuous non-condensation. Despite of the fact that these genes are transcribed with a low rate, they are expressed continuously (see chapter two). There are some other examples but they are out of the scope of this book.
- (c) Specific expression pattern; many genes are expressed in only certain tissues and there are many examples show this fact. The frog *Xenopus* provides a good example of regulation of 5S genes. *Xenopus borealis* possesses 19,000 copies of the oocyte specific 5S rRNA genes, and these genes are active only in the oocyte and in no other cell.

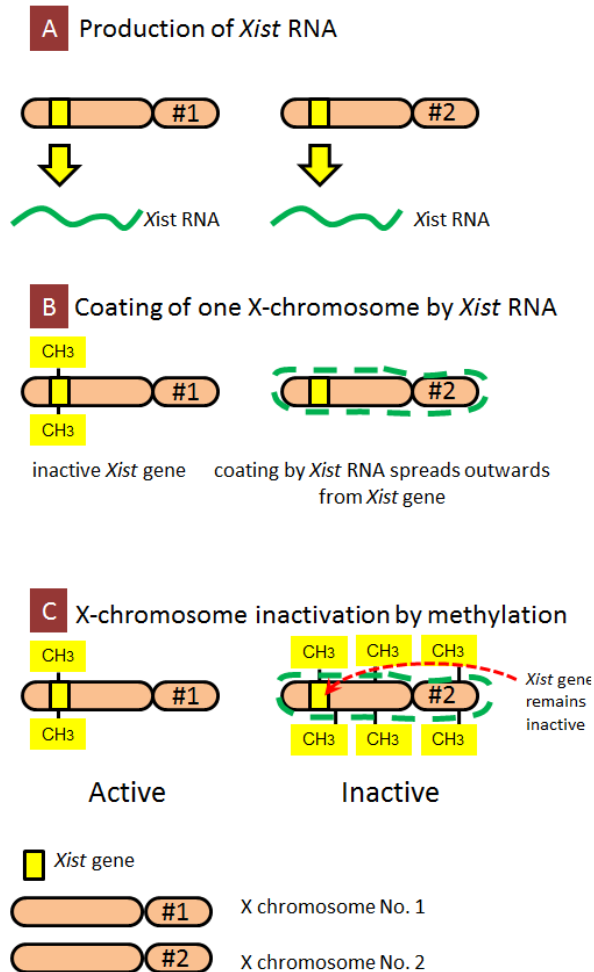


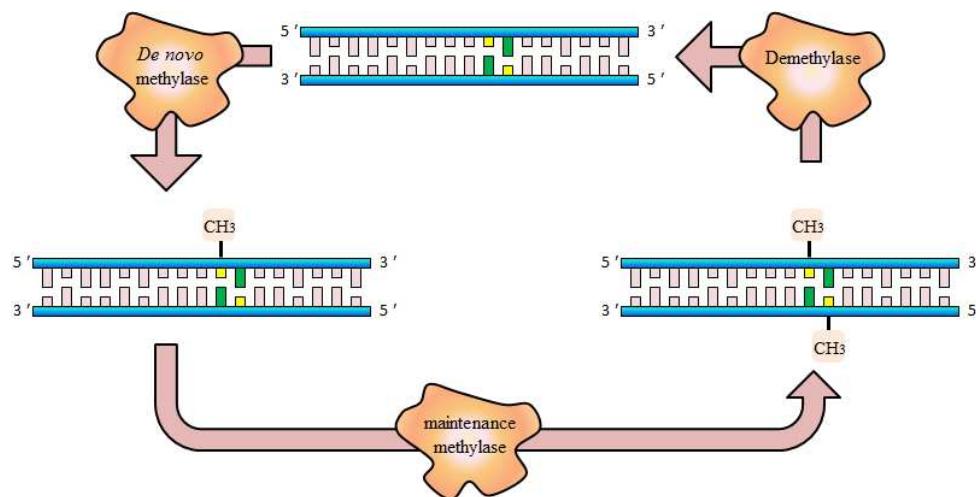
Figure (6.17); X-inactivation which involves the *Xist* Gene and *Xist* RNA. A) Originally, both X chromosomes transcribe *Xist* RNA from the *Xist* gene. B) The X chromosome that will remain active is methylated in the *Xist* gene region, which inactivates the *Xist* gene. The *Xist* RNA coats the other X chromosomes and inactivates it. C) The inactive X chromosome is almost entirely methylated, except for the *Xist* gene (together with a few aberrant loci that are exempted from X inactivation and are not shown here). This causes the X chromosome to transform largely into heterochromatin. Only the *Xist* gene itself remains active.

- (d) Non-genetic elements expression pattern; some DNA is never transcribed in any cell. Some eukaryotic DNA is comprised of tandemly repeated short sequences that are concentrated in heterochromatin, such as centromere of the chromosomes and Y chromosome. Much of this DNA is never transcribed in any cell. That's mean they are never genetically express about themselves, therefore they are called sometimes non-genetic elements. Add to that, there are certain sequences occur between genes are called spacer sequences. Such spacer sequences are often taken to be un-transcribed. The very large size of the genomes of the higher eukaryotes certainly indicates that much of the DNA is repetitious and probably not utilized as coding or regulatory sequence. There is another example provided by pseudogenes (see chapter two).
- (e) Gene rearrangement expression pattern; the splicing phenomenon (that is described in chapter four) causes genetic rearrangement. This mechanism is taking place during immunoglobulin Ig genes expression. The splicing here differs from this process that is described in chapter two since the later is taking place for RNA molecules that never leads to DNA rearrangement.

Consequently, this mechanism may give a reasonable explanation about the abundance of variation on the gene level in some mammalian genes which reflected into highly variable regions in protein level. This, in turn, usually reflected on the function of the resulting antibodies.

Second; Regulation of Gene Expression at Transcriptional Level

- (a) Modification of DNA sequences (DNA methylation); methylation exists in both prokaryotic and eukaryotic organisms, although the purpose is quite different. Prokaryotes use methylation to distinguish the newly synthesized DNA. While in eukaryotes, newly synthesized DNA is recognized by other means that are still unclear. Nonetheless, many eukaryotes do methylate their DNA as a marker for regulating gene expression. In these multi-cellular organisms DNA methylation is often used as a marker for genes whose expression is involved in tissue differentiation. The recognition sequences are extremely short; typically CG for animals. There are two types of methylases. Maintenance methylases add methyl groups to newly made DNA at locations opposite methyl groups on the old, parental DNA strand. This ensures that the pattern of methylation is inherited during chromosome division. Changing the pattern of methylation involves *de novo* methylases to add new methyl groups and demethylases to remove methyl groups (figure 6.18). In this methylation, the change that taken place in the chromatin structure that connected with transcription is represented by the methylation of C residues, which yields 5-methylcytosine. The methylated DNA is widely associated with the repression of transcription, in which the transcriptionally active DNA is non-methylated in these organisms.



1. **Figure (6.18);** control of methylation of DNA in eukaryotes. Three enzymes control methylation of DNA. De novo methylase adds methyl groups to non-methylated CG-islands. Maintenance methylase adds a second methyl group on the opposite strand of hemimethylated sites. Demethylase removes methyl groups.

Usually, DNA methylation is most common on cytosine bases adjacent to guanine nucleotides on the same strand (CpG); so two methylated cytosines sit diagonally across from each other on opposing strands: DNA regions with many CpG sequences are called **CpG islands** and are commonly found near transcription start sites. While genes are not being transcribed, these CpG islands are often methylated, but the methyl groups are removed before the initiation of transcription.

Methylation in eukaryotes silence gene expression. The scientific evidence for this are **Housekeeping genes (see chapter two)**, which are expressed in all tissues, possess non-methylated CG islands. In contrast, the CG-islands of tissue specific genes are only non-methylated in those particular tissues where the genes are expressed. The maintenance of the pattern of methylation therefore makes sure that the pattern of gene expression stays constant among the cells of a particular tissue. The modern researches refer to the binding between DNA methylation and deacetylation of histones both repress transcription. Certain proteins that bind firmly with methylated G and C sequences form complexes with other proteins which act as removers for histone deacetylases, which remove the acetyl groups from histone tails in order to stabilize nucleosomes structure and repress transcription.

- (b) Histone modification. The histone component of chromatin is exposed to three types of modifications after the synthesis which either have direct or indirect effect on gene regulation in eukaryotes. 1) Histone methylation affect on H3 and H4 that involved in irreversible methylation of the few lysine residues, this modifies the hydrophobic nature of side chains of these histones; 2) histone phosphorylation involves histone H1. Phosphorylation affects serines and threonines, changing them from a state of neutral charge to one of negative charge and is a reversible reaction. The state of phosphorylation of H1 protein varies through the eukaryotic cell cycle, and after H1 phosphorylation, chromatin becomes much more strongly condensed, as it does in mitotic chromosomes. Evidently, activation of the histone kinase enzyme that is responsible for H1 phosphorylation may be the first step in the chain of events that leads to eventual chromatin condensation prior to mitotic cell division. 3) One factor affecting chromatin structure is acetylation, the addition of acetyl groups (CH₃CO) to histone proteins. Histones in the octamer core of the nucleosome have two domains: (1) a globular domain that associates with other histones and the DNA and (2) a positively charged tail domain that probably interacts with the negatively charged phosphates on the backbone of DNA (figure 6.19).

Acetyl groups are added to histone proteins by acetyltransferase enzymes; the acetyl groups destabilize the nucleosomes structure, perhaps by neutralizing the positive charges on the histone tails and allowing the DNA to separate from the histones. Other enzymes called deacetylases strip acetyl groups from histones and restore chromatin repression. Certain transcription factors and other proteins that regulate transcription either have acetyltransferase activity or attract acetyltransferases to the DNA. Some transcription factors and other regulatory proteins are known to alter chromatin structure without acetylating histone proteins. These chromatin-remodeling complexes bind directly to particular sites on DNA and reposition the nucleosomes, allowing transcription factors to bind to promoters and initiate transcription.

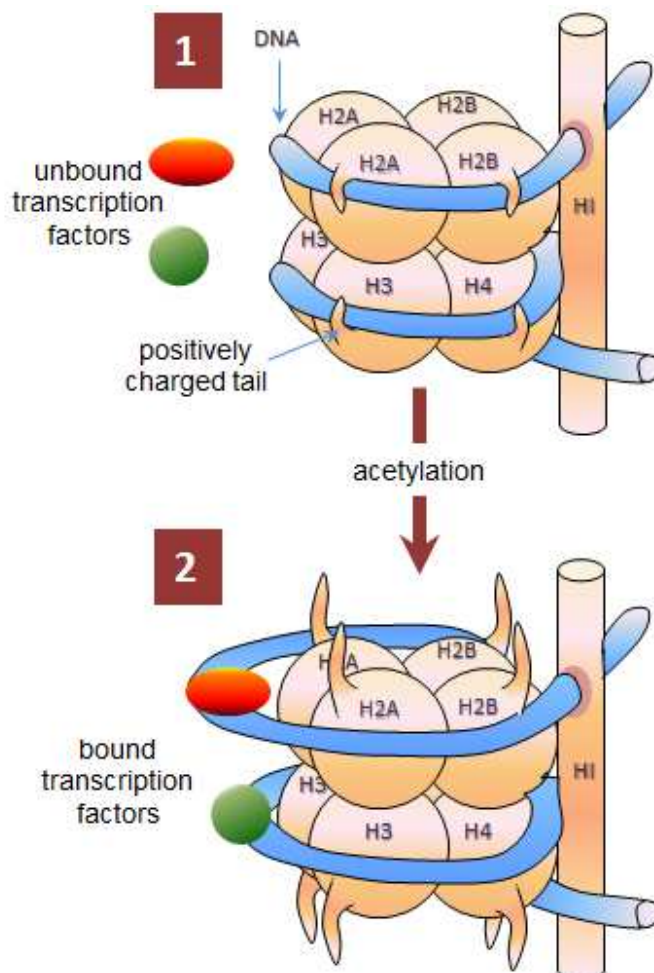


Figure (6.19); the role of acetylation of histone proteins in the modification of chromatin structure and in allowing for some transcription factors to bind with DNA, (1) Positively charged tails of nucleosomal histone proteins probably interact with the negatively charged phosphates of DNA. (2) Acetylation of the tails weakens their interaction with DNA and may permit some transcription factors to bind to DNA.

Third; Regulation of Gene Expression at Post-Transcriptional Level

Steps that come between transcription and translation are described as post-transcriptional modifications. That's mean, any modification for these steps leads to a modification in the final product shape. This, can be clarified with an example called differential splicing. The removing of introns and the attachment of exons together for the rest of exons during heteronuclear RNA (hnRNA) processing must be absolutely precise. This is partially engineered by a distinct group of nuclear particles which are Sn RNPs that containing U1, U2 , U3 , U4 , U5 and U6sn RNAs (see chapter four). For example, differential splicing is used in different lymphocyte cells to produce different proteins from the same hn RNA molecule. The two types of mRNA molecules are produced by part of an intron being omitted from one of the mRNAs but included in the exon splice used to produce the other mRNA. This allows production of two distinct proteins both immunoglobulins (Ig), but *one* with a long strand of hydrophobic amino acids at its carboxyl terminus, and the *other* with only a short length of relatively hydrophilic amino acids. The Ig molecule with the long hydrophobic peptide is membrane bounded within a lymphocyte, whilst the molecule with the terminal hydrophilic peptide is secreted from the cell. This change in splicing takes place within the life of a single lymphocyte cell and clearly explains the following observation. Immature lymphocytes retain antibody and simply insert the Ig molecules into their plasma membranes, whereas following stimulation with antigen, the same lymphocyte becomes secretory, releasing antibody molecules into circulation. From this example, the central role that differential splicing plays is clarified in changing the functions of many cells.

Fourth; Regulation of Gene Expression at Translational Level

In bacteria, most mRNA molecules are translated about the same number of times with only fairly small variation from gene to gene. In eukaryotes translational regulation occurs in which a mRNA molecule is not translated at all until a signal is received. We are limited in only one mechanism which is the extension of lifetime of the mRNA. An important example of translational regulation is that of informosomes or masked mRNA. Unfertilized eggs are biologically static, but shortly after fertilization many new proteins must be synthesized, for example, the proteins of the mitotic apparatus, the cell membranes, histones for nucleosome formation as well as others. Unfertilized sea urchin eggs store large quantities of mRNA for many months in the form of mRNA protein particles (= masked mRNA) made during formation of the egg. This mRNA is translationally inactive, but within minutes after fertilization, translation of these molecules begins. Here, the timing of translation is regulated.

Fifth; Regulation of Gene Expression at Post-Translational Level

Some proteins are altered after synthesis, usually this is done by three-dimensional structure modification, chemical modification of protein's side chains, or by backbone modification (see chapter five). Any change in these modifications that

taken place in the proteins after their modifications have dangerous clinical consequences.

Chapter Six Appendix

Improving Tomato Quality by Preventing the Expression of polygalacturonase

Fruit ripening involves several biochemical and physiological changes that are primary factors influencing quality attributes, such as color, flavor and texture of the product. Tissue softening of fruits during ripening is the result of solubilization of the cell wall by a group of enzymes. One of the key enzymes is polygalacturonase (PG), functioning in the breakdown of pectin, a polymer of galacturonic acids that forms part of the structural support in cell wall. The texture of tomato fruit is a major quality consideration of commercial importance in both fresh market and commercial processing. Tomatoes sold in markets are picked from the field when they are green, stored under low temperature, and gassed with ethylene to trigger fruit coloration and ripening. Recombinant DNA technology is now capable of engineering tomatoes that soften slowly and can be left to ripe on the vine, with full development of color and flavor by inhibition of the expression of the PG gene by antisense RNA. The basic idea of this technique is to introduce into the plant an RNA molecule that is complementary to the mRNA of the PG gene. By this way, scientists are able in antisense technology to introduce a fragment of RNA that is complementary in its sequence to PG mRNA. This introduced RNA fragment can bind to mRNA to prevent its translation. Consequently, it prevents the production of this enzyme (figure 6.20). Since the expression of PG gene is inhibited the plant losses its ability to produce PG enzyme that causes the unfavorable ripening in the walls of tomato fruits.

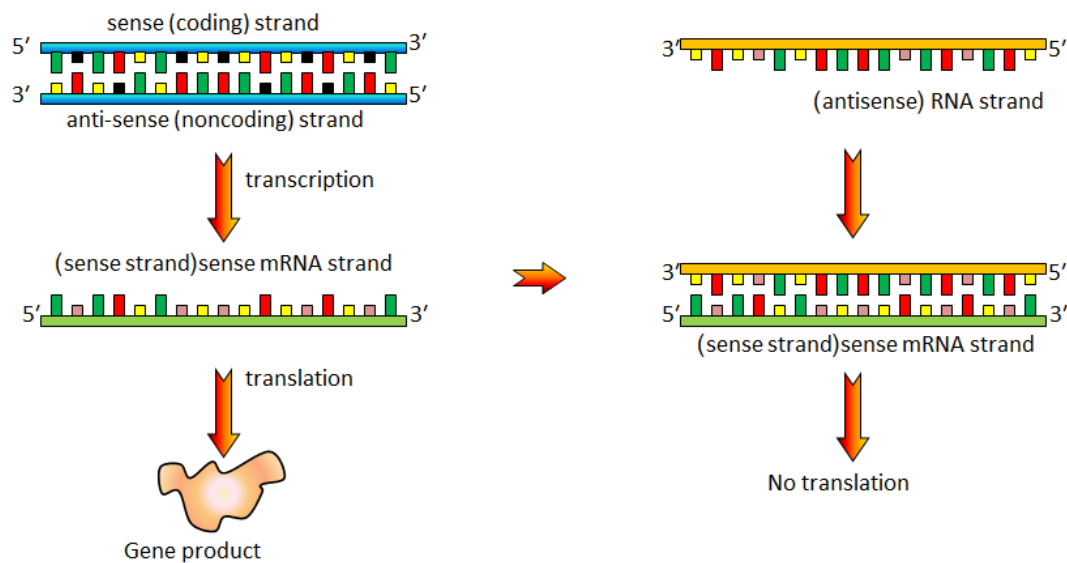


Figure (6.20); the role of inhibitory mechanism of antisense RNA in preventing the synthesis of PG enzyme.

Chapter Six Revision Questions

Q1- Justify the following:

1. When both glucose and lactose available in the medium of *E. coli*, glucose is exhausted first?
2. *Lac* operon is not working when glucose available and lactose is unavailable in the medium of *E. coli*?
3. The regulator protein of *lac* operon CAP is described as apoactivator instead of coactivator?
4. The hypothesis one gene one enzyme is incorrect?
5. Antisense RNA technology is used to prevent polygalacturonase synthesis in tomato?
6. Barr bodies are exclusively available in females of human beings?
7. The connection that is taking place between the methylation of DNA and the deacetylation of histones both repress transcription?
8. It's possible for mRNA to life – as in sea urchin – for a period estimated more than two months?

Q2- choose the correct answer:

1. The typ repressor is a regulatory protein takes a motif called
a- helix-turn helix b- helix-loop helix c- helix-up helix d- helix-to helix e- helix-and helix
2. I the pathway in which phage remains in their dormant stage.
a- lysogenic pathway b- temperate pathway c- lytic pathway d- dormant pathway e- protein synthesis pathway
3. The operon is a group of genes under the control of one
a- promoter b- cistrone c- protein d- polypeptide e- operator
4. The effector molecular (tryptophan) in the operon system is sometimes called
a. antirepressor b. pararepressor c. corepressor d. (a) and (c) e. non
5. *Lac* operon express in very low level in case of
a. glucose and lactose presence b. glucose presence & lactose absence c. glucose absence & lactose presence d. glucose and lactose absence e. none
6. Depending on, either λ phage integrates into the bacterial chromosome and remain dormant, or it synthesizes large number of copies to lyse the bacterial cell.
a. temperature b. nutrition c. DNA sequence d. species e. behavior
7. The, is the state at which all phagal genes are repressed except only one gene which is *ci*, that prevents other phagal genes working.
a. temperate stage b. lytic state c. prophage state d. none
8. the ability of antibodies of the body to bind with very large number of antigens to the phenomenon in gene expression.
a. genetic breakdown b. DNA methylation c. histone acetylation d. DNA rearrangement
9. Attenuation is a phenomenon that taken place when trp level is high, in which structure is formed, and transcription is terminated within 5'-UTR region.
a. 1-2 structure b. 2-3 structure c. 3-4 structure d. 2-4 structure
10. The of T₄ phage consist of promoter sequences that can be recognized by the bacterial σ factor.
a. early genes b. delayed genes c. late genes d. overlapping genes
11. The gene expression of *Xist* gene of the female X chromosome by antisense – RNA, which is called....., that transcribed from the same gene, but in opposite direction.
a. *istX* b. *Xsti* c. *tisX* d. *tsiX* e. *Xtis*
12. The binds largely with the repression of transcription.
a. DNA acetylation b. DNA glycosylation c. DNA phosphorylation d. DNA methylation
13. the chromatin that is sometimes condensate and appears sometimes in a non-condensated state is called
a. euchromatin b. constitutive heterochromatin c. facultative heterochromatin d. only “b” & “c”

Q3- Define the following:

lac repressor, constitutive genes, antisense RNA technology, temperate phages, translational frameshifting, spacer sequences

Q4- Connect terms exist in right with its suitable counterparts in left

genetic breakdown	Ovalbumin gene
constitutive expression	sn RNA
specific expression	mRNA-protein particles
non-genetic elements	Housekeeping genes
gene rearrangement	X- inactivation
DNA methylation	GpC islands
Differential splicing	Telomeres
informosomes	Luxury genes

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Bell C.** and Lewis M. 2001. The *lac* repressor: A second generation of structural and functional studies *Curr. Opin. Struct. Biol.* 11: 19-25.
- Berk A.**, Zipursky S., Baltimore D., Darnell J., and Lodish H. Molecular Biology. Fourth edition. Paul Matuataira, USA, 1998.
- Bird A.** and Wolffe A. 1999. Methylation-induced repression: Belts, braces, and chromatin *Cell* 99: 451-454.
- Busby S.** and Ebright R. 1999. Transcription activation by catabolite activator protein (CAP) *J. Mol. Biol.* 293: 199-213.
- Cairns B.** 1998. Chromatin remodeling machines: Similar motors, ulterior motives *Trends Biochem. Sci.* 23: 20-25.
- Calladine, C. R.** Understanding DNA: The Molecule and How It Works. New York: Academic Press, 1997.
- Clark D.** Molecular biology: understanding the genetics revolution. Elsevier, 2010.
- Dale J.**, and Schantz M. From genes to genomes, concepts and applications in gene technology. John Wiley and Sons, 2002.
- Kolter R.** and Yanofsky C. 1982. Attenuation in amino acid biosynthetic operons *Annu. Rev. Genet.* 16: 113-134.
- Kramer, M.**, Sanders, R. A., Sheehy, R. E., Melis, M., Kuehn, M., and Hiatt, W. R. Field evaluation of tomatoes with reduced polygalacturonase by antisense RNA. In: *Horticultural Biotechnology*, eds. A. B. Bennett and S. D. O'Neil, Wiley-Liss Inc., New York, 1990.
- Lewis M.**, Chang G., Horton N., Kercher M., Pace H., Schumacher M., Brennan R., and Lu P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer *Science* 271: 1247-1254.
- Lodish, H.**, Berk, A., Zipursky, L., and Matsudaira, P., Molecular Cell Biology (fifth edition). W. H. Freeman and Company, 2000.
- Niu W.**, Kim Y., Tau G., Heyduk T., and Ebright R. 1996. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase *Cell* 87: 1123-1134.
- Robinson R.** Genetics. Volume 3, MacMillan Reference USA, 2003.
- Schultz S.**, Shields G., and Steitz T. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees *Science* 253: 1001-1007.
- Verma P.**, and Agarwal V. molecular biology. First Edition Schand group, India 2004.
- Watson J.**, Baker T., Bell S., Gann A., Levine M., Losick R. Molecular biology of the gene. Fifth edition, Pearson Prentice Hall, 2004.
- Wong D.W.S.** The ABCs of gene cloning. Second edition, Springer, 2006.

Chapter Seven

DNA Mutation and Repair



Skull mutation (classical) - 1. Meto picture records.

Introduction

There are several types and forms of the damage that taken place for the genetic material. The most important and common one may be the mutations. The DNA damage the cell unable to repair may leads to mutation. And whether these reasons leads to mutagenesis are spontaneous (natural) reasons or artificial, the mutations can be known in their universal concept as follows:

The mutation is a permanent and inherited change in the base sequences of the genetic material (the DNA) of the living organisms. This altered sequence can be reflected with changes in the base sequence of the mRNA and sometimes in changes in the amino acid sequence of the protein. It's possible for mutations to cause genetic diseases. It's also causes changes in many features of the cell. Mutations provide evidence shows that DNA is the genetic material. The change in the DNA sequence causes an alteration in the protein sequence. Hence, it's an additional evidence show to us how DNA encodes to protein. Moreover, the change in the shape of the organism

may allow us to identify the protein function. The presence of several mutations in certain gene may allow comparing of many of variable forms of the protein.

Types of Mutations

There exists a lot of controversy about the possible kinds of mutations among geneticists. They have been classified variously according to different criteria as follows:

First; Classification of Mutation According to Type of Cells

According to their occurrence in somatic and germinal cells following types of mutations have been classified into somatic mutations and gametic mutations. The somatic mutations are known as the mutations occurring in non-reproductive body cells. The genetical and evolutionary consequences of somatic mutations are insignificant, since only single cells and their daughter cells are involved. If, however, a somatic mutation occurs early during embryonic life, the mutant cells may constitute a large proportion of body cells and the animal body may be a mosaic for different types of cells. Somatic mutations have been often related with malignant growth (see the appendix of this chapter). As for gametic mutations, they identified as the mutations occurring in gamete cells (*e.g.*, sperms and ova). Such mutations are heritable and of immense genetical significance. The gametic mutations only form the raw material for the natural selection.

2. Classification of Mutations According to the Size and Quality

According to size following two types of mutations have been recognized: Point mutation. When heritable alterations occur in a very small segment of DNA molecule, *i.e.*, a single nucleotide or nucleotide pair, then this type of mutations are called “point mutations”. As for the non-point mutations which include larger segments of DNA, that are big mutations on the level of chromatin or the whole chromosome are therefore out of the scope of this book. The point mutations may occur due to following three types of subnucleotide change in the DNA and RNA. 1. Deletion mutations. The point mutation which is caused due to loss or deletion of some portion (single nucleotide pair) in a triplet codon of a cistron or gene is called deletion mutation. Deletion mutations have been frequently reported in some bacteriophages (Phage T4). 2. Insertion or addition mutation. The point mutations which occur due to addition of one or more extra nucleotides to a gene or cistron are called insertion mutations. The insertion mutations can be artificially induced by certain chemical substances called mutagens such as acridine dye and proflavin. A proflavin molecule, it is believed, insert between two successive bases of a DNA strand, thereby stretching the strand lengthwise. At replication, this situation would allow the insertion of an extra nucleotide in the complementary chain at the position occupied by the proflavin molecule. The mutations which arise from the insertion or deletion of individual nucleotides and cause the rest of the message downstream of the mutation to be read

out of phase, are called frameshift mutations. They result in the production of an incorrect, hence, inactive protein, due to which the death of the cell may occur. 3. Substitution mutation. A point mutation in which a nucleotide of a triplet is replaced by another nucleotide, is called substitution mutation. The substitution mutation affects only a particular triplet codon. Such an altered code word (triplet codon) may designate a different amino acid and may result in the production of a protein with a single amino acid substitution. The substitution mutations alter the phenotype of an organism variously and are of great genetical significance. The alteration mutation in a single base is considered the most common mutations, which is a change in a single base pair, and it can be transition or transversion. In the first type, a given pyrimidine changes into the other pyrimidine, or a given purine changes into the other purine. Transversion mutations are changes from purine into either one of two pyrimidines, or a change from pyrimidine into either one of the two purines (figure 7.1). The presence of transverse mutations for the first time was revealed at 1959 by E. Freese.

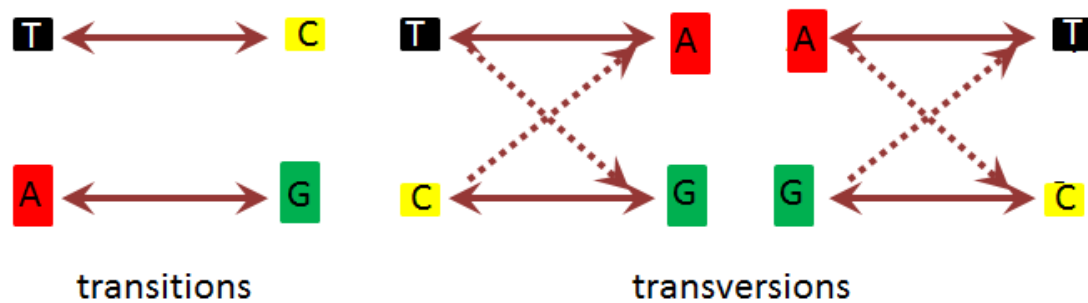


Figure (7.1); illustration shows the difference between transition and transversion mutation.

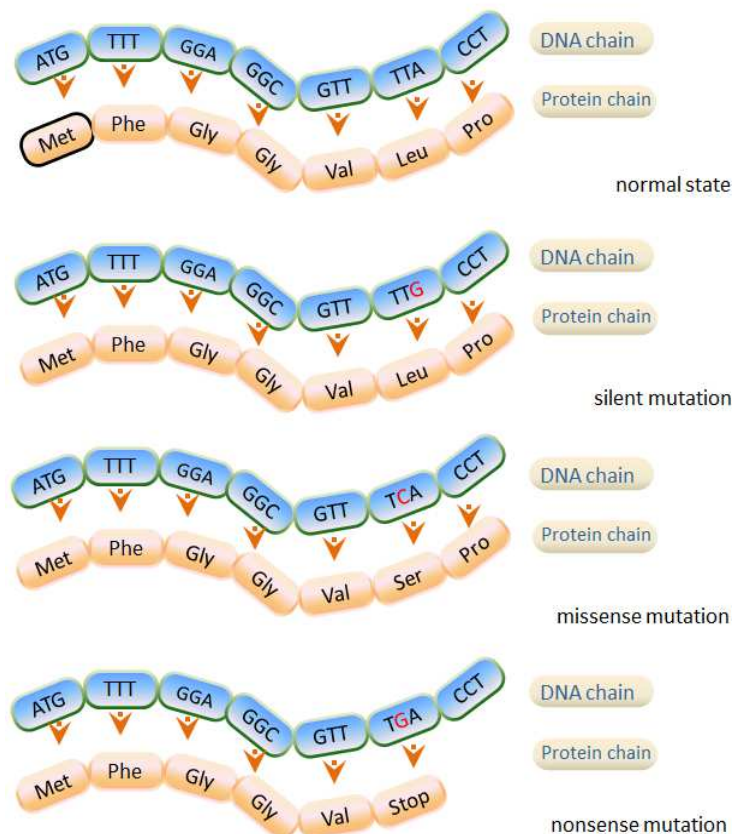
Third; classification of mutations according to their effect on the amino acid sequence of the resulting protein

The change in a single base pair of mRNA molecule may have one out of variable effects when its translated into protein.

(1) Silent mutations; in which there may be no detectable effect because of the degeneracy of the code. This would be more likely if the changed base in the mRNA molecule were to be at the third nucleotide of a codon. This occurs because of the degeneracy of the genetic code (each amino acid can be encoded by more than one code whereas each code encodes only single amino acid). For example, leucine, which is encoded by six different genetic codes, when UUA codon is changed into UUG, both codons encodes the same amino acid (leucine), this made some researchers to call this effect a non-detectable.

(2) **Missense mutations;** which occur when a different amino acid is incorporated at the corresponding site in the protein molecule. This mistaken amino acid—or missense, depending upon its location in the specific protein—might be acceptable, partially acceptable, or unacceptable to the function of that protein molecule. From a careful examination of the genetic code, one can conclude that most single-base changes would result in the replacement of one amino acid by another with rather similar functional groups. This is an effective mechanism to avoid drastic change in the physical properties of a protein molecule. If an acceptable missense effect occurs, the resulting protein molecule may not be distinguishable from the normal one. A partially acceptable missense will result in a protein molecule with partial but abnormal function. If an unacceptable missense effect occurs, then the protein molecule will not be capable of functioning in its assigned role.

(3) **Nonsense mutations;** which occur when a nonsense codon may appear that would then result in the premature termination of amino acid incorporation into a peptide chain and the production of only a fragment of the intended protein molecule. The probability is high that a prematurely terminated protein molecule or peptide fragment will not function in its assigned role.



(4) **Temperature sensitive mutations:** which occur when the alteration that takes place in the codon leads to production of active protein in certain temperature (usually at 30°C) and inactive at higher temperatures (usually at 40 – 42°C). This alteration in the protein structure that is resulting from such mutations has been indicated using figure (7.2) and list (7.1).

Figure (7.2); the classification of mutations according to their effect on amino acids sequence in the resulting protein.

List (7.1); the effect of some common types of mutations on the structure of the protein.

Type of mutation	The effect on the protein
Silent: the new codon encodes the same amino acid	Nothing
Missense: the new codon encodes different amino acid	A possible reduction in the function, variable effects
Nonsense mutation: the new codon is the stop codon	Shorter than the normal length, usually inactive

Fourth; Classification of Mutation According to Magnitude of Phenotypic Effect

According to their phenotypic effects following kinds of mutations may occur:

1. **Dominant mutations.** The mutations which have dominant phenotypic expression are called dominant mutations.
2. **Recessive mutations.** Most types of mutations are recessive in nature and so they are not expressed phenotypically immediately. The phenotypic effects of mutations of a recessive gene is seen only after one or more generations, when the mutant gene is able to recombine with another similar recessive gene.
3. **Isoallele mutations.** Some mutations alter the phenotype of an organism so slightly that they can be detected only by special techniques. Mutant genes that give slightly modified phenotypes are called **isoalleles**. They produce identical phenotypes in homozygous or heterozygous combinations.
4. **Lethal mutations.** According to their effects on the phenotype mutations may be classified as lethals. Lethal mutations result in the death of the cells or organisms in which they occur.

Fifth; Classification of Mutation According to the Direction

According to their mode of direction following types of mutations have been recognized:

1. **Forward mutations.** In an organism when mutations create a change from wild type to abnormal phenotype, then that type of mutations are known as forward mutations. Most mutations are forward type.
2. **Reverse or back mutations.** The forward mutations are often corrected by error correcting mechanism, so that an abnormal phenotype changes into wild type phenotype.

Sixth; Classification of Mutation According to the Origin

According to the mode of origin, following two kinds of mutations have been recognized:

1. **Spontaneous mutations.** The spontaneous mutations occur suddenly in the nature and their origin is unknown. They are also called “background mutation” and have been reported in many organisms but they are understood in prokaryotes. The major cause of spontaneous mutation results from the presence of an unusual base in the DNA. In addition to the four bases that are inserted into DNA when it is synthesized, modified bases are sometimes found. The name reflects their origin; they are produced by chemically modifying one of the four bases already present in DNA. The most common modified base is 5-methylcytosine, which is generated by a methylases enzyme that adds a methyl group to certain cytosine residues at specific sites in the DNA (see chapter six). Sites containing 5-methylcytosine provide hotspots for spontaneous point mutation in *E. coli*. In each case, the mutation takes the form of a G-C to A-T transition. The hotspots are not found in strains of *E. coli* that cannot methylate cytosine. The reason for the existence of the hotspots is that cytosine bases suffer spontaneous deamination at an appreciable frequency. The replacement of a keto group instead of amino group converts 5-methylcytosine into thymine. The figure (7.3) shows why the rare 5-methylcytosine deamination causes a mutation, while the most common cytosine deamination doesn't have such effect.

The deamination of cytosine generates uracil. *E. coli* contains an enzyme, uracil DNA glycosidase that removes uracil residues from DNA (base flipping mechanism around bases; see DNA repair in the next section). This action leaves unpaired G residues, and a "repair system" that inserts a C base to partner it. The net result of these reactions is to restore the original sequence of the DNA. This system protects DNA against the consequences of spontaneous deamination of cytosine. But the deamination of 5-methylcytosine leaves thymine. And because this base is a respectable component of DNA, it's impossible for the system to recognize this change, therefore, the mutation is generated. This conversion creates a mismatched G with T. the consequence of their separation produces a normal G≡C pairing of one site, and a mutant A=T pairing of the other site (figure 7.4).

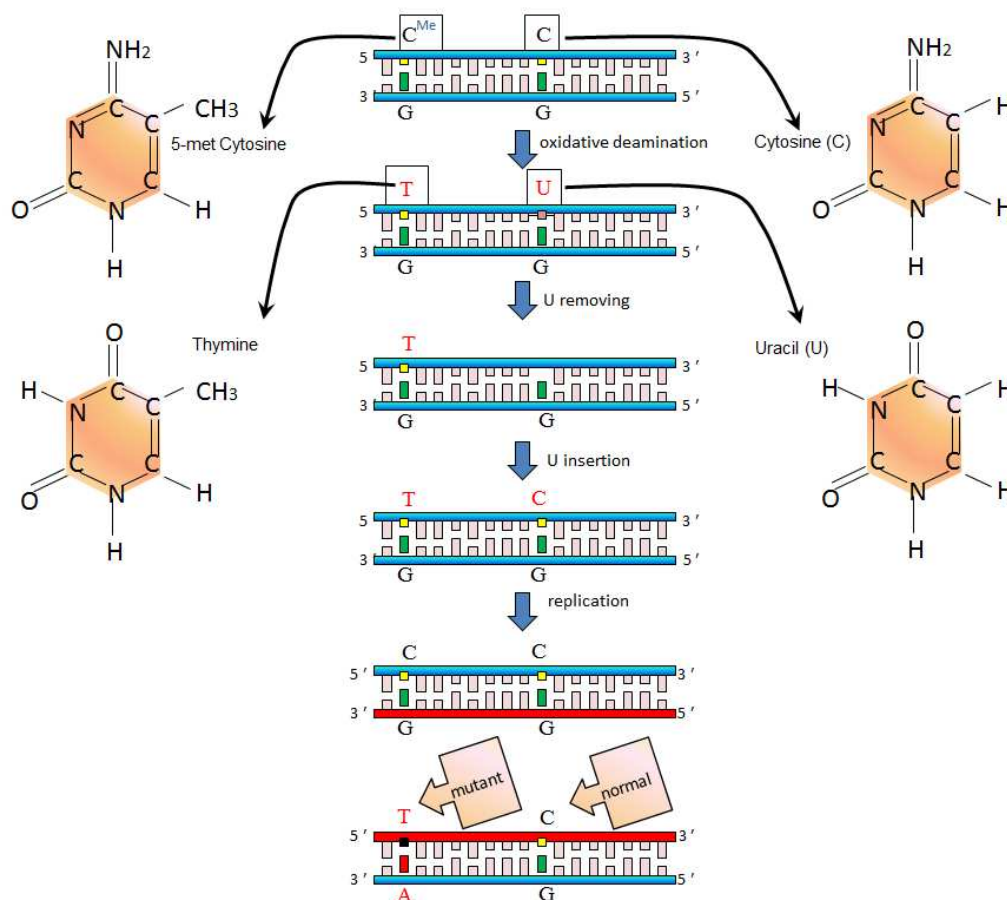


Figure (7.3); deamination for 5-methylcytosine producing to thymine (transition mutation from C≡G into T=A), whereas deamination of cytosine produces uracil (which is usually removed and replaced with cytosine).

The action of this system focuses on T utilization in DNA compared with U in RNA. This may be connected with the requirement of DNA to the stability in its sequence: the utilization of T means the ability for immediate recognition for any deamination for C, because C generates U which doesn't usually exist in DNA. Thus, there are may be several major reasons explaining why RNA has U instead of T, whereas DNA has T instead of U. when amino group is removed spontaneously from C to form U, it's about 100 spontaneous action for each mammalian cell per day, this rate is very high. Uracil DNA glycosylase recognizes these "mutations" and replaces these U bases with C counterparts. But it doesn't has the ability to recognize between the wild type and mutant U. and because of the importance of DNA, these cells solve this problem by using T (which is 5-methyl uracil) instead of U in DNA. Moreover, C deamination can be taken place in RNA. However, each gene has large number of mRNA copies, thus, the random mutation in any RNA copy cannot causes great effect. Add to that, A and G deamination can be taken place in 100 folds slower rate than C deamination.

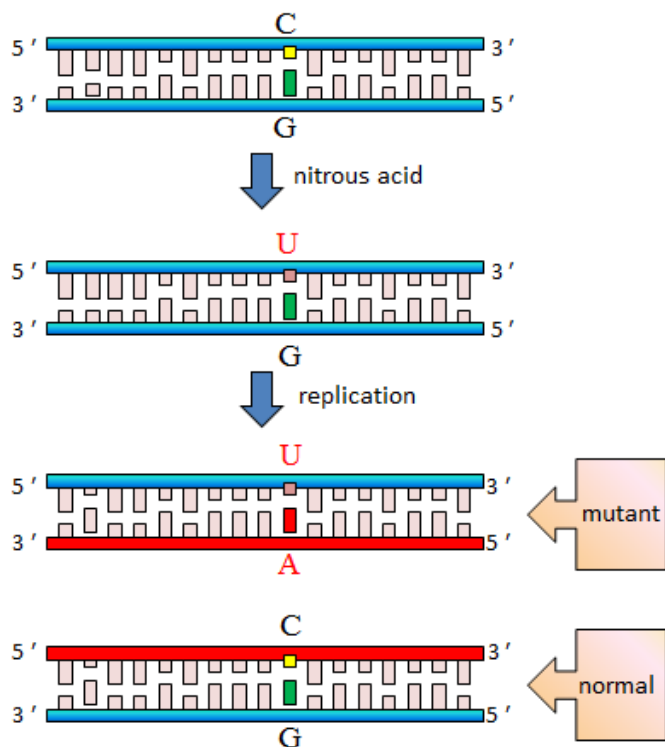


Figure (7.4); the ability to induce mutations by chemical base modification.

2. Induced mutations.

Besides naturally occurring spontaneous mutations, the mutations can be induced artificially in the living organisms by exposing them to abnormal environment such as radiation, certain physical conditions (*i.e.*, temperature) and chemicals. The substances or agents which induce artificial mutations are called mutagenes or mutagenic agents.

Mutagenes

Mutagenes are divided into two main categories:

1. **Physical mutagenes;** the radiations which are important in mutagenesis are of two categories: one type is ionizing radiations such as X-rays and gamma rays which cause damages to the target molecule. Such mutagenes can cause condensed chemical alterations for DNA, which include degradation in the strand and breakage in the ribose and nitrogen bases. Or, these mutagenes can be non-ionizing radiations which cause molecular agitation or induction for electrons into higher levels of energy within the target molecule (excitation). The excited molecule becomes reactive and mutated and is called photoproducts. Add to that, the exposure for non-ionizing radiation can lead the formation of new chemical bonds. The most important forms that cause damage for DNA are UV light, which produces pyrimidine dimers by the adjacent pyrimidine pairs.
2. **Chemical mutagenes;** there are wide range of organic and inorganic chemical materials, whether natural or artificial which can react with DNA and modify its characters. It's possible for chemical mutagenes to cause mutation only when they penetrate the nucleus. Hence, they can affect on the chromosomal DNA by two ways. The first one is called direct gene change; in which chemical mutagenes affect on DNA directly. They affect the constituents of DNA only when DNA is not replicating. For example, nitrous acid converts adenine into

hypoxanthine and cytosine to uracil by deamination. Like the nitrous acid, nitrogen mustard, formaldehyde, epoxides, and nitrosoguanidine and other chemical materials also have direct mutagenic effect on the DNA molecule. The other method is called copy error. In which, certain chemical compounds, called base analogues (e.g., 5-bromouracil, 2-aminopurine, etc.) closely resemble with certain DNA bases and are, therefore, act as mutagens. During DNA replication, they are incorporated by DNA in place of the normal DNA bases. Certain other base analogues such as caffeine (in coffee, tea and soft drinks), phenol and carcinogens, acridines and others materials that have mutagenic effects.

Effect of Chemical Mutagens on Nucleotide Sequence

The effects that can be done by chemical mutagenes which affect on the nature of nucleotide sequences are summarized by following points;

(a) Alteration in Resting Nucleic Acid; which can be done in the following forms:

1. **Deamination.** Some chemical substances such as nitrous acid causes transitional mutation due to oxidative deamination of DNA bases. In the process of oxidative deamination, the amino group (NH₂) of a DNA base is replaced by hydroxyl (OH) group by the chemical mutagen. Thus, adenine By tautomeric shift the hypoxanthine is converted into more common or keto-tautomer which pairs with cytosine. The A: T pair, thus, can be converted to a G : C pair. Similarly, deamination converts C to U, which has pairing properties similar to thymine and in such a case G: C pair would be changed into A : T pair.
2. **Hydroxylamination.** When DNA is treated with hydroxylamine, its cytosine base is the strongest reacting base. Hydroxylamine probably causes hydroxylation of cytosine at amino group giving rise to hydroxylcytosine, which then subsequently pair with adenine. Thus, hydroxylamine induces in DNA a GC into AT base pair transition.
3. **Pyrimidine rings breaking.** The hydrazine affects DNA by breaking of rings of uracil and cytosine giving rise to pyrazolone and 3-aminopyrazole, respectively. The treatment of RNA or DNA with anhydrous hydrazine results in the destruction of their pyrimidines.
4. **Alkylating agents.** Some alkylating agents carry one, two, or more alkyl groups in a reactive form and act as strong mutagens. Examples of some most extensively studied alkylating agents include diethyl sulphate, dimethyl sulphate, methyl methane sulphonate, ethyl ethane sulphonate and ethyl methane sulphonate. These mutagens produce mutations in the following ways : (a) They add ethyl or methyl groups to guanine. This makes guanine the base analogue to adenine. (b) They remove the alkylated guanine. This is known as depurination. The loss of the base produces gaps

in the DNA chain which may be filled with a wrong base, thus, producing mutation. (c) The gap may also produce a deletion, causing mutation.

(b) Alteration during Replication of Nucleic Acid: which can be done in the following forms:

1. **Base analogues.** Certain chemical substances have molecular structure similar to the usual DNA bases that, if they are available, such analogues may be incorporated into a replicating DNA strand. For example, 5-bromouracil (5BU) or its nucleoside 5-bromodeoxyuridine (5-BUdR) in its usual (keto) form is a structural analogue of thymine (5-methyl uracil) and it will substitute for thymine. Thus, an AT pair becomes and remains A-BU. There is some in vitro evidence to indicate the BU immediately adjacent to an adenine in one of DNA strands causes the latter to pair with guanine. But, in its rare (enol) state, 5BU behaves similar to the tautomer of thymine and pairs with guanine. This converts A : T to G : C. the figure (7.5) shows an example to this state which is represented by bromouracil (BrdU), shows an analog of thymine that contains a bromine atom in place of the methyl group of thymine. BrdU is incorporated into DNA in place of thymine. But it has ambiguous pairing properties, because the presence of the bromine atom allows a shift to occur in which the base changes structure from a keto (=O) form to an enol (-OH) form. The enol form can base pair with guanine, which leads to substitution of the original AT pair by a GC pair. The mistaken pairing can occur either during the original incorporation of the base or in a subsequent replication cycle. The transition is induced with a certain probability in each replication cycle, so the incorporation of BrdU has continuing effects on the sequence of DNA.
2. **Inhibition of precursors of nucleic acids.** There are some mutagens which interfere with the synthesis of nitrogen bases of nucleic acids such as purines or pyrimidines. Often lack of one base either causes breaks or causes pairing mistakes. For example, azaserine that inhibits purine synthesis and urethane that is an inhibitor of pyrimidine synthesis. However, urethane induced chromosome breaks are inhibited by thymine.
3. **Transversion.** (see classification of mutations according to size and quality) it usually includes the substitution mutation when involves the substitution or replacement of a purine with a pyrimidine or vice versa. And as it mentioned above, these mutations want replication to be accomplished.

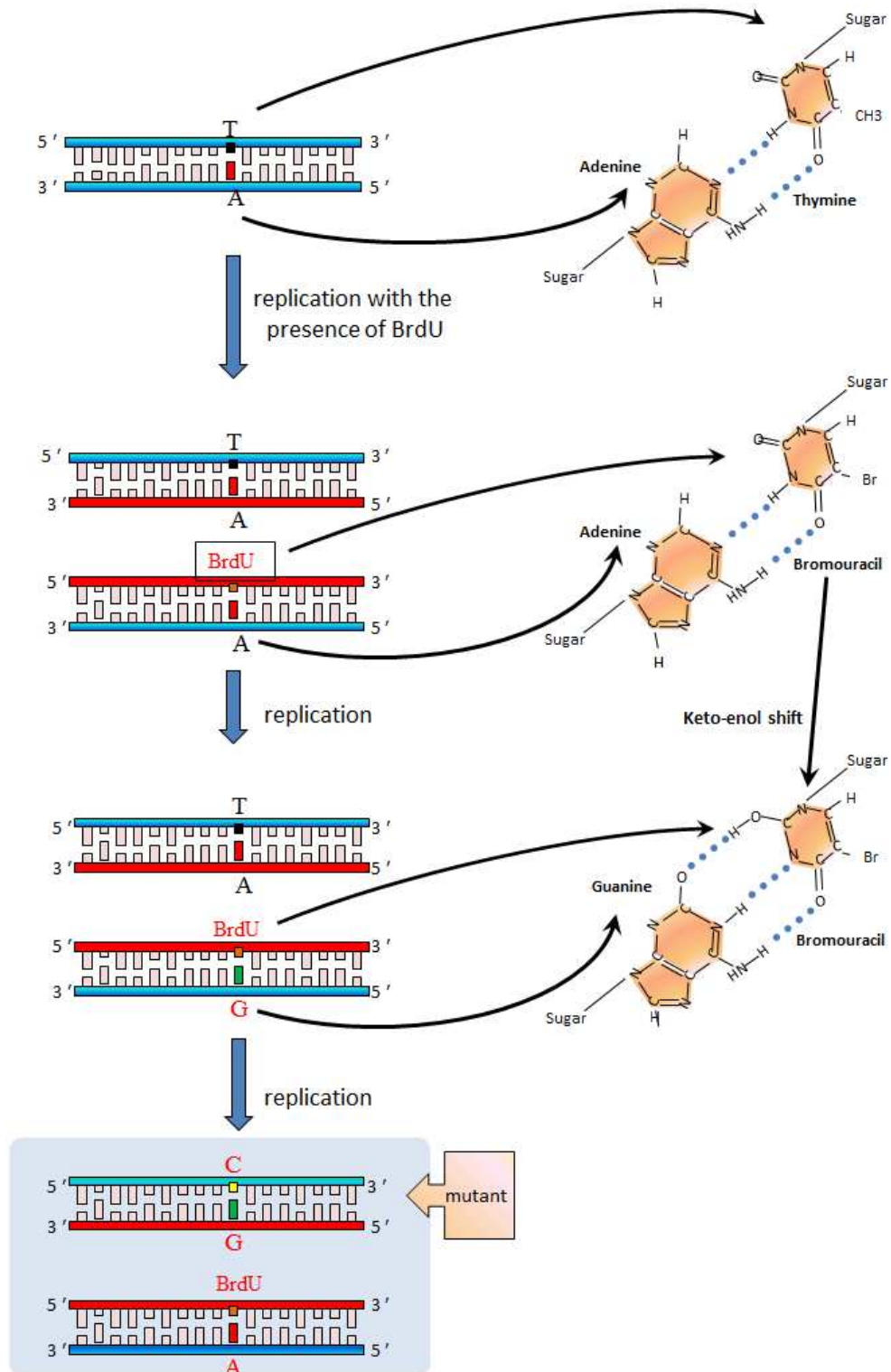


Figure (7.5); the ability to induce mutations by inserting the base analogs in DNA.

Problems of Dealing with Chemical Mutagenes

Experiments involving chemical mutagens suffer from a number of drawbacks such as safety is considered as a concern in this aspect, as mutagens are usually carcinogens and are also frequently highly toxic. They are very unpleasant compounds to work with. The dose of mutagen used must be chosen carefully to give an average of 0.1 mutation per genome; otherwise, the resultant viruses will contain multiple mutations which can complicate interpretation of the phenotype. Therefore, most of the viruses that result will not contain any mutations, which is inefficient as screening for mutants can be very laborious. There is no control over where mutations occur, and it is sometimes difficult or impossible to isolate mutations in a particular gene or region of interest. For these reasons, site-specific molecular biological methods such as oligonucleotide- directed mutagenesis or PCR-based mutagenesis are now much more commonly used since it provides many features represented by its specificity, its low biohazard, its capability to give reliable results in DNA rearrangement experiments.

DNA Repair

As long as life proceeded and complicated, cells must face the mutagenes by many means that Allah Almighty given to them. But, it's necessary to maintain the mutation rate within certain limits could not be crossed. This is taken place through repairing the damage that happened in the non-replicating DNA from time to time and also for DNA during its replication.

DNA is packaged usually within the chromosomes. Chromosomes are large molecules can be an easy target for breakage. A single damage on a DNA molecule can prevent a whole chromosome from replication. So, this can leads to the death of the cell. Thus, it's obviously so important for the living system to repair the damage. The concern is directed toward DNA repair since most types of repairable damages are mutagenic or carcinogenic. Despite of the DNA is very stable material – this feature is required for storage of genetic information – but it's very complicated organic material that is exposed, even in the normal cellular conditions into spontaneous changes which leads to mutations if they are left in its un-repaired conditions. For example, DNA bases are damaged sometimes with UV light to form thymine dimers if left uncorrected when the DNA is replicated, most of these changes would be expected to lead either to the deletion of one or more base pairs or to a base-pair substitution in the daughter DNA chain. The mutations would then be propagated throughout subsequent cell generations as the DNA is replicated. Such a high rate of random changes in the DNA sequence would have disastrous consequences for an organism.

The chemistry of DNA Bases Facilitates the Detection if it's Damage

The double-helical structure of DNA is ideally suited for repair because it carries two separate copies of all the genetic information—one in each of its two strands. Thus, when one strand is damaged, the complementary strand retains an intact copy of the same information, and this copy is generally used to restore the correct

nucleotide sequences to the damaged strand. Thus, each possible deamination in DNA produces abnormal nitrogen bases. Therefore, they can be recognized directly and removed by specialized DNA glycosylase.

DNA Repair Mechanisms

The most important DNA repair mechanisms are:

1. **Excision repair mechanism;** There are two pathways for DNA repair, using different enzymes that act upon different kinds of lesions. In both, the damage is excised, the original DNA sequence is restored by a DNA polymerase that uses the undamaged strand as its template, and the remaining break in the double helix is sealed by DNA ligase (figure 7.6). The two pathways differ in the way in which the damage is removed from DNA. The first pathway, called base excision repair, involves a battery of enzymes called DNA glycosylases. As an example of the general mechanism of base excision repair, the removal of a deaminated C by uracil DNA glycosylase that shown in the figure (7.6.A). How is the altered base detected within the context of the double helix? A key step is an enzyme-mediated “flipping-out” of the altered nucleotide from the helix, which allows the enzyme to probe all faces of the base for damage. Once a damaged base is recognized, the DNA glycosylase reaction creates a deoxyribose sugar that lacks its base. This “missing tooth” is recognized by an enzyme called AP endonuclease, which cuts the phosphodiester backbone, and the damage is then removed and repaired by DNA ligase (7.6 A).

The second major repair pathway is called nucleotide excision repair. This mechanism can repair the damage caused by almost any large change in the structure of the DNA double helix. Such “bulky lesions” include those created by the covalent reaction of various pyrimidine dimers (T-T, T-C, and C-C) caused by sunlight. In this pathway, a large multienzyme complex scans the DNA for a distortion in the double helix, rather than for a specific base change. Once a bulky lesion has been found, the phosphodiester backbone of the abnormal strand is cleaved on both sides of the distortion, and an oligonucleotides containing the lesion is peeled away from the DNA double helix by a DNA helicase enzyme. The large gap produced in the DNA helix is then repaired by DNA polymerase and DNA ligase (figure 7.6. B).

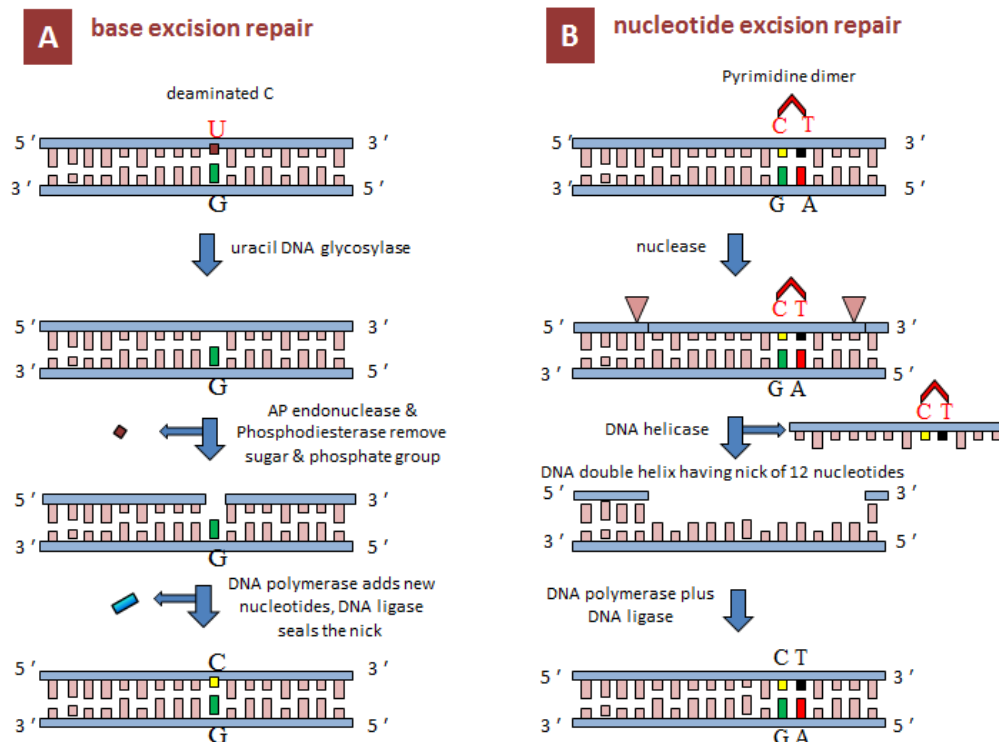


Figure (7.6); excision repair mechanism. (a) base excision repair, (b) nucleotide excision repair.

2. **Mismatch excision repair**; before we talk about this mechanism we have to shed some light again on DNA polymerase III since it's a typical example for the enzyme that can corrects the mistakes (edits) because it has 3' - 5' exonuclease activity that exists in ϵ subunit (see chapter three). Since the specificity of nucleotide addition by DNA polymerase is determined by Watson – Crick base pairing rule, this basis has a ratio of mistake, in which a mistaken base (as in A instead of G) is sometimes inserted during DNA replication. Actually, the α subunit that is responsible on polymerization in *E. coli* DNA polymerase III produces about one mistaken base per one hundred thousand (10^4) during DNA replication. But the rate of mutations in bacterial cells is largely less than this, and its approximately one mistake per 10^9 nucleotides. The increasing of the efficiency initially refers to the proofreading function in *E. coli* DNA polymerases. In DNA polymerase III, this functions held in ϵ subunit in the core polymerase. When a mistaken nucleotide is incorporated during DNA replication, the polymerase is temporarily stops, and then the 3' end of the growing DNA chain is transferred into the 3' - 5' exonuclease activity, in which the mistakenly paired base is removed. Then, the 3' end is transferred backward into the 5' - 3' polymerizing site (figure 7.7), where nucleotides are added sequentially during polymerization (see chapter three).

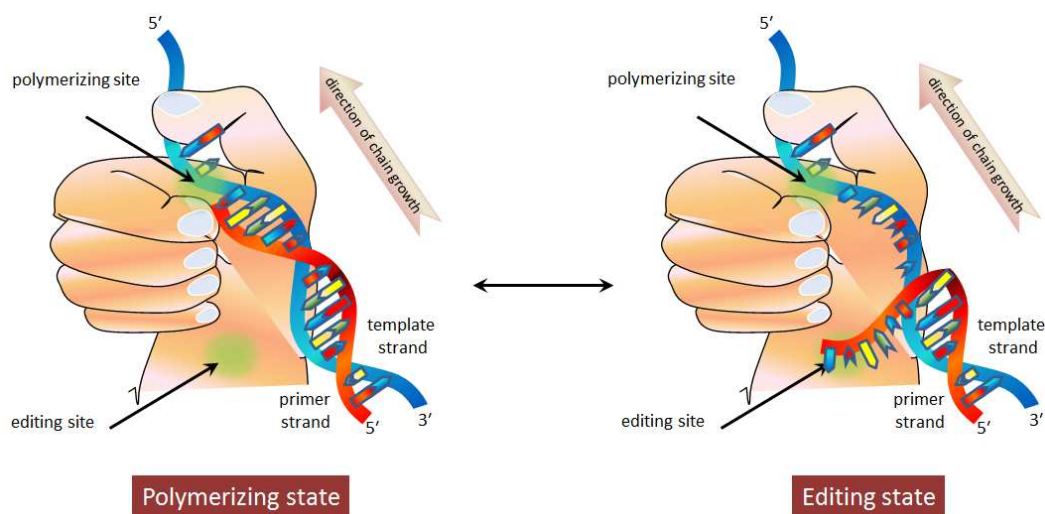


Figure (7.7); Editing by DNA polymerase. Outline of the structures of DNA polymerase complexed with the DNA template in the polymerizing mode (left) and the editing mode (right). E refers to the catalytic site that has the exonucleolytic activity; P refers to the catalytic site that has the polymerizing activity. The catalytic site for the exonucleolytic (E) and the polymerization (P) reactions are indicated.

Despite these superior ability of DNA polymerase III in editing reactions, occasionally certain bases are inserted incorrectly which if it is not corrected the mutation occurs somehow. Hence, this system is called mismatch repair, which is the simplest repair systems that actively removes such miss-incorporation and replace it with a correct nucleotide. If the cell wants to do so, it has to distinguish between what is the old strand and what is the new strand, through its ability to distinguish between the strand that have the correct information (the parental or old strand) and the other strand that have the incorrect information (the new strand) through the presence of methyl group in the old strand. Therefore, the system removes the mismatched nucleotide in a small region of the new strand (the strand that yet has added methyl group) and replaces it with other non-mistaken one through fillings in the resulting gaps by specialized enzymes which are DNA polymerase I and DNA ligase. Therefore, this process called methyl-directed mismatch repair.

The repair of the incorrect nucleotide or mismatch repair is may be a specialized form of excision repair which deals with any base pairing that taken place during replication after its escape from the proofreading mechanism for the replicating enzymes. In this case, the incorrect base is represented by the daughter strand. Therefore, this system has a mean to differentiate between the parental and daughter strands after passing the replication fork to ensure the replacement only of the incorrect base from the daughter strand. In prokaryotes, methyl groups are added to A in the sequence GATC in both strands. The methylation of daughter strands is delayed for several minutes behind the replication fork. Thus, the newly replicated

DNA is hemimethylated, since the parental strands are methylated and the daughter strands are not. Consequently, it's easy to differentiate between them. The mismatched base pair (such as GT or CA) can easily be distinguished in order to bind with protein complexes called MutS and MutL that bind after that with MutH endonucleases which professionally nicks the daughter strand in close proximity to the GATC site. The nick initiates the breaking of the region that contains the incorrect bases. The differentiation mechanism in eukaryotes is still not exactly unknown yet.

3. **Recombination (post-replication) repair;** There is another type of mutant that is abnormally sensitive to UV, although possessing a functional excision repair system. These bacteria are defective in a gene (*recA*) that is responsible for general recombination (see chapter eight). This indicates that excision repair is not the only mechanism for dealing with UV damage, but that there is a further repair mechanism involving general recombination (figure 7.8). The forms of DNA damage that interfere with the base pairing between the strands will normally prevent replication, due in part to the requirement of the DNA polymerase for replication. The replication fork will therefore pause. It is possible however for replication to restart beyond the lesion, thus leaving a gap in the newly synthesized strand. This region of DNA cannot be repaired by excision repair, which requires one intact strand. The gap can however be filled using a portion of DNA from the other pair of strands by a recombination process (i.e. cutting and rejoining the DNA). Although this merely re-assorts the damage rather than directly repairing it, it does achieve a situation where the damage is repairable. The original damage can now be repaired by excision repair while the gap in the other DNA molecule can be filled in by DNA polymerase I and DNA ligase.
4. **SOS Response Repair;** The SOS repair system is a complex cellular response mechanism activated by certain types of DNA damage. Because the repairing is done in the absence of proper template instructions, many errors are made, so it is referred to as an error-prone mechanism. Therefore, this mechanism is not considered as repair mechanism strictly. The SOS response is induced by DNA damage that blocks replication (e. g. UV-produced thymine dimers). This cellular response is used when overwhelming levels of DNA damage is faced which prevent the normal replication to modify the DNA temporarily or abolish the specificity of DNA polymerase. Consequently, this method enables the continuity of synthesis of new DNA strand despite of the absence of the necessary accuracy in replication. The SOS response is a reason for the mutations that generated from irradiation with UV light.

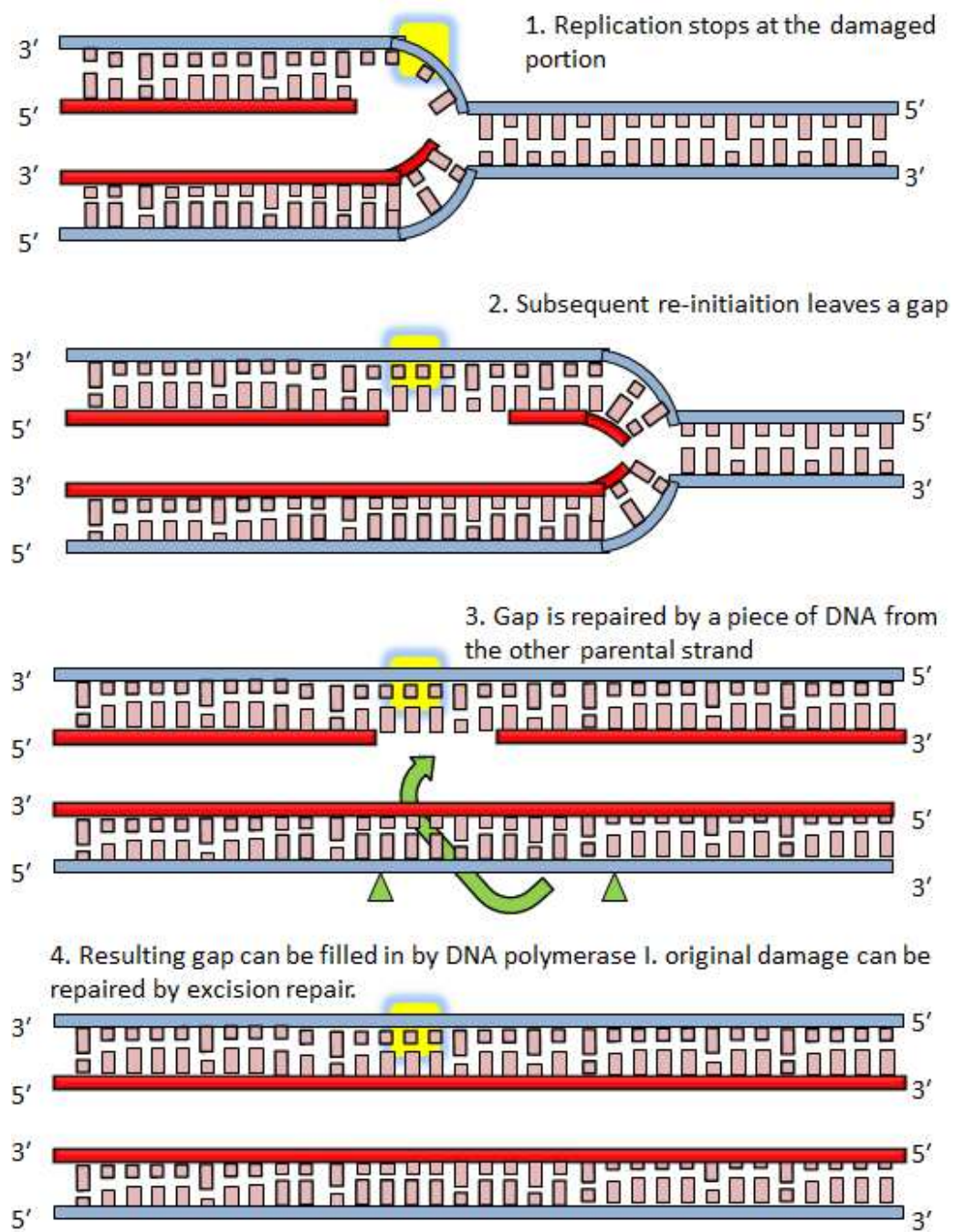


Figure (7.8); post-replication repair mechanism. After replication stops at the damaged site (1) subsequent re-initiation leaves a gap in the new strand, (2) this can be repaired by DNA exchange, (3) allowing for the original lesion to be repaired, (4) which is taken place by excision repair.

The SOS repair mechanism in *E. coli* is the best studied examples about the repair systems that resulted from the emergency response for the acute damage in DNA. In the core of SOS system there is two genes, *lexA* and *recA*. Under the normal conditions, *lexA* acts as repressor by its binding with approximately 17 genes which encode proteins involved in multiple types of DNA repair system. These genes are altogether called damage inducible genes (*din*) (figure 7.9). In this bacteria, any stall in DNA replication because of DNA damage generate a certain signal, this signal (an increase in single stranded DNA) activates *recA* protein at the beginning. This protein breaks down *lexA* through its promoting to *lexA* self cleavage of *lexA*. The reaction of protein breakdown is entirely abnormal, since *lexA* is a protein has latent protease activity which is easily activated when it's exposed to *recA*. This protein activity hydrolyzes the protein that produces it – *lexA* – autonomously. This acts on destruction of the repressor function of *lexA*, and this promotes all the proteins that *lexA* bind with (figure 7.9). This lead to the expression of genes that are repressed by *lexA* through the activated *recA*. After releasing of these proteins, the mutation rates are temporarily increased by increasing the number of errors in copying DNA sequences. These errors are resulted from the production of low fidelity DNA polymerases that can utilize the damaged DNA as a template for DNA synthesis.

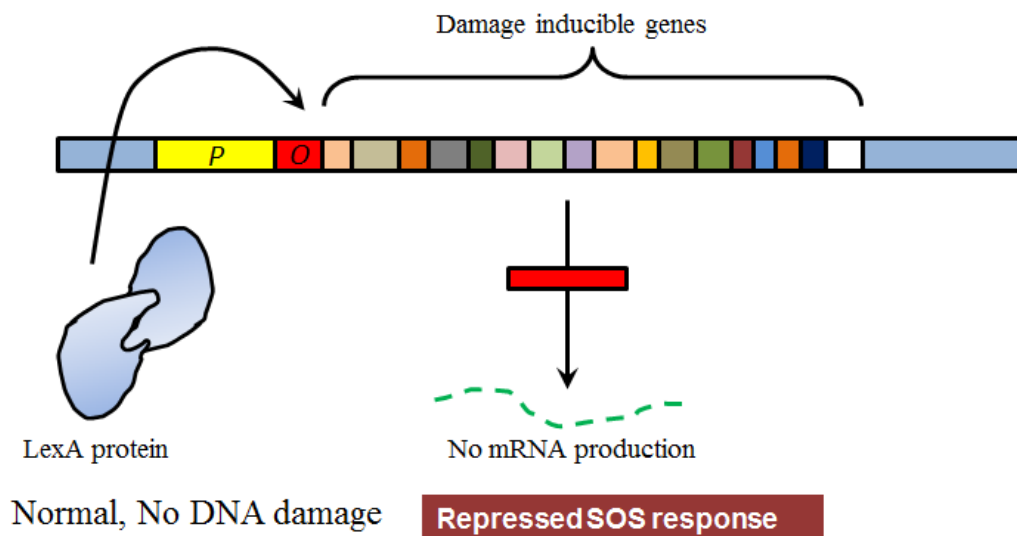


Figure (7.9); SOS operon repression mechanism (*din* refers to damage inducible genes).

This story is not ends yet, since SOS response is a circuit its importance is held on its ability to come back into its normal state. When the inducing signal is removed, *recA* losses its ability to destabilize *lexA*. In this glance, *lexA* gene is expressed in high level, and when the activated *recA* is absent, *lexA* is rapidly accumulated in its un-cleaved form and blocks SOS genes (figure 7.10). This explains why SOS is a reversible. Although it's possible for such "error prone" repair mechanism to be

harmful for bacterial cells, it's important on long term because of its production for enormous number of genetic variations in bacteria which increase the possibility of increasing of the ability of the SOS resulting mutated cells to survive under the harsh circumstances.

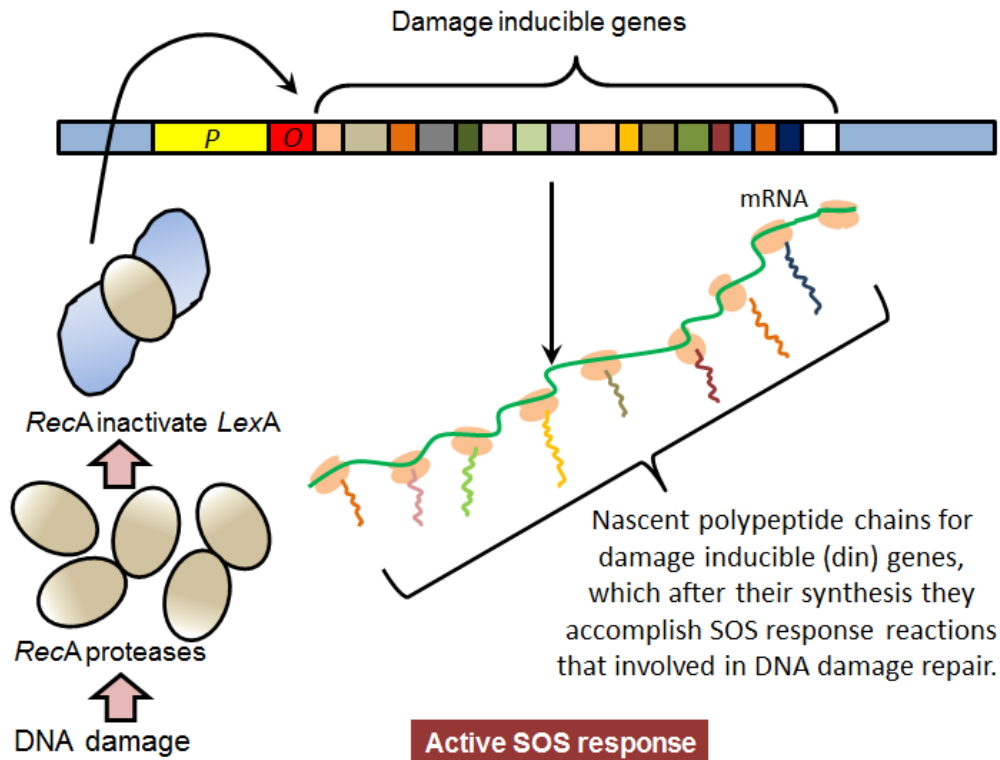
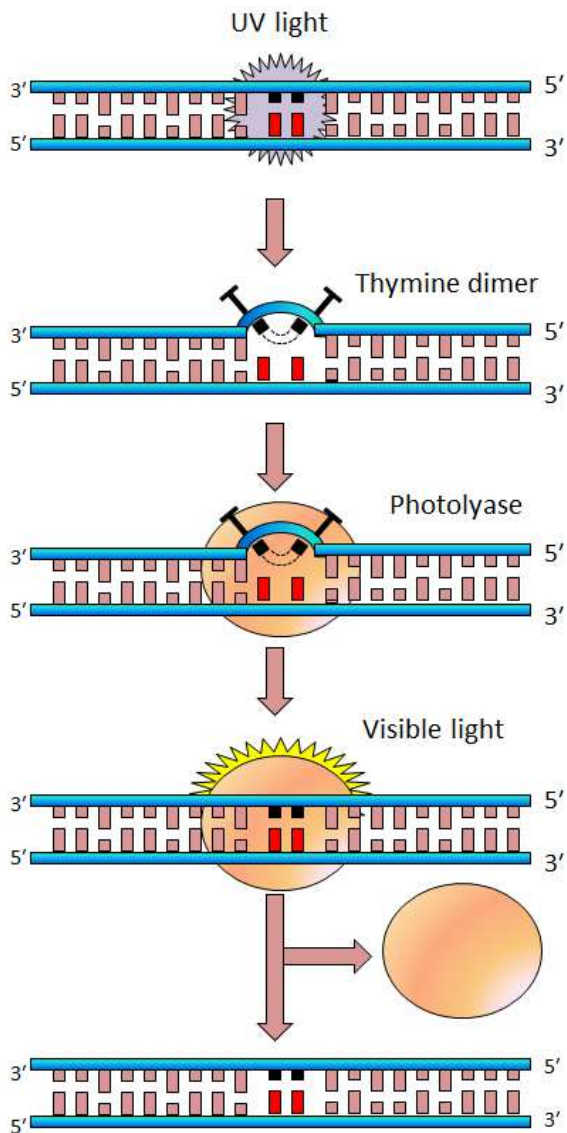


Figure (7.10); SOS operon repression mechanism (*din* refers to damage inducible genes).

5. **Photoreactivation Repair;** the effect of UV light cannot be repaired by single mechanism. When the adjacent pyrimidine dimers of DNA strand form pairs with high efficiency after the absorption of UV light, they called cyclobutane pyrimidine dimers (figure 7.11). there is another mechanism to repair this damage rather than excision repair, the presence of this mechanism is excluded on the pyrimidine dimers only, in which purine bases are highly resistant to the damage that induced by UV light. Pyrimidine dimers or other forms of DNA damage can be removed by excision repair. By this form of repair, the damage of DNA is removed and repaired. The excision repair is a multistep system, but there is a system that can fix these dimers directly by the direct reversal which can be done by the bacterial photolyase. In which the Photoreactivation enzymes are bound with DNA that containing the pyrimidine dimer and use the visible light to cleave the dimer without breaking any phosphodiester bond to get rid of the cyclobutane pyrimidine formed between the two pyrimidine bases, and restoring these two bases into the their original state before being exposed to UV light, so, the damage is repaired by this way. As for human beings, the inability to repair DNA damage is connected



with rare genetic syndromes, such as *Xeroderma pigmentosum* (XP). Since the people that infected with this syndrome have oversensitivity against the exposure to UV light, because of their inability to repair certain DNA photoproducts that are resulted from radiation, where pigmented lesions on areas of the skin exposed to sun (figure 7.12). And an elevated incidence of skin cancer with an incidence of about 1000 times that found in un-infected people. This error in the repair produces an increase in the rate of mutation, and this leads to an increase of certain cancers ratios.

Figure (7.11); Photoreactivation mechanism that is responsible on the repair of the damage induced by exposure to UV light.

The photoreactivation mechanism in mammals is taken place with a low frequency. Therefore, when both strands of DNA are damaged, there is an incidence for the appearance of dangerous consequences. Therefore, another mechanism must exist that deals with the repair of the damage in both strands.



Figure (7.12); syndromes of patients with *Xeroderma pigmentosum* (Google image).

6. **Double stranded Breakage Repair**; the breaks that occur in only one strand of DNA were discussed with the way of their repair. And in all the above mentioned mechanisms, the intact strand is used to repair its damaged counterpart. But, what happened if both strands are damaged in the same time, such as in case of exposure to the ionizing radiation and some metabolic products of the cell? If these lesions are left without repair, they will lead quickly to chromosomal degradation into smaller fragments. There are two identified mechanisms to repair this type of damage. The most easily understood mechanism is known as non-homologous end joining, by which the broken ends are aligned and rejoined with each other by DNA ligation. Usually, this leads to the loss of only one nucleotide or more at the site of joining (figure 7.13 A).

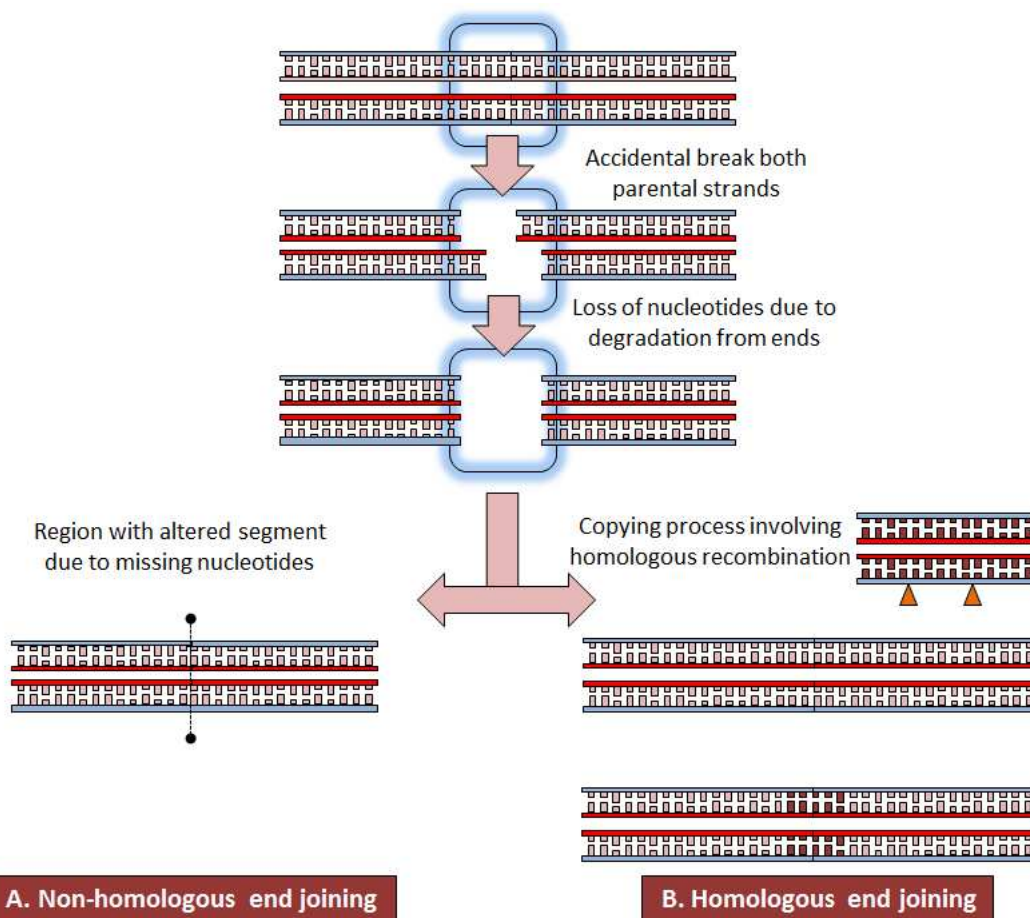
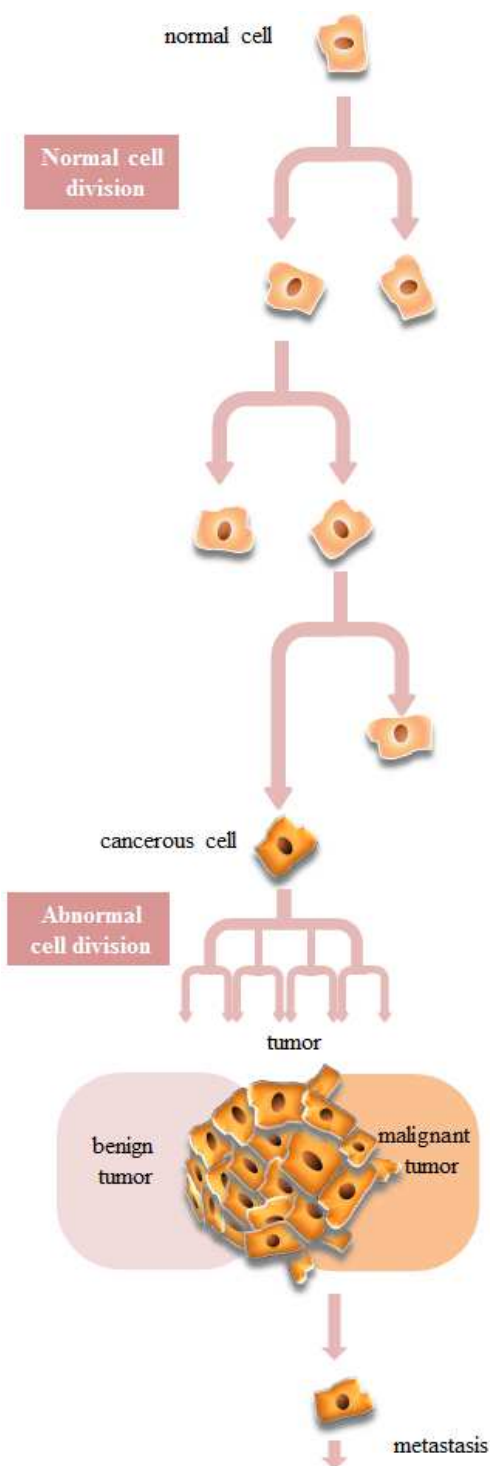


Figure (7.13); double stranded DNA breakage repair. (A) non-homologous end joining, (B) homologous end joining.

7. The end joining mechanism, which can be imagined as an emergency solution to repair double stranded breaks, is a common solution mechanism in mammalian cells. There is a more efficient type in repairing the double stranded breaks utilizes



the fact that diploid cells consist of two copies of each double helix. The second repair pathway is called homologous end joining. Hence, the general recombination mechanisms are recruited to transfer the nucleotide sequence information from the intact DNA double helix into the legion of the broken DNA double helix (figure 7.13 B). This type of reaction requires special recombination proteins which recognize corresponding DNA sequences regions between chromosomes and join them together. Then, the non-damaged chromosomal replication is used as a template to transfer the genetic information the broken chromosome, repairing it without any change in its sequence.

Chapter Seven Appendix

Cancer An Accumulation of Mutations

Cancer is a disease that begins when a single cell escapes from the regulation of its own division. This division is normally regulated so that a cell divides only when more cells are required, and when conditions are favorable for division. A cancerous cell is a rebellious cell that divides without instructions from the body. Unregulated cell division leads to a pile of cells that form a lump or tumor. A tumor is a mass of tissue that has no apparent function in the body. Tumors that stay in one place and do not affect surrounding tissues are said to be benign. Some benign tumors remain benign, others become cancerous. Tumors that invade surrounding tissues are no longer benign—they are malignant, and these invasive tumors are cancers. Cells from the original tumor can break away and start new cancers at distant locations, a process called metastasis (Figure 7.14). While it's possible to remove benign tumor surgically, malignant tumor spreading for distant parts of the body made it repeatedly resistant for these local treatments.

Figure (7.14); the meaning of cancer on the cellular level. The tumor is a group of nonfunctional cells. Tumors may be benign and they may invade the surrounding tissues and become malignant.

There are number of carcinogens that originated from the food and chemical pollutants. Moreover, the ionizing radiations are considered as carcinogenic as well. For example, there are strong evidences confirm that the ionizing radiation that come from the sun are basic cause for skin cancer. Consequently, the

ratio of skin cancer incidence in the northern part of United States is higher than its ratio in the southern states. Add to that, the high doses of radioactive materials or X-rays are connected firmly with the incidence of many forms of cancer.

In many cases, the exposure to certain types of viruses is connected with certain types of cancers. For example, the liver cancer, which is 300 times more common with persons that were previously infected with Hepatitis B virus, where the infection with Hepatitis B is the almost common cause for the human liver cancer. As for the mechanisms for conversion of proto-oncogenes into oncogenes, we will discuss four mechanisms that alter the expression or the structure of the proto-oncogenes to convert them into oncogenes. For simplicity, the process by which the transcription of a gene is increased (from zero or from the relatively low level) is called "activation". There is simplified explanation for these four mechanisms that participate in carcinogenesis.

- 1- **Chromosomal rearrangement or chromosomal translocation;** many tumor cells exert chromosomal abnormalities. One of these chromosomal changes that are seen in the cells is known as the translocation from one location to other one. The basis of this translocation is the cleavage of one chromosomal segment and its connection afterward with other chromosome. If the second chromosome offers any part of its own to the first chromosome, this process can be described as reciprocal. The translocation feature is found in many tumor cells. The "Philadelphia chromosome" is the most important types of translocations, which includes the translocation between the chromosomes 22 and 9, and this is taken place in one of the types of leukemia (figure 7.15). The translocation occurs in Burkett's lymphoma, which is a rapidly growing cancer of the human B lymphocytes. In certain cases, the activation of some oncogenes is taken place when this reciprocal translocation occurs between two chromosomes (8 and 14). In certain cases, the activation of certain oncogenes is taken place. When this reciprocal translocation occurs between the number 8 chromosome and number 14 chromosome. A portion of the number 8 chromosome carrying the proto-oncogene *c-myc*, is moved to the number 14 chromosome. This translocation put the inactive *c-myc* near a cluster of antibody genes where located. Therefore, it will be under the control of the enhancer sequences of the genes encoding for heavy chains of immunoglobulins. This proximity results in activation of *c-myc* transcription. Apparently, the increased synthesis of the protein encoded by *c-myc*, which has DNA binding activity, leads or "forces" the cell to be malignant. This may occurs through it effect on mitosis. This mechanism resembles the enhancer insertion except that this chromosomal translocation (instead of the proviral insertion) is responsible on the proto-oncogene insertion (as in *c-myc*) under the effect of enhancer.

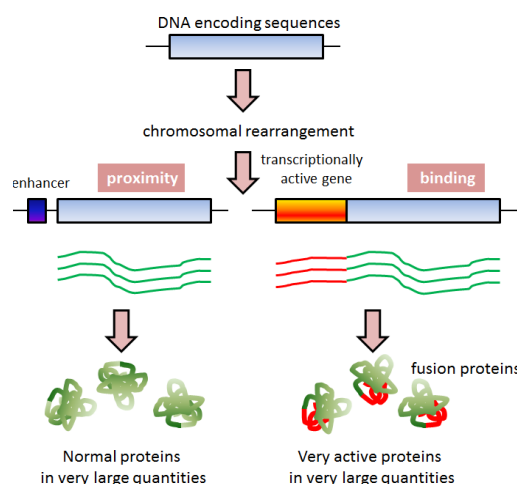
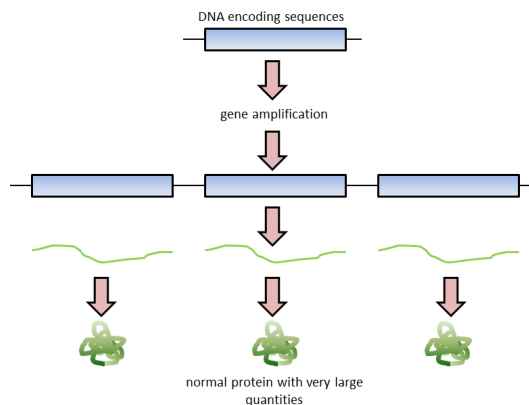


Figure (7.15); one of the mechanisms for proto-oncogene conversion into oncogenes through chromosomal rearrangement. In the right of this figure, there was binding between the transcriptionally active gene and oncogenes, this binding cause the production of very active proteins in large quantities. In the left part of this figure, there was proximity between a strong enhancer of the proto-oncogene; this binding causes the production of natural proteins but in very large quantities. This leads to carcinogenesis in both cases.

- 2- **Recombination between retroviral DNA and Proto-oncogene;** initially, this mechanism is similar to the above mentioned recombination (figure 7.15). This mechanism of two types; the first one involved

the promoter insertion. Retroviruses differ in their ability to induce cancer. Those that don't have oncogenes (such as avian leukemia virus), are able to induce cancer after long period of time – months instead of days – compared with those that contain oncogenes. When other retroviruses infect host cells, their own resulting cDNA copy is inserted into the host genome by reverse transcriptase. The inserted double stranded cDNA copy is called provirus. The retroviral cDNA copies are flanked by long terminal repeats (LTRs), such as in certain transposons that exist in bacteria and plants (see chapter eight). Seemingly, the LTR sequences have an importance in the insertion of provirus. Moreover, it's possible to act as promoter sequences for the transcription after chickens B-lymphocytes infection by avian leukemia virus (ALV) which form an inserted provirus near *c-myc*. The *c-myc* is activated by upstream LTR sequences that act as promoter to result in both transcription of the corresponding *c-myc* mRNA and its translation in these cells. Consequently, a tumor is the outcome in B-cells. The second type is enhancer insertion; the provirus is inserted in some cases downstream of or upstream of *c-myc*. But, the direction of insertion is reversed. However, the *c-myc* is activated. This activation cannot lead to promoter insertion, since promoter sequences must be upstream of the gene of interest, and the sequence must be in the proper direction (from 5' to 3'). It's possible for one or two mechanisms (promoter and enhancer insertion) to explain the viral origin carcinogenesis.

- 3- **Gene amplification;** it was found that there was an amplification for certain genes in a number of



tumors. One of the examples of gene amplification in tumors is the administration of anticancer drugs, such as methotrexate; which is an inhibitor for dihydrofolate reductase. It's possible for tumor cells to be resistant to the action of this drug. The basis of this phenomenon is that the gene that encodes for dihydrofolate reductase is became amplified. This amplification results in increasing of the enzyme activity for more than 400 times (figure 7.16). These amplified regions are recognized by their homogeneous staining. Alternatively, the amplified regions are recognized as double – minute chromosomes which are short chromosomes with missing centromeres.

Figurer (7.16); one of the mechanisms of proto-oncogenes conversion into oncogenes through gene amplification.

- 4- **Single point mutation;** the *v-ras* gene is originally discovered in certain organisms (such as mice and rats). The product of this gene is called p21; which is called according to its molecular weight, and it appears that it has a correlation with G proteins that modulate the adenylate cyclase activity. Thus, they play a major role in the cellular responses for many hormones and drugs. The analysis DNA sequences results for primary *c-ras* oncogenes that isolated from the normal human cells and *c-ras* that isolated from human bladder cancer have shown that they are differ in only one nucleotide base, resulting from the substitution of amino acid in the location number 12 for p21. The evidence for that is the gene that isolated from the tumor cells has exerted only one point mutation compared with the primary *c-ras* that isolated from the normal cells. It seems that this mutation that occurs in p21 affects on its conformation and abolish its activity as GTPase. The reduction in GTPase protein activity results in a continuous stimulation for its adenylate cyclase activity which usually abolished when GDP is formed from GTP. It's possible for the stimulation for GTPase activity to result in a group of effects on the cellular metabolism since the increased amounts of cAMP affect on a number of enzymes of kinase type (that depend on their activity on cAMP). These events may help in the disturbance in the cellular balance that may lead the cell toward malignant tumorigenesis (figure 7.17).

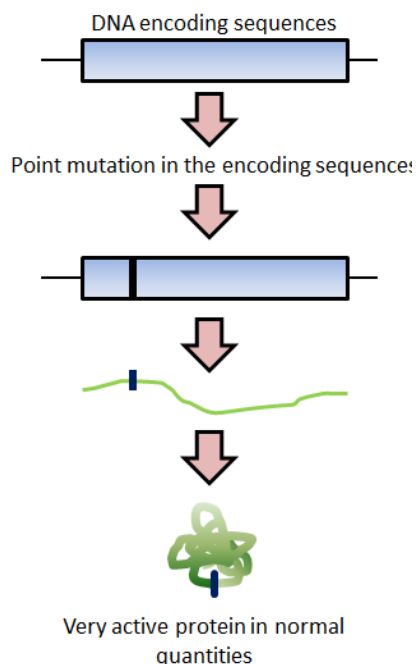


Figure (7.17); one of the mechanisms of proto-oncogene conversion into oncogenes by point mutation.

In the four above discussed mechanisms, the first three mechanisms (the recombination between retroviral DNA and proto-oncogene and chromosomal translocation form one place to the other and gene amplification) are involved in the increasing of oncogenes product as an indication for increasing transcription but not in the alteration of the oncogenes final product structure. Thus, the increasing amounts of the oncogenes product may be sufficient to push the cell toward carcinogenesis. While, the fourth mechanism, the point mutation, involves a change in the oncogenes product structure and not necessary in its quantity. This notion refers to the fact that the presence of the structurally abnormal key regulatory protein in the cell may be sufficient as well to push the cell toward carcinogenesis.

When we keep focus on the role of oncogenes, it's important to put in our mind that the activation of proto-oncogenes is not the only pathway that is responsible about carcinogenesis. Instead, the grouping of their activation with the repression (inactivation) of other genes that known as tumor suppressor genes is involved in the carcinogenesis of the most clinical cases.

Chapter Seven Revision Questions

Q1- Justify the following:

1. The 5-methylcytosine containing regions are hotspot for spontaneous point mutation?
2. The DNA contains T, whereas RNA contains U?
3. The PCR mutagenesis is better than the classic mutagenes?
4. The DNA double helix is a typical molecule to repair its own damage?
5. The mismatch repair is characterized as a mechanism that is directed by methyl?
6. The SOS response is not strictly considered as a repair mechanism?
7. The SOS response resembles the circuit?
8. Some nucleotides are lost during the repair of the damaged DNA in both of its strands?

Q2- Choose the correct answer:

1. The reading frame is not changed in
a. deletion mutation d. insertion mutation c. addition mutation d. none
2. It's possible for polypeptide to be released before the termination of translation because of the change in the reading frame of
a. deletion mutation d. insertion mutation c. addition mutation d. all
3. There are no detectable effect in silent mutation because of character of the genetic code.
a. unambiguity b. universality c. overlapping d. degeneracy
4. The singular base change may leads in missense mutation to the substitution of amino acid with another amino acid of similar functional groups for the original amino acid, this case is called
a. acceptable b. partially acceptable c. unacceptable d. a and b
5. The transversion mutation is take place by the conversion of
a. purine to purine b. pyrimidine to pyrimidine c. pyrimidine to purine d. purine to pyrimidine e. "c" and "d"
6. The removing of amino group from C causes formation of
a. adenine b. guanine c. cytosine d. thymine e. uracil
7. The nitrous acid induces mutation through mechanism.

- a. base analogue c. copy error c. direct gene change d. hydroxylation e. alkylation
8. It's usually to the polypeptide chain to be terminated before its being mature in mutation.
- a. silent b. accepted missense c. partially accepted missense d. non-accepted missense e. nonsense
9. The rate of mutation in bacterial cells is estimated about one error per each nucleotide.
- a. 103 b. 105 c. 107 d. 109 e. 1012
10. The cell uses mechanism in distinguishing between the parental and daughter strand.
- a. mismatch repair b. base excision repair c. nucleotide excision repair d. post-replication repair e. photoreactivation repair, cyclobutane pyrimidine
11. The mechanism is the most error prone repair mechanisms.
- a. mismatch repair b. excision repair c. SOS response repair d. post-replication repair e. photoreactivation repair
12. People with Xeroderma pigmentosum have excessive sensitivity for rays.
- a. chemical b. physical c. cosmic d. ultra violet e. laser

Q3- Define the following:

Transversion mutation, temperature sensitive mutation, backward mutation, depurination, isoallele mutation, photolyase.

Q4- Connect terms exist in right with its suitable counterparts in left

Missing tooth	Mutagen
Stop codon	Xeroderma pigmentosum
Mosaic	Base excision repair
Proflavin	Double stranded breakage repair
Flipping out	Nonsense mutation
Skin cancer	Somatic mutation
X-ray	Uracil DNA glycosylase
Damage inducible genes	lexA

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Aravind, L.** Walker D. and Koonin E. 1999. Conserved domains in DNA repair proteins and evolution of repair systems *Nucleic Acids Res.* 27: 1223-1242.
- Belk C.** and Borden V. Biology Science for life. Printed in USA, 2007.
- Berk A.**, Zipursky S., Baltimore D., Darnell J., and Lodish H. Molecular Biology. Fourth edition. Paul Matuataira, USA, 1998.
- Berneburg M.** and Lehmann A. 2001. Xeroderma pigmentosum and related disorders: Defects in DNA repair and transcription *Adv. Genet.* 43: 71-102.
- Cann A.** Principles of Molecular Virology. Fourth edition, Elsevier Academic Press, 2005.
- Cooper J.** and Hausman R. The cell, A Molecular Approach. Fourth edition, Sinauer Associates, 2007.
- Dale J.**, Park S. Molecular Genetics of Bacteria. Fourth edition, Wiley and Sons, 2004.
- Friedberg, E. C.**, Walker, G. C., Suede, W., 1995. *DNA Repair and Mutagenesis*. American Society for Microbiology.
- Guttmann B.**, Griffiths A., Suzuki D., Cullis T. Genetics: a beginner guide. Oneworld Oxford, 2004.
- Hanson B.**, and Jorde L. Kaplan USML Step 1 lecture notes. Kaplan Incorporation, 2001.
- Kresina T.** An Introduction to Molecular Medicine and Gene Therapy. Wiley-Liss, 2001.
- Lewin B.** Genes. VII, Oxford University Press, 2000.
- Michelson R.** and Weinert. T. 2000. Closing the gaps among a web of DNA repair disorders *Bioessays* 22: 966-969.
- Mol C.** Parikh S. Putnam C. Lo T. and Tainer J. 1999. DNA repair mechanisms for the recognition and removal of damaged DNA bases *Annu. Rev. Biophys. Biomol. Struct.* 28: 101-128.
- Murray R.**, Bender D., Botham K., Kennelly P., Rodwell V., Weil P. Harper's Illustrated Biochemistry. 28th edition, McGrawHill Medical, 2009.
- Paolella. P.** Introduction to molecular biology. First edition, WCB/McGrawHill, 1998.
- Parikh S.** Mol C. and Tainer J. 1997. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway *Structure* 5: 1543-1550.
- Turner P.**, McLennan A., Bates A., White M. Instant notes in molecular biology. Third edition, Taylor & Francis Group/UK. 2005.
- Verma P.**, and Agarwal V. molecular biology. First Edition Schand group, India 2004.

Chapter Eight

DNA recombination and Transposition

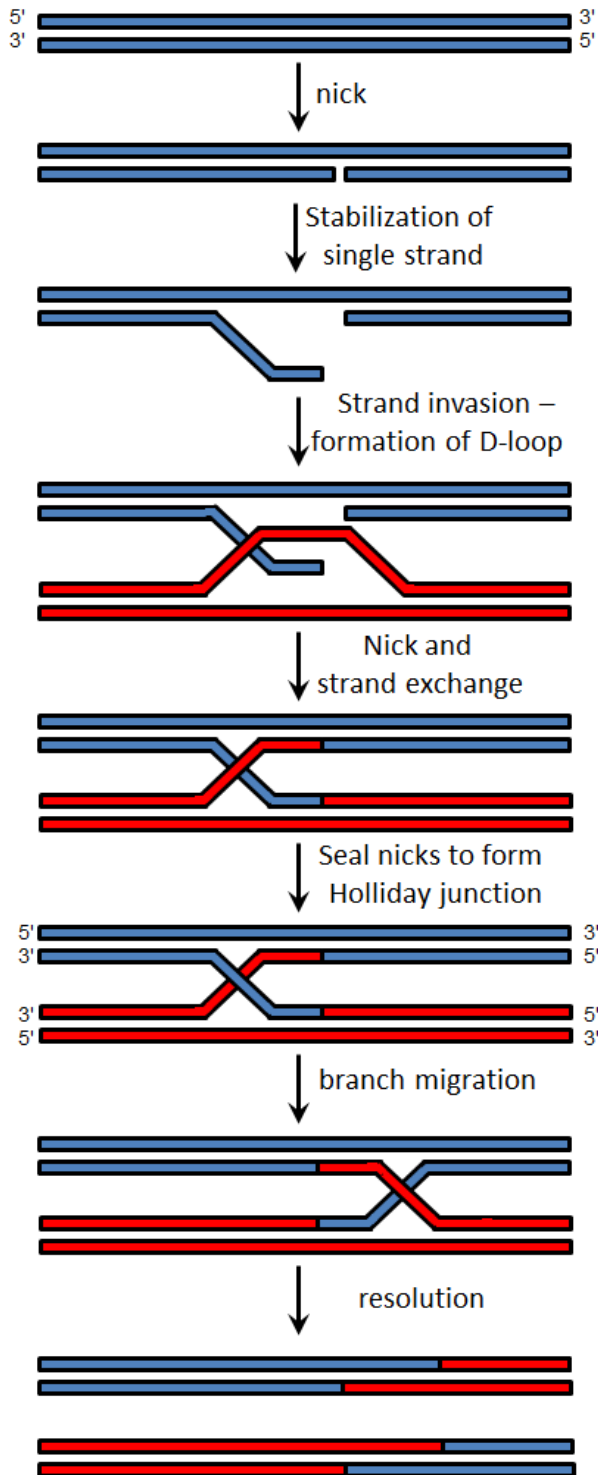


Variegated (spotted) kernels (Pierce, 2002)

Introduction

Initially, DNA should never be considered as a static molecule. Instead, these molecules have a significant dynamic movement (see chapter two). It is constantly changing, and one mechanism by which this change is brought about is recombination. Recombination is a DNA re-arrangement between two molecules of DNA having homologous sequences. Recombination provides the tools through which genetic information are reformed. This, of course, provides an evolutionary advantage to eliminate the unwanted mutations, and to maintain the spreading of the wanted mutations. Moreover, it has major role in conferring each individual a unique group of genetic information. Recombination must be taken place between precisely corresponding sequences to ensure no loss or addition for any base pair. It should for DNA extensions that are instantly recombined to be retained their original structure to act precisely. There are two types of recombination; the first type is homologous recombination (which is called general recombination) that involves genetic exchanges usually between two different DNA molecules, which participate in vast regions of homologous sequences. The second type is called site specific recombination; this is happened only at very short DNA sequences. As for DNA transposition, it participates with the first two classes in the incidence of recombination between two regions of DNA, where a short segment of DNA is transported from one location to the other counterpart in the same chromosome or in the other chromosome. But, it differs from recombination in that it's not necessary of homologous sequences to be present between two DNA fragments to be recombined together (see below). Thus, the important feature of DNA in the cells is its ability to perform rearrangement which can be varied according to the groups of genes that available in the genome of any organism. By this way, the highly important rearrangement in the survival of the organisms is caused by a genetic recombination event.

Homologous Recombination



The key to understanding the molecular processes involved in recombination was first articulated in 1964 by Robin Holliday (Holliday, 1964). His model for recombination required that a nick (cleavage at the phosphodiester bond between two nucleotides) occurred at the same site in two homologous DNA molecules (called the donor and recipient DNA). A strand exchange reaction then occurred, followed by sealing of the nicks using DNA ligase. The exchange of donor and recipient DNA strands leads to the formation of a Holliday junction – a structure in which the crossed DNA strands connect the donor and recipient DNA molecules. The Holliday junction can undergo branch migration, leading to more recombination between strands from different DNA molecules. Finally, resolution of the junction gives rise to the recombined DNA molecules. Then, this model has been developed by other scientists to give a simplified form that shown in the figure (8.1). The formation of the Holliday junction and its resolution, together with some of the *E. coli* proteins known to be involved in the process, are outlined as follows.

Figure (8.1); a simplified model for homologous recombination.

- The nick is created by the *RecBCD* endonuclease, which cleaves DNA strands at sequences called chi (χ) sites (5'-GCTGGTGG-3').
- Strand invasion is catalyzed by the *RecA* protein, along with single-strand DNA binding protein (SSB), which

stabilizes the free single DNA strand.

- The 'D-loop' that is formed after strand invasion is cut, again by the *RecBCD* endonuclease.
- Strand exchange occurs and the ends are sealed using DNA ligase. This forms the Holliday junction.
- The formation of the Holliday junction is followed by branch migration, catalyzed by the proteins *RuvA* and *RuvB*.
- The resolution of the Holliday junction requires the *RuvC* protein, which nicks two of the DNA strands. DNA ligase then seals the strands after they have been cut.

At the end of these processes, the cell will have two molecules that have exchanged DNA. The same process outlined above also takes place in eukaryotes, using similar kinds of enzymatic activity. It's noteworthy that cutting and ligation reactions here are very precise in which even one nucleotide isn't lost. But, despite of this precision, homologous recombination creates DNA molecules with new sequences.

The Role of Rec A Protein in Homologous Recombination

General recombination requires several types of specialized proteins, in particular, the *E. coli* RecA protein (or its homologs in yeasts, mice, and human) which has a central role in the recombination between chromosomes (figure 8.2). Like a single-strand DNA-binding protein (see chapter four), the RecA type of protein binds tightly and in long cooperative clusters to single-stranded DNA to form a DNA binding site, a RecA filament can hold a single strand and a double helix together. This allows it to catalyze a multistep DNA synapsis reaction between a DNA double helix and a homologous region of single-stranded DNA. The region of homology is identified before the duplex DNA target has been opened up, through a three-stranded intermediate in which the DNA single strand forms transient base pairs with bases that flip out from the helix in the major groove of the double-stranded DNA molecule (figure 8.2). This reaction begins the pairing shown previously in figure 8.2, and it thereby initiates the exchange of strands between two recombining DNA double helices. Once DNA synapsis has occurred, the short heteroduplex region where the strands from two different DNA molecules have begun to pair are enlarged through a process called branch migration.

In addition to having two DNA-binding sites, the RecA protein is a DNA-dependent ATPase, with an additional site for binding and hydrolyzing ATP. The protein associates much more tightly with DNA when it has ATP bound than when it has ADP bound. Moreover, new RecA molecules with ATP bound are preferentially added at one end of the RecA protein filament, and the ATP is then hydrolyzed to ADP. The RecA protein filaments that form on DNA could drive the branch migration reaction.

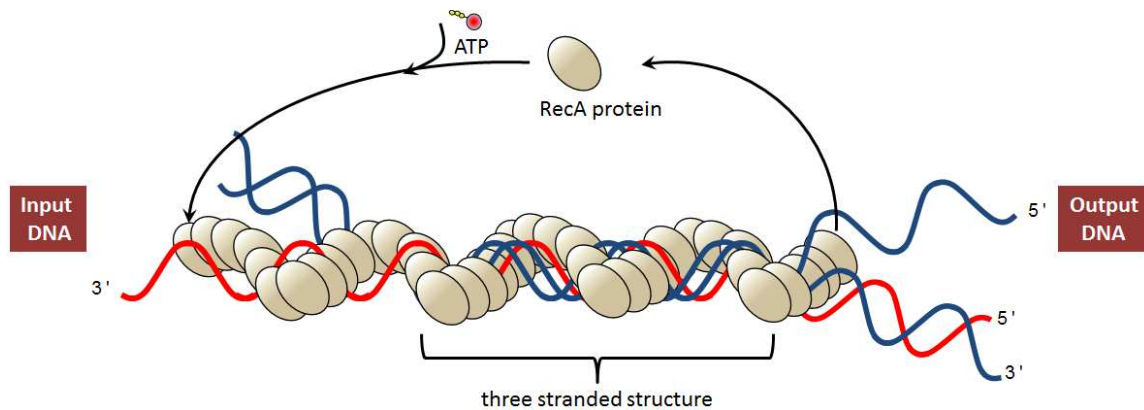


Figure (8.2); DNA synapsis catalyzed by the RecA protein. *In vitro* experiments show that several types of complex are formed between a DNA single strand covered with RecA protein and a DNA double helix (blue colour in the left side). First a non-base-paired complex is formed, which is converted through transient base-flipping (see Figure 5–51) to a three stranded structure as soon as a region of homologous sequence is found. This complex is unstable because it involves an unusual form of DNA, and it spins out a DNA heteroduplex (one strand *green* and the other strand *red*) plus a displaced single strand from the original helix. Thus the structure shown in this diagram migrates to the left, reeling in the “input DNAs” while producing the “output DNAs.” The net result is the exchange DNA strand.

Holliday Junction

Actually, the slow and difficult step in the homologous recombination is the formation of Holliday junction. After this step, the pairing region extends and there is a greater exchange of strands between the two paired DNA helices. Once it forms, Holliday junction proceeds quickly. In most cases, Holliday junction is formed, which is a key recombination intermediate, the Holliday junction (also called a cross-strand exchange) forms as a result.

In a Holliday junction, the two homologous DNA helices that have initially paired are held together by the reciprocal exchange of two of the four strands present, one originating from each of the helices. As shown in figure 8.3 A, a Holliday junction can be considered to contain two pairs of strands: one pair of crossing strands and one pair of noncrossing strands. The structure can isomerizes, however, by undergoing a series of rotational movements, catalyzed by specialized proteins, to form a more open structure in which both pairs of strands occupy equivalent positions (figure 8.3). This structure can, in turn, isomerizes to a conformation that closely resembles the original junction, except that the crossing strands have been converted into noncrossing strands, and vice versa (figure 8.3C).

Once the Holliday junction has formed an open structure, a special set of proteins can engage with the junction: one of these proteins uses the energy of ATP hydrolysis to move the crossover point (the point at which the two DNA helices are joined) rapidly along the two helices, extending the region of heteroduplex DNA (figure 8.3).

To regenerate two separate DNA helices, and thus end the exchange process, the strands connecting the two helices in a Holliday junction must eventually be cut, a process referred to as resolution. There are two ways in which a Holliday junction can be resolved. In one, the original pair of crossing strands is cut (the invading, or inside, strands in Figure 8.3A). In this case, the two original DNA helices separate from each other nearly unaltered, exchanging the single stranded DNA that formed the heteroduplex. In the other way, the original pair of noncrossing strands is cut (the inside strands in Figure 8.3C). Now the outcome is far more profound: two recombinant chromosomes are formed, having reciprocally exchanged major segments of double-stranded DNA with each other through a crossover event (Figure 8.3).

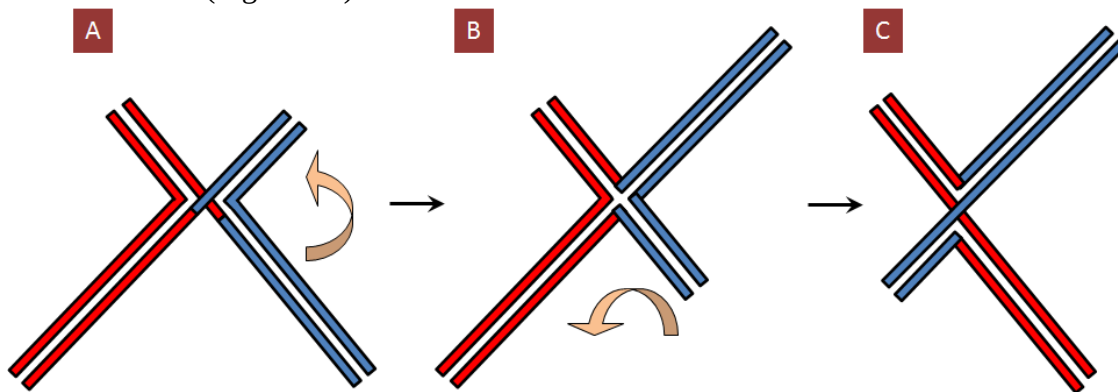


figure (8.3); Holliday junction and its isomerization.

Genetic analyses reveal that heteroduplex regions of several thousand base pairs are readily formed during recombination. As described next, the processing of these heteroduplexes—which generally consist of nearly identical paired complementary strands—can further change the information in each resulting DNA helix.

The Role of Homologous Recombination

In bacteria, homologous recombination is initially considered as a DNA repair. In this context, this process is called recombinational DNA repair. This type of DNA repair constructs the stalled replication forks at DNA replication site (see chapter seven). This type of recombination can occur during conjugation or mating, when the chromosomal DNA is transferred from the donor bacterial cell into the recipient counterpart (see chapter nine). Though the recombination during conjugation is a rare event, it participates in the genetic diversity of the wild type bacterial cell.

In eukaryotes, it seems that homologous recombination has multiple roles in the replication and cellular division, and in the repair of the stalled replication forks at the site of lesion of DNA. The recombination is taken place with the highest frequency during meiosis, the process by which the diploid germ – line cells are divided to give haploid gametes – sperm cells or oocytes in the higher eukaryotic organisms – each gamete contains only one member for each pair of chromosome. Meiosis begins by replicating DNA in the germ-line cell. Consequently, each molecule of DNA exists with four copies.

Then, the cell goes with two rounds of cellular divisions without interphase of DNA replication. This reduces the DNA content into the haploid level in each gamete.

After replicating DNA during the prophase of meiosis I, the resulting sister chromatids are remain connected at their own centromeres. In this stage, each class of homologous chromosomes are exists as a two pairs of chromatids. Now, the genetic information has been exchanged between the homologous chromatids that connected largely with each other by homologous recombination, in a process involves breaking and rejoining the DNA. This exchange, which is referred as crossing over, can be noticed under the light microscope. The genetic exchange connects the sister chromatids pairs with each other in regions called chiasmata (singular chiasma).

Thus, homologous recombination has at least three significant roles; 1) participation in the repair of many DNA damages, 2) providing – in eukaryotic cells – a temporal physical connection between the chromatids which enhances the organized separation of the chromosomes during meiosis I. and 3) increase the speed of the genetic diversity in the living organisms.

Site Specific Recombination

Unlike homologous recombination, site-specific recombination occurs when a particular DNA, that is not at all homologous to another DNA, has a sequence that acts as a target for the recombination of that other DNA. In prokaryotes, this is the case when bacteriophage λ integrates its DNA into the *E. coli* chromosome, where it will reside in the lysogenic state. This recombination event occurs between sites in the bacteriophage DNA (*attP*) that have a sequence homology with sites in the bacterial DNA (*attB*), even though the two DNA molecules are completely different in all other sequences. (figure 8.4).

The enzyme that catalyses this event is called integrase (Int), along with a host protein called integration host factor (IHF). When the bacteriophage DNA is excised from the *E. coli* genome to begin the formation of new phage particles, a phage protein called Xis is required. This site-specific recombination still requires the base pairing of a short homologous region (the *attP* with the *attB* sites), but does not involve any of the proteins of the homologous recombination pathways.

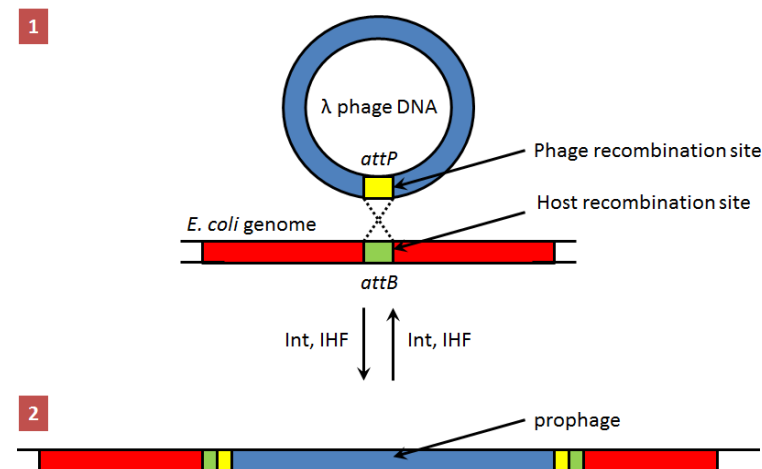


Figure (8.4); site specific recombination of λ genome in *E. coli*. In step number 1; the specialized sequences of the recombination site that belonging to the phage are aligned with their complementary recombination sequences that belong to the host. In step number 2; the phage genome is inserted in the host genome. The host integration factor (IHF) and integrase (Int) facilitate the phage integration or its release from the host genome.

Transposons

Transposons, also called transposable elements, are mobile genetic elements that generally have only modest target site selectivity and can thus insert themselves into many different DNA sites. In transposition, a specific enzyme, usually encoded by the transposon and called a transposase, acts on a specific DNA sequence at each end of the transposon - first disconnecting it from the flanking DNA and then inserting it into a new target DNA site. There is no requirement for homology between the ends of the element and the insertion site. Most transposons move only very rarely, and for this reason it are often difficult to distinguish them from non-mobile parts of the chromosome. In most cases, it is not known what suddenly triggers their movement. Thus, let us unmask these interesting genetic elements.

The term "transposons" is used for the first time at 1974 by R. W. Hedge and A. E. Jacob to describe the genetic elements that can be moved from one molecule to other one which carry the resistance to ampicillin in the bacterial cells. But this doesn't mean that these transposons are discovered for the first time at 1970s, since Barbara McClintock discovered transposons in corns at 1951, which she called the controlling elements. Despite of the fact that McClintock at that time never recognizes the fine structure of the genetic material, she was able to document – through precise notices – the movement of transposons in the genome of the corn, depending of the way of the staining of the corn seeds. Although she published paper talks about these transposons at 1948, the light was never shed on her work until the golden era of molecular biology has come at 1970s. Consequently, McClintock deserved Nobel Prize at 1983. Now, the interest of transposons has increased. This interest is exposed through the multiple terms of transposons, such as mobile elements, transposable elements, and insertion sequences. Transposons are also called junk DNA, since it doesn't has no known useful role and selfish DNA because they invest the genetic mechanisms of the cell, where they are spreading within the genome without being participating in the development of the living organism.



Barbara McClintock

We have to know that not each DNA is capable on transposition, instead, only transposons are capable to insert themselves in new locations without the help of enzymes or gene products of homologous recombination (such as recBC and recA). The target site of transposon, the sequence at which transposon insert itself on it, doesn't require any homology with the transposon. Structurally, transposons are nucleotide sequences immediately flanked on either ends by inverted repeat sequences that are themselves flanked by direct repeat sequences, for example;

```
5' ATGCA/CCGTAA[GGTTCCATTT]AATGCC/ATGCA 3'
3' TACGT/GGCATT[CCAAGGTAAA]TTACGG/TACGT 5'
```

We have here in a 5' to 3' reading frame direction a transposon in brackets [GGTTCCATTT] that is flanked by an inverted repeat sequence of CCGTAA inverted repeat to AATGCC, which is then flanked by a direct repeat sequence of ATGCA.

Transposons don't usually contain the genes that encode for their autonomous replication. They possibly considered as "a parasitic" DNA elements since they particularly depend on the host replication in their replication. They are not like bacterial replicons that contain the genes encoding for DNA replication functions. The case is the same for chromosomes, plasmids and phages.

When transposon inserted in a new target site, it moves like separate fragment of DNA. The transposon doesn't change in the size. It almost said that transposons "jump" from one location to other one, and in some cases this expression is the right one.

General Features of Transposons

Usually, most transposons have the following general features;

1. They were found to be DNA sequences that code for enzymes which bring about the insertion of an identical copy of themselves into a new DNA site.
2. Transposition events involve both recombination and replication processes which frequently generate two daughter copies of the original transposable elements. One copy remains at the parent site while the other appears at the target site (on the host chromosome).
3. The insertion of transposable elements invariably disrupts the integrity of their target genes. For example, if an IS (= insertion sequence) becomes inserted into an operon, it interrupts the coding sequence and inactivates the expression of the target gene into which it inserts as well as any gene downstream in that same operon. This is because an IS contains transcription and/or translation termination signals that block the expression of other genes downstream in an operon. This "one-way" mutational effect (or polarity) is referred to as a polar mutation.
4. It's not conditional for transposons to have negative effect on the expression of the target site. Instead, it may cause the activation of genes existing in the target site. Since transposable elements carry signals for the initiation of RNA synthesis, they sometimes activate previously dormant genes.
5. A transposable element is not a replicon - a sequence that contains a site for the origin of replication (see chapter four) - , thus, it cannot replicate apart from the host chromosome the way that plasmids and phage can (see chapter nine).
6. No homology exists between the transposon and the target site for its insertion. Many transposons can insert at virtually any position in the host chromosome or into a plasmid. Some transposons seem to be more likely to insert at certain positions (hot spots), but rarely at base-specific target sites.
7. The DNA sequences at either ends of the transposon are identical or nearly identical (as it shown above), and they are inverted. These terminal sequences are called inverted repeats or IR. The only exception for this general feature is the Mu phage. The IR is varied in size from 9 to 40 base pairs.

Transposons Classification in Prokaryotes

Transposons are classified according to their complexity in prokaryotes into simple transposons or what is known as insertion sequences (IS), and composite transposons.

Insertion Sequences (IS)

The simplest transposons are called insertion sequences, small fragments of DNA, about 1 kb in size. Each of which codes only for the proteins needed to sponsor its own transposition and don't code for additional transpositional functions (figure 8.5). Each IS element is different in sequence, but there are some common features in organization. An IS element ends in short inverted terminal repeats; usually the two copies of the repeat are closely related rather than identical.

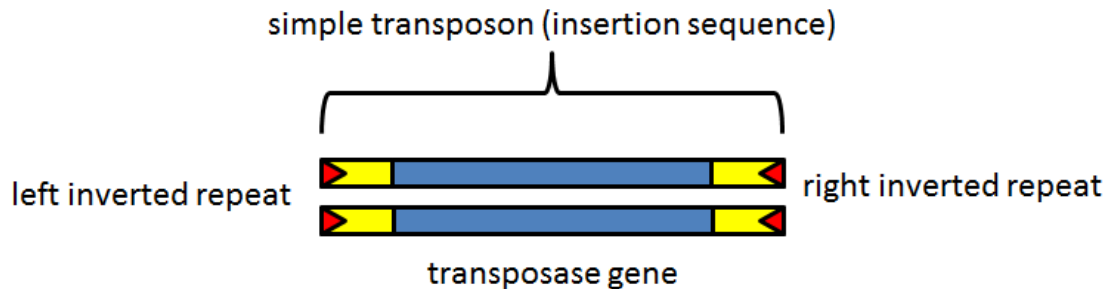


Figure (8.5); the simple transposon consists of the transposase gene that is flanked by inverted repeats.

It seems that most of IS elements transpose via conservative mode. The IS elements code for transposase functions that is responsible for the duplication of the target site and about the recognition of transposon ends. When an IS element transposes, a sequence of host DNA at the site of insertion is duplicated. The nature of the duplication is revealed by comparing the sequence of the target site before and after an insertion has occurred.

It has been shown (figure 8.6) that at the site of insertion, the IS DNA is always flanked by very short direct repeats. (In this context, "direct" indicates that two copies of a sequence are repeated in the same orientation, not that the repeats are adjacent.) But in the original gene (prior to insertion), the target site has the sequence of only one of these repeats. In the figure, the target site consists of the sequence $\begin{smallmatrix} \text{ATGCA} \\ \text{TACGT} \end{smallmatrix}$. After transposition, one copy of this sequence is present on either side of the transposon (figure 8.6). An IS element therefore displays a characteristic structure in which its ends are identified by the inverted terminal repeats, while the adjacent ends of the flanking host DNA are identified by the short direct repeats.

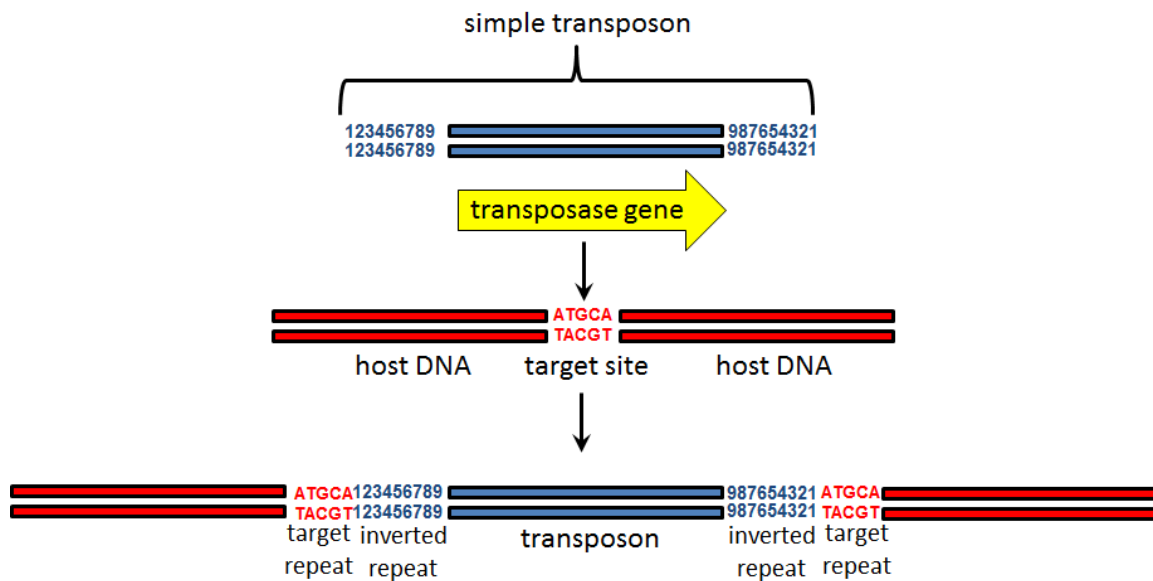


Figure (8.6); Transposons have inverted terminal repeats and generate direct repeats of flanking DNA at the target site. In this example, the target is a 5 bp sequence. The ends of the transposon consist of inverted repeats of 9 bp, where the numbers 1 through 9 indicate a sequence of base pairs.

Composite Transposons

The composite transposons have two copies of IS elements which flank a non-related DNA fragment. This foreign fragment may be resistant for drugs, or it may be represented by other marker that carried in the central region of this transposons, that is IS elements that codes for transposition functions which are like "arms" which flank them, that may be in the same or (more commonly) in the inverted orientation. Such rearrangements jump like one unit, in which it's possible for the two IS elements actually to transpose any sequence that is located between them in addition to their own transposition. These transposons are called with the symbol Tn followed by number (figure 8.7).

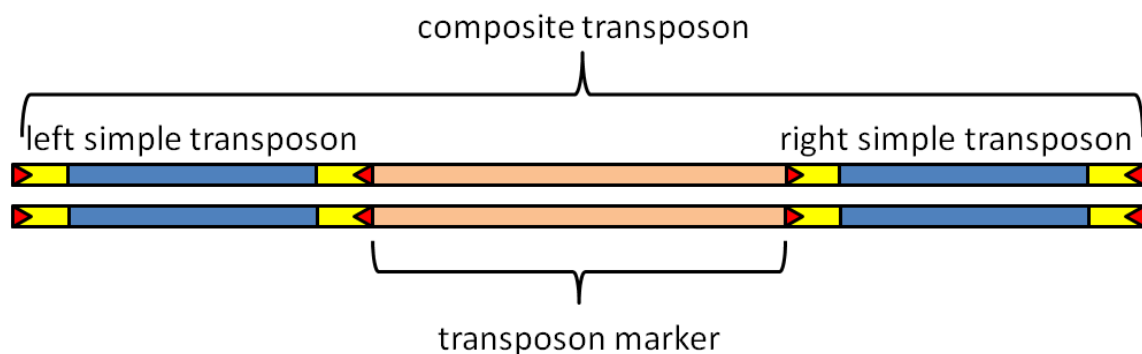


Figure (8.7); the structure of composite transposons. The composite transposon has an intermediate region carrying the markers (such as drug resistance) and this region is flanked by insertion sequences (simple transposons).

Despite the fact that composite transposons have two copies of transposase, only one copy of them is necessary for transposition. For example, in Tn10, only IS element at the right side have the transposase gene. The transposase that is encoded by other IS elements (left side) is defective. In order to transpose (jump) the composite transposon, the outer IR sequences for both IS elements are necessary for this process. As in the case of IS elements that composite transposons from which are derived, the composite transposons jump through conservative (non-replicative) mechanism.

Transposons Classification in Eukaryotes

In eukaryotes, transposons are classified into two classes according to the mechanism through which the transposon jump from the donor molecule into the recipient counterpart. The first class jumps through a copy of RNA, whereas the second class jumps directly.

Class I

The class I can be called RNA elements because of the transcription of DNA that exist in the genome of the donor site into RNA copy. The RNA copy is reversely converted into a DNA copy again, but this copy is inserted into the recipient (target) site (figure 8.8). It's possible for class I transposons to be called retro-elements because of its movement that is recognized in its reversal flow of genetic information, from RNA to DNA in a process called retrotransposition. There is a vast number of class I copies in the host genome. For example, there is a millions of copies of class I transposons that called *Alu* family in the human genome (see chapter two). Since the class I transposons need an mRNA copy as intermediate copy in their own transposition while they are never perform transposition by themselves, therefore, the persistence of these transposons insertion can be noticed in the genome. This means, these transposons cannot be excised from the donor site. Nevertheless, these sequences are considered as transposons for the ability of their copies to be inserted in new sites. The prototype of this class is Ty elements that are available in yeasts.

Class II

In contrast with the first class, the transposons of the second class are called DNA elements, since they move by themselves from one genomic site into the other one (figure 8.8). in contrast also with the class I transposons, it's possible for the class II transposons to be excised for the donor site. It's noteworthy that the first transposon that discovered in the corn belongs to this class. This transposon causes unusual variegated (spotted) kernels of corn because of the excision or cutting to the gene that is responsible about the production of kernels.

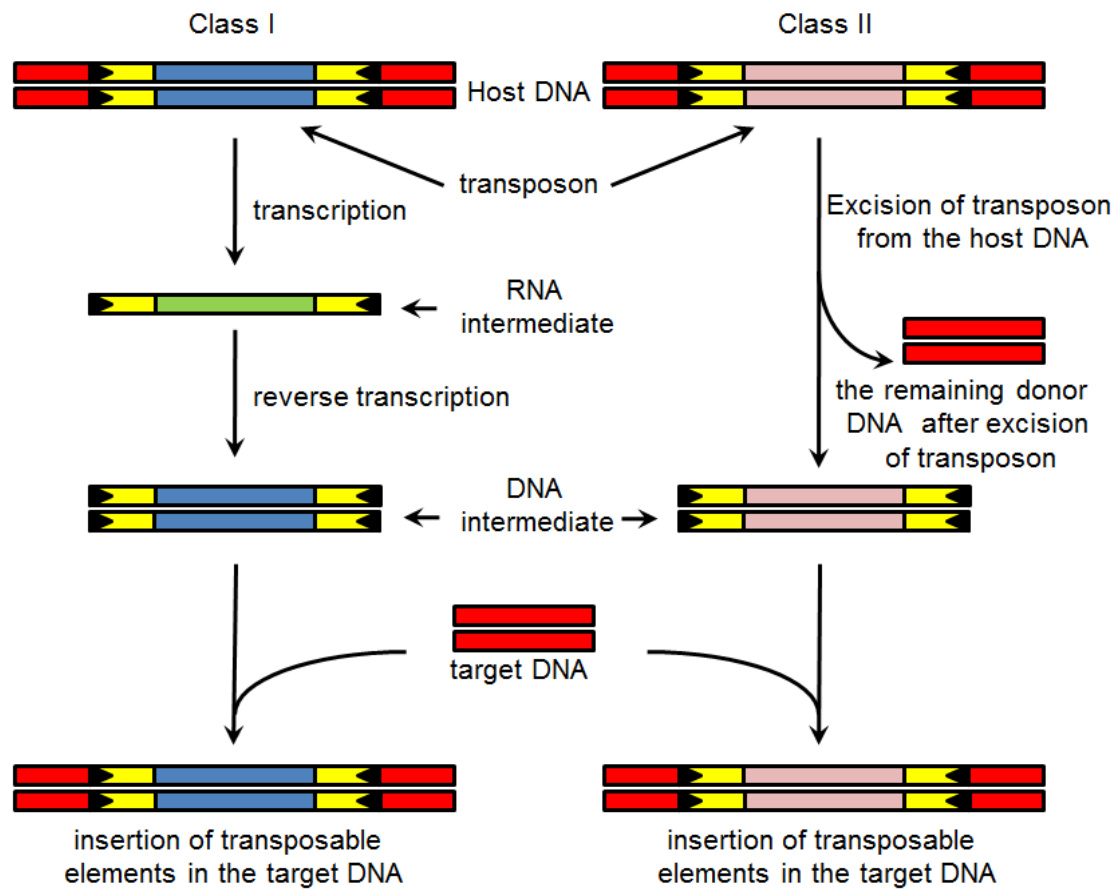


Figure (8.8); illustration shows two classes of transposons in eukaryotes. Class I transposons remains permanent once it's being inserted. It's possible to call class I sequences with transposons since the resulting RNA copies are submitted to reverse transcription into DNA which inserts into a new genomic site. The class II moves by removing itself from one site and reinserting itself into another site.

Mechanisms of Transposition

Usually, there are two mechanisms transposons used to, whether in prokaryotes or in eukaryotes, in the transposition through its jumping from one site to the other site.

1. Direct transposition; which occurs either through replicating or cointegrate-forming form, or through conservative (non-replicating) form. In transposition that depends on DNA intermediate is either replicative or non-replicative. In replicative transposition, a new copy of in a new site is entered while the old copy is being in the old site. Consequently, the number of the transposable element is increased. In the non-replicative transposition, the transposable element is excised from the old site then inserted into a new site without any increasing in its copy number. The non-replicative transposition requires few nucleotides duplication where constitute direct repeats (see below).

2. Non-direct transposition; the transposition mechanism through RNA intermediates (see below).

Direct Transposition Mechanism

1. Replicative Transposition

It's sometimes called copy-and-paste transposition, which can be either between two different DNA molecules or between two parts of the same DNA molecule. The figure (8.9) shows the steps of transposition between two circular DNA molecules. Before transposition, the transposable element is on one molecule. In the first step, the two DNA molecules are joined, and the transposable element is replicated, producing the cointegrate structure that consists of the donor and recipient molecules fused together (figure 8.9). After the cointegrate has formed, crossing over at regions within the transposable elements produces two molecules, this second step is known as resolution of the cointegrate.

How are the steps of replicative transposition (cointegrate formation and resolution) brought about? Cointegrate formation requires four events. First, a transposase enzyme (often encoded by the transposable element) makes single strand breaks at each end of the transposable element and on either side of the target sequence where the element inserts. Second, the free ends of the transposable element attach to the free ends of the target sequence. Third, replication takes place on the single-stranded templates, beginning at the 3' ends of the single strands and proceeding through the transposable element. This replication creates the cointegrate, with its two copies of both the transposable element and the sequence at the target site, which is now on one side of each copy (figure 8.9). The enzymes that perform the replication and ligation functions are cellular enzymes that function in replication and DNA repair. Fourth, after the cointegrate has formed, it undergoes resolution, which requires crossing over between sites located within the transposon. Resolution gives rise to two copies of the transposable element. The resolution step is brought about by resolvase enzymes (encoded in some cases by the transposable element and in other cases by a cellular gene) that function in homologous recombination.

In this replicative mechanism, in which the transposon pairs during the transposition: only one copy is move, but one copy remains. Therefore, the original host DNA is never demolished. In this case, second cuts are not generated. Instead, the second transposon strand is stay in order to be copied by DNA polymerase, which starts at the end of the first cut. The resulting structure is called cointegrate, its existence is a strong clue for such pathway of transposition. Then, this structure is resolved by the action of other transposon coded enzyme that is called resolvase. It's possible for resolvase to induce site specific recombination through two copies of transposons in a reaction resembles the integration of λ phage, producing separated DNA molecules, each of which has only one copy of transposon. The results of this two steps transposition reaction include two copies of transposon, one copy in its original site of the donor replicon, and the other is on the target replicon. The members of Tn3 family (simple transposons), Tn1, and Tn501 are examples of this group.

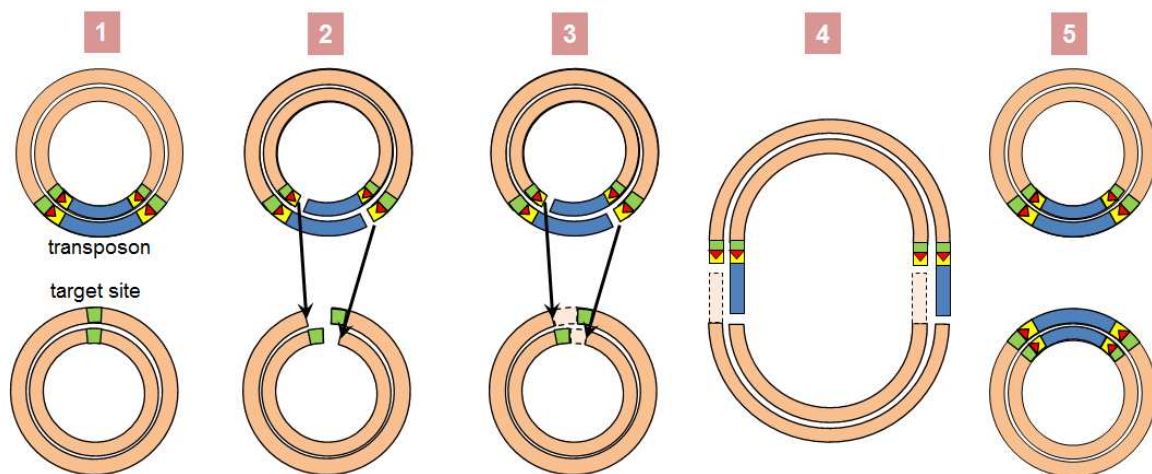


Figure (8.9); replicative transposition. In this type of mechanisms, DNA sequences of transposons are taken place by DNA replication. The end products are represented by a DNA molecule that is homologous to the donor and the target DNA on which the transposon is inserted. These steps includes; 1) transposase aligns the transposon sequence with the target sequence together, 2) this enzyme makes a staggered cuts (see chapter ten) for both transposon and target site, 3) DNA polymerase synthesizes new strand in the target site, 4) the transposon synthesis continues by DNA polymerase after the formation of the cointegrate, 5) finally, the resolvase resolve this cointegrate to produce two molecules, each of which contains the transposon.

2. Non-replicative Transposition

In this type of transposition, the transposon element moves from one site to another one without replicating the transposon element completely (figure 8.10), and this occurs despite of the replication of short DNA sequences in the target DNA, generating flanking direct repeats.

Despite of the ability of transposase to cut DNA strands, it cuts the DNA in a unique and different way of endonucleases that are used in the genetic engineering experiments. Transposase binds either transposons ends in addition to the target DNA sequence on which it inserts, then, the enzyme makes staggered cuts in the same way by which endonucleases do in the target site, in addition to its cutting to the either transposons ends simultaneously.

But, the different thing here is that transposase make these cuttings in different strands, a strand from the donor DNA and other strand from the recipient DNA, while endonucleases used in the genetic engineering experiments do these cuttings in one DNA molecule, in the either strands of the same molecule (see chapter ten). The class II transposons perform in eukaryotes rely on this mechanism. In prokaryotes, Tn5, Tn9, and Tn10 transposons relies on this mechanism as well.

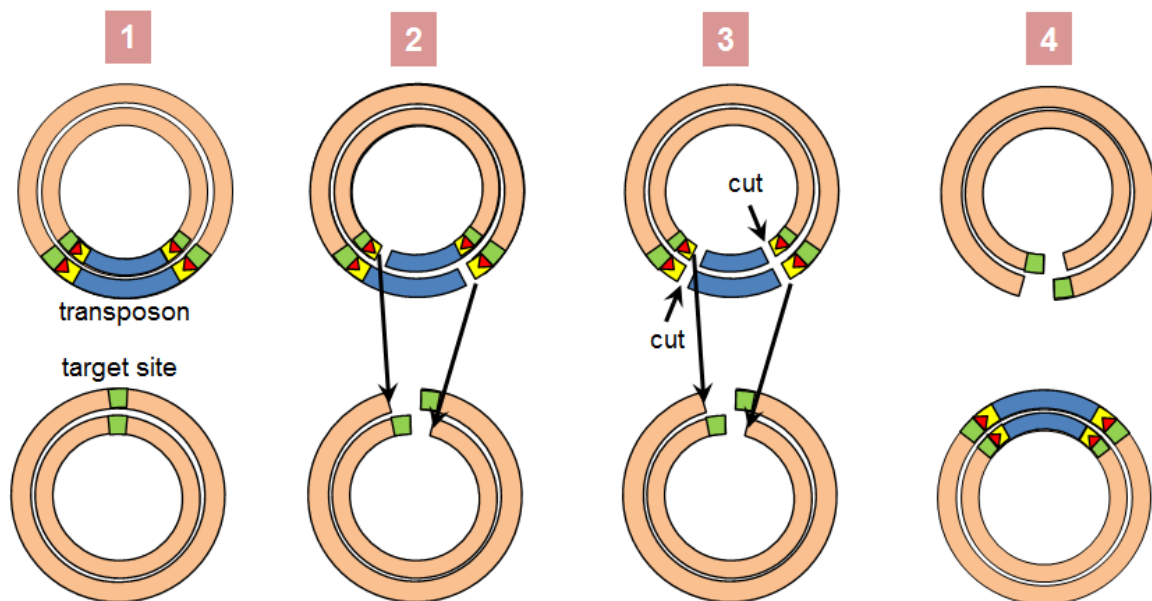


Figure (8.10); non-replicative (cut and paste) transposition. The end products are represented by DNA molecule homologous to the donor and the target DNA on which the transposon is inserted. These steps includes; 1) transposase aligns the transposon sequence with the sequence of the target site together, 2) this enzyme makes staggered cuts (see chapter ten) for both transposon and target site, 3) second group of cuts undertake the separation of transposon from its old sequence, 4) finally, the transposon containing molecule is losses the transposons, and the later is inserted in the new target sequence.

Indirect Transposition Mechanism

Transposition via RNA intermediates

Eukaryotic transposable elements that transpose through RNA intermediates are called retrotransposons. A retrotransposon in DNA is first transcribed into an RNA sequence (see figure 8.8), which may be processed. The processed RNA undergoes reverse transcription by a reverse transcriptase enzyme to produce a double-stranded DNA copy of the RNA. Staggered cuts are made in the target DNA, and the DNA copy of the retrotransposon inserts into the genome (see figure 8.8). Replication fills in the short gaps produced by the staggered cuts, generating flanking direct repeats on both sides of the retrotransposon.

There are several examples for this mechanism; many experiments have shown that some eukaryotic transposons, especially these that belong to retroviruses, move in very different way compared with bacterial transposons such as Tn3. Thus, Ty elements of yeasts will be transcribed first into RNA, the later is transcribed into DNA by reverse transcriptase. Then, the DNA copy is recombined into a new site of the host chromosomes. Not all Ty elements are the only elements that transpose by this mechanism; instead, retroviruses have single stranded RNA that replicated through double stranded DNA intermediate. The life cycle of the virus involves an obligatory stage in which the double stranded DNA is inserted into the host genome by a transposition-like event that generates

short direct repeats of target DNA. The crucial steps are that the viral RNA is converted into DNA, the DNA becomes integrated into the host genome, and then the DNA provirus is transcribed into RNA. The enzyme responsible for generating the initial DNA copy of the RNA is reverse transcriptase. The enzyme converts the RNA into a linear duplex of DNA in the cytoplasm of the infected cell. The linear DNA makes its way to the nucleus. One or more DNA copies become integrated into the host genome. A single enzyme, called integrase, is responsible for integration.

We have to know some facts correlated with the general mechanism of transposition, such as; 1) the transposons move precisely, they carry all sequences integrated into the old site and not in the adjacent DNA. The transposase recognizes either ends of transposons. These ends are almost homologous since they are the common binding sites for the transposase. 2) Precise replication made of 3 to 12 bases of target DNA (according to the type of transposon) is formed, and only one copy of short sequence on the either sides of the inserted transposon (see figure 8.8). This may be attributed to the sort of cutting the transposase made (see below). Therefore, some DNA synthesis reactions occur during transposition to lead to the replication of target site. This is an obvious difference from site specific recombination for λ phage in which the replication of target site is never takes place. 3) Despite that most transposons move for any region in the new genome, but they never entirely move randomly, favoring invasion of certain DNA sequences. Some of them attack specialized tetranucleotides or hexanucleotides of specialized symmetrical DNA sequences as edonuclease do (see chapter ten).

Chapter Eight Appendix

The Advantages of Transposons

When transposable elements are discovered by Barbara McClintock at 1950s, they were considered as genetic elements that can cause unstable mutations, to produce chromosomal aberrations (that include breaks) that change the position in the genome. When rediscovered in bacteria at 1970s "as insertion sequences", it was noticed that transposons move from one genetic site to other site and cause premature transcriptional termination in the bacterial operons.

The largest transposons often contain genes confer evolutionary benefits for the transposon having cell which is the resistance for antibiotics that are the most popular, and resistant to mercury, moreover, toxins production genes and fermentation genes are found. Add to that, transposable elements facilitate horizontal gene transfer between non-relative organisms. Transposons cause mutation in the bacteria either by its insertion in the gene coding region, losing some of the genes in the process, or by its insertion upstream of gene coding region in an important region in the determination of gene expression, as in the region of the DNA binding with the transcription factor. The insertion in the gene can cause alternative splicing causing into a greater variation in the proteins that are resulted from single gene. Transposons can cause replication of fragments of DNA, which is the first important stage in the evolution of new information. Moreover, transposons can cause increasing in the genome size, since they leave multi-copies of their own in the genome. Transposons may play a role in causing diseases including cancer and hemophilia.

The development of multiple resistance plasmids become easy by the ability of transposons to recombine without homology in the sequence. Thus, these are collected with each other in a single plasmid, which ensure their ability to survive through the multiple using to multiple antibiotics with each others. The R1 plasmid that is existing in the nature and evolved from the transposons encoding to resistance to chloramphenicol (Cm) and kanamycin (Km). add to that, it contains the intact transposon Tn4 which carries the resistance to streptomycin (Sm), sulfonamide (Su), and ampicillin (Amp) (figure 8.11). the Tn4 contains itself the resistance for ampicillin. The entire assembly of R – determinants bind with the other fragments

containing genes for plasmid replication and for transfer function, that induce the conjugate contact of the bacterial cell with the other bacteria. This permit the plasmid the ability to transfer quickly among cells. The plastic nature of this assembly gives many plasmids variations in both laboratory and nature.

Most of the transposable elements can be inserted into the bacterial chromosomes at several locations. If the transposon inserted within the gene, the continuity of this linearity for this gene can be dispersed and it function is lost. Thus, the transposons can be considered as "biological mutagenes" in contrast with "chemical mutagenes" as in nitrosoguanadine. The thing that distinguishes this biological mutagenes is that each mutant is rarely suffering more than one mutation. Thus, it's possible to avoid the multiple mutations problem that is connected with chemical mutagenes. Consequently, these mutations have great advantages in multiple types of genetic manipulations in bacteria.

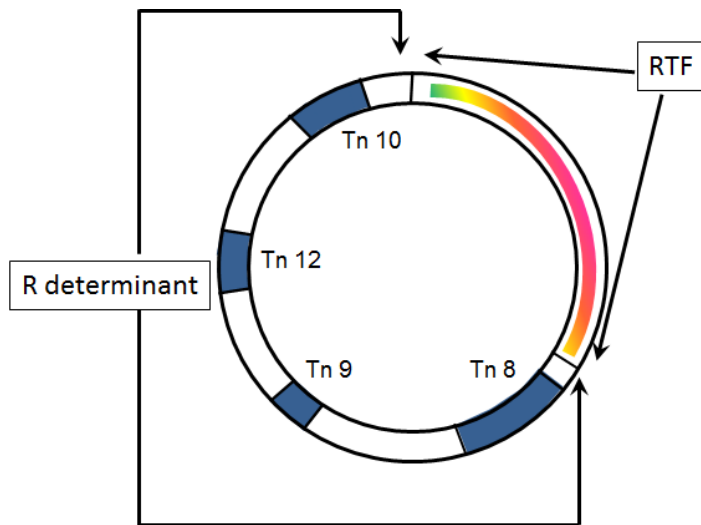


Figure (8.11); the structure of R plasmid; the RTF means resistance transfer factor that carries the genes of transfer. The R determinant carries the resistance genes.

Chapter Eight Revision Questions

Q1- Justify the following:

1. It's possible for recA protein to drive the branch migration reaction in recombination?
2. The sperm cell or the mature oocyte don't have the same information of the body they resulting from?
3. The positions that transposons recognize to bind with are characterized as "modest" positions?
4. The target site sequence replicates on the either sides of the transposon?
5. The formation of a structure called cointegrate is considered as clue for transposition?
6. Retroviruses are characterized as diploid?
7. The transposable elements facilitate the horizontal gene transfer between non-related organisms?
8. Transposons are called sometimes as biological mutagenes, and they are favored in comparison with the chemical mutagenes?
9. Both of transposase and restriction enzyme make staggered cuts in the DNA, but, the first enzyme is used in transposition of DNA while the second enzyme is used in the cutting of DNA?

Q2- Choose the correct answer:

1. To ensure recombination event between any two DNA molecules, the must be available.
a. insertion sequences b. inverted sequences c. heterologous sequences d. homologous sequences
2. The recombination between two molecules of DNA with a loss of one or more nucleotide pairs is called
a. homologous recombination b. non-homologous recombination c. site specific recombination d. general recombination

3. The formation of D-loop is not restricted to recombination alone, but it can be noticed in certain replicative structures, such as
- a. chloroplast b. Golgi apparatus c. plasmids d. mitochondria
4. The step is the most exciting recombination steps, since to what its require to rotational movement to occur.
- a. strand invasion b. D-loop formation c. Holliday junction d. single stranded stabilization
5. The site specific recombination between λ phage and E. coli is taken place when the phage being in the
- a. lytic stage b. lysogenic stage c. both lytic and lysogenic stages d. non
6. The proteins participate in the site specific recombination.
- a. RecBCD c. RecA c. RuvB d. all e. non
7. In order to be transported from one position to another, transposons usually require sequences.
- a. long region of homology b. intermediate region of homology c. short region of homology d. no region of homology
8. Special type of topoisomerase type I is participate in
- a. homologous recombination b. site specific recombination c. transposition d. non
9. All transposons have inverted sequences at both ends except
- a. Mu phage b. Tn 10 c. Tn 9 d. Ty elements
10. The mechanism by which the transposon recognizes the target site is
- a. completely random b. completely specific c. similar to restriction enzymes d. all
11. is one of the transposition mechanisms, in which the donor DNA is demolished.
- a. non-replicative transposition b. replicative transposition c. retrotransposition d. non

Q3- Define the following:

Holliday junction, Chi sequence, chiasmata, junk DNA, reverse transcriptase

Q4- Connect terms exist in right with its suitable counterparts in left

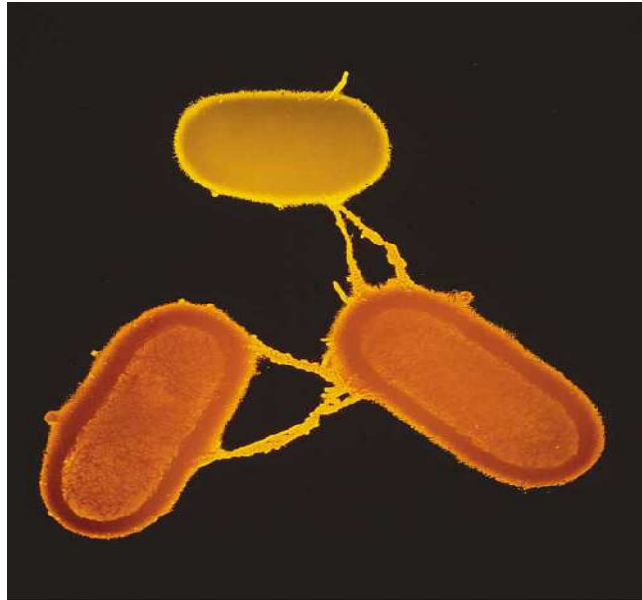
RecA	nick
RuvC	strand invasion
RecBCD	D loop
RuvA	single stranded stabilization
DNA ligase	Holliday junction
SSBs	Branch migration
RuvB	resolution

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Craig N.** Transposition, in *Escherichia coli and Salmonella*, Washington, DC: ASM Press, 1996.
- Craig N.** (1997). Target site selection in transposition *Annu. Rev. Biochem.* 66: 437-474.
- Gopaul D.** Guo F. and Van Duyne G. 1998. Structure of the Holliday junction intermediate in Cre-loxP site specific recombination *EMBO J.* 17: 4175-4187.
- Griffiths A.**, Wessler S., Lewontin R., Gelbart W., Suzuki D., Miller J. An Introduction to Genetic Analysis. Eighth edition, 2004.
- Kumar H. D.** Molecular genetics and biotechnology. Second edition, Vikas Publishing House, India, 1997.
- Lewin B.** Genes. VIII, Pearson Prentice Hall, 2004.
- Mizuuchi K.** (1992). Transpositional recombination: mechanistic insights from studies of mu and other elements *Annu. Rev. Biochem.* 61: 1011-1051.
- Murray et al.** Medical Microbiology. Third edition, 2005.
- Pierce B.** Genetics: A conceptual approach. 2002
- Passarge E.** Color atlas of Genetics. Second edition, Thieme, 2001.
- Reece R.J.** Analysis of genes and genomes. John Wiley & Sons, 2004.
- Schleif R.** Genetics and molecular biology. Second edition, The Johns Hopkins University Press, 1993.
- Turner P.**, McLennan A., Bates A., White M. Instant notes in molecular biology. Third edition, Taylor & Francis Group/UK. 2005.
- Verma P.**, and Agarwal V. molecular biology. First Edition Schandgoup, India (2004).
- Watson J.**, Baker T., Bell S., Gann A., Levine M., Losick R. Molecular biology of the gene. Fifth edition, Pearson Prentice Hall, 2004.
- Robinson R.** Genetics. Volume 3, MacMillan Reference USA, 2003.

Chapter Nine

Systems of DNA Transfer in Bacteria



More than 11 folds magnified picture taken by transmission electron microscope (TEM) of conjugated *E. coli* cells, in which one male cell (right below) with either two female cells (Griffiths *et al.*, 2004).

Introduction

The bacterial world and their viruses are not distant from molecular genetics, in such away a large part of this science is involved in the studying these prokaryotic organisms intensively and revealed many basics on which molecular biology is built. Although bacteria have genes composed of DNA arranged in a long series on a “chromosome,” their genetic material is not organized in the same way as that of eukaryotes (see chapter one). Bacteria belong to a class of organisms known as prokaryotes. Viruses are also very different from the organisms we have been studying so far. Viruses share some of the properties of organisms; for example, their genetic material is DNA or RNA, constituting a short “chromosome.” However, most biologists regard viruses as nonliving because they cannot grow or multiply alone. To reproduce, they must parasitize living cells and use the molecular machinery of these cells. The viruses that parasitize bacteria are called bacteriophages, or simply phages.

This chapter focuses on the understanding of the principles of bacterial genetics. This may be done through focusing on the mode of transfer of the genetic material between these organisms, which are transformation, conjugation, and transduction (figure 9.1), because these processes are major source for the great genetic variability in bacteria. Even though bacteria and phages do not undergo meiosis, the genetic variability that occurs in bacteria is a consequence of one of these processes. Therefore, bacteria diversify their genetic material by horizontal gene transfer, in which the bacteria uptake DNA from other organisms by these means. This is considered as very important method in bacterial evolution.

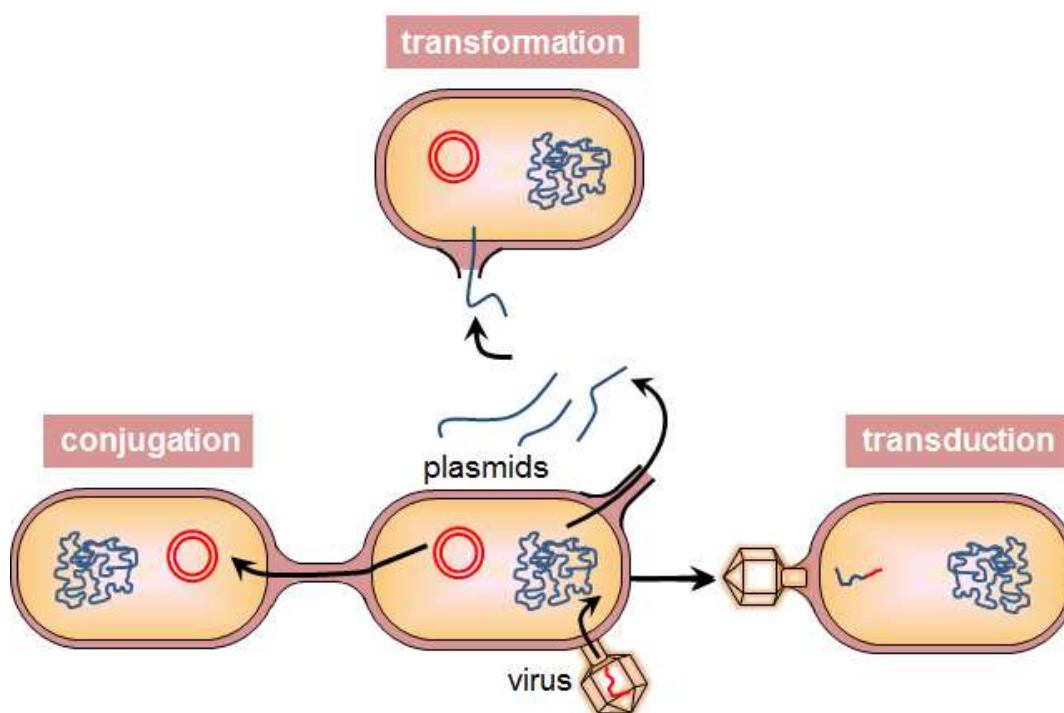


Figure (9.1); three ways by which DNA can be transferred between bacterial cells. The transformation method is shown in the upper portion of this figure, while the transduction method that is mediated by phage is shown in the right portion, and the conjugation method is shown in the left portion of this figure.

In bacterial conjugation, the bacterial cell binds with another one directly. In bacterial transformation, the bacterial cell can uptake a piece of DNA from the environment and converts it inserts it into the chromosome after its entry. In transduction, certain phages uptake the DNA from a bacterial cell and inject it into another cell, in which it's possible for this piece to be inserted into the chromosome (figure 9.1). Before we take the three major methods by which DNA can be transferred between bacterial cells, we must know an important concept which is related to microbial genetics which is the plasmids.

Plasmids

The bacterial genome consists of a chromosome in a single circular molecule. In addition to the chromosome, many bacteria have circular molecules known as plasmids. The term "plasmid" is originally used by Lederberg at 1952 to describe extra chromosomal genetic elements which have been developed to be: extra-chromosomal genetic elements have the ability to replicate autonomously. That is, they are autonomous units, and stably inherited. These molecules exist in Gram positive and Gram negative bacteria in addition to some yeasts and fungi. Most plasmids are covalently closed circular double stranded DNA molecules (figure 9.2).

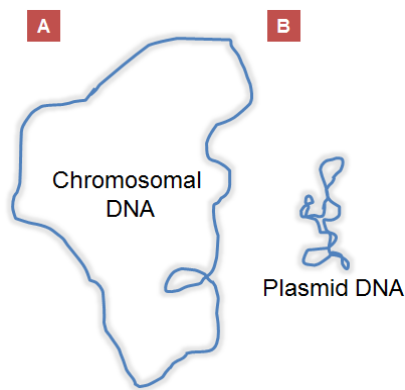


Figure (9.2); (A) big bacterial chromosome (the big circle) with, (B) it's plasmid (the small circle) as it's seen under electron microscope.

Some plasmids exist with only one copy per cell, while other plasmids in a multicopy form. The first type of control is called single copy control; which resembles the bacterial chromosome and replicates only once per each generation. The single copy plasmid maintain efficiently on its homology with the bacterial chromosome. The multicopy control systems allow for several replications per one generation, this result in several copies of plasmids for each bacterium. The multicopy plasmids exist in a noticeable number (typically 10 to 20 copies) per each bacterial chromosome. It has to be note that low copy number plasmids are under more strict regulation for their replication than the high copy number plasmids.

Some plasmids and episomes are transferred between cell by a conjugative process (which includes a direct contact between donor cells and recipient cells). The transfer feature has sometimes bacterial host genes are transferred with plasmid DNA (or episome's DNA). Thus, these events play a major role in the exchange of the genetic material between bacteria. In general, plasmids carry the genes that are not essential for the bacterial function, but they play important role in the life cycle and the growth of bacterial hosts. Some plasmids speed the mating between bacteria, and they called F factors (see below), whereas the other carry genes that kill other bacteria, which called Col factors (some bacterial strains produces bacteriocin that can kill there competing strains) (figure 9.3).

The bacteriocin that kills *E. coli* strains is called colicin, which is synthesized only from these strains that carry specialized plasmids: these strains called colicinogenic,

and the plasmids called Col factors. It's possible for some plasmids, as in *Col* E1, to be transferred for other host cells by conjugation.

Some of antibiotics resistant bacteria have plasmids called R-factors that carry resistance for some antibiotics such as streptomycin, sulphonamide, and tetracycline. One of the most important features that make the plasmids in the superior of microbiology is their ability to carry and transfer the genes that codes for the resistance of antimicrobial agents. This type of plasmids is spreading in the bacteria and can be transferred between different types of microorganisms. This is the genetic feature that represent very dangerous feature in human and animals. These plasmids called R plasmids that contain many of genes that code for the resistance for wide range of antimicrobial agents such as ethidium bromide. From the structural aspect, these plasmids are conjugative, in which replication and transfer genes are located in certain part of them, while the resistance genes are located in another part (see the appendix of this chapter).

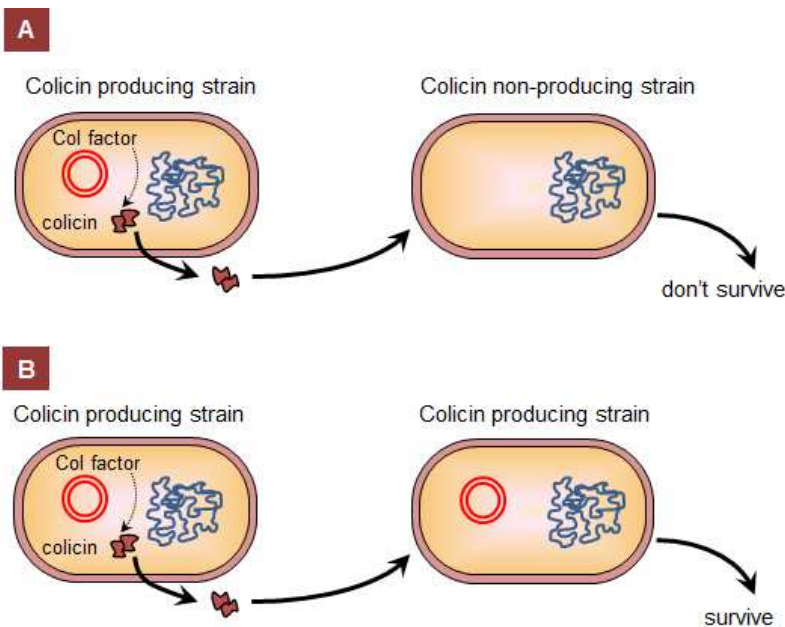


Figure (9.3); rare production of colicin, rendering the environment toxic for plasmid-free cells. Although synthesis and release of colicin is invariably lethal, plasmid containing cells are immune to external colicin.

The important thing to be mentioned in the application field is the wide utilization of some plasmids in the genetic engineering experiments (see chapter eleven). Although some plasmids have hundreds thousands of base pairs, most of plasmids have circular form with a length of only several thousands of base pairs. The possession of plasmids to origin of replication made them capable on replication independently of the bacterial chromosomes (see chapter three).

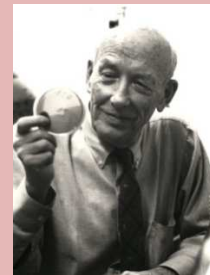
First: Conjugation

It is the transfer of DNA from the donor to the recipient by direct physical contact between cells. In bacteria, there are two types of mating between the donor (male) and the recipient (female), and the direction of the transfer of genetic information is only one direction, which is the transfer of DNA from the donor to the recipient. Conjugation differs from the sexual reproduction in two essential considerations; the first one, the genetic participation from parents is non-equal. Therefore, it refers to the parents with donor cells and recipient cells. The second consideration, the resulting offspring resemble only to the recipient cell and not the donor cell in most features.

The excited experiment that performed by Joshua Lederberg and Edward Tatum at 1946 was answered the question that is asked from time to time whether bacteria possess any processes similar to sexual reproduction or not. They discovered a sex-like process in bacteria. Their experiment was summarized by using of two strains of *Escherichia coli* with different sets of auxotrophic mutations. Strain A would grow only if the medium were supplemented with methionine and biotin; accordingly it was strain A is called (*met- bio-thr+ leu+ thi+*), while strain B would grow only if it were supplemented with threonine, leucine, and thiamine; accordingly, strain B was called (*met+bio+thr-leu-thi-*). The figure (9.4) shows the details of this experiment.



Joshua Lederberg



Edward Tatum

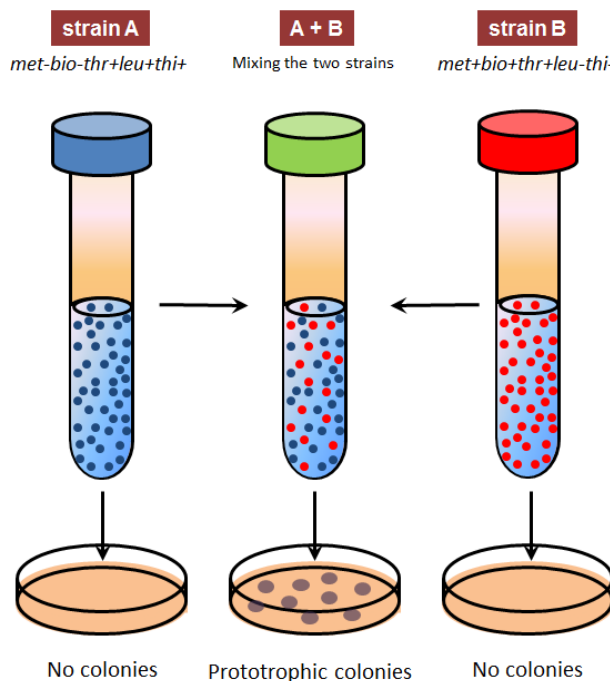


Figure (9.4); Lederberg and Tatum's demonstration of genetic recombination between bacterial cells. The basic concept: two auxotrophic cultures (A and B) are mixed, yielding prototrophic wild types (WT). (b) Cells of type A or type B cannot grow on an unsupplemented (minimal) medium (MM), because A and B each carry mutations that cause the inability to synthesize constituents needed for cell growth. When A and B are mixed for a few hours and then plated, however, a few colonies appear on the agar plate. These colonies derive from single cells in which an exchange of genetic material has occurred; they are therefore capable of synthesizing all the required constituents of metabolism.

Strains A and B were mixed together, incubated for a while, and then plated on minimal medium, on which neither auxotroph could grow. A small minority of the cells ($1 \text{ in } 10^7$) was found to grow as prototrophs and hence must have been wild type, having regained the ability to grow without added nutrients. Some of the dishes were plated only with strain A bacteria and some only with strain B bacteria to act as controls, but from these no prototrophs arose (figure 9.4). Lederberg and Joshua suggested that some form of recombination of genes had taken place between the genomes of the two strains to produce the prototrophs.

This story was not end, since Bernard Davis noticed that the two strains were never exchange genes between each other in the form imagined by Lederberg and Tatum. Instead, certain materials bound with these cells achieve this. He proved this by his contraction for a U-tube in which the two arms were separated by a fine filter. The pores of the filter were too small to allow bacteria to pass through but large enough to allow easy passage of any dissolved substances (Figure 9.5). Strain A was put in one arm, strain B in the other. After the strains had been incubated for a while, Davis tested the contents of each arm to see if there were any prototrophic cells, but none were found. In other words, *physical contact* between the two strains was needed for wild-type cells to form. It looked as though some kind of genome union had taken place, and genuine recombinants produced. The physical union of bacterial cells can be confirmed under an electron microscope, and is now called conjugation.

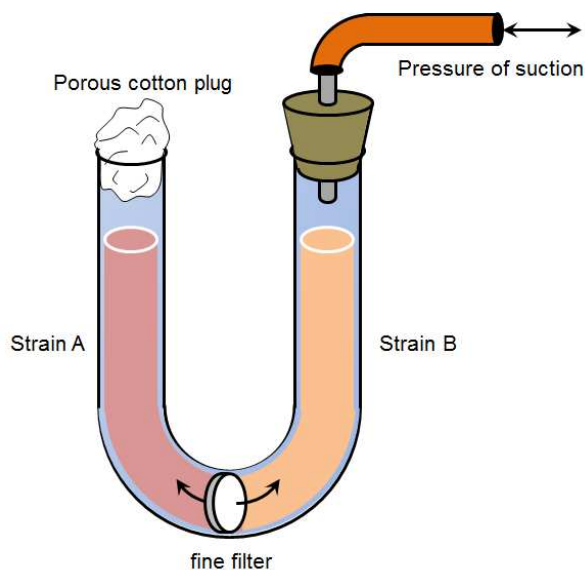


Figure (9.5); Davis experiment to prove the necessity of Physical contact between bacterial for genetic recombination. Auxotrophic bacterial strains A and B are grown on either side of a U-shaped tube. Liquid may be passed between the arms by applying pressure or suction, but the bacterial cells cannot pass through the filter. After incubation and plating, no recombinant colonies grow on minimal medium.

The bacterial conjugation is done by F plasmids, that are classic example for episomes, the elements that can exist either a free circular plasmids or integrated as a linear sequence (as in the lysogenic bacteriophages) in the bacterial chromosome. The F factor can integrate in multiple locations of the *E. coli* chromosome, and when the factor (the plasmids) is free, it can be preserved as one copy per each bacterial

chromosome. When it's integrated into the bacterial chromosome, its genes are submitted to repression and its DNA (F DNA) replicates as a part of the chromosome.

When F plasmid present, whether free or integrated, it has important consequences for the host bacteria, since the bacteria that have fertility factor (F^+ bacteria) are able to recombine (to mate) with the bacteria that loss this factor (F^- bacteria). The conjugation includes direct contact between the donor bacterium (F^+), and the recipient bacterium (F^-), followed by the transfer of F factor. If the F factor is exist as a free plasmid in the donor bacterium, at which its transferred as a plasmid, and this process converts the recipient cell that don't have F factor (F^- bacterium) into a recipient cell having F factor (F^+ bacterium).

There is a large region of the F plasmid, called the transfer region (abbreviated as *tra* region), is required for conjugation. It contains about 40 genes that are required for the transmission of DNA. F-positive bacteria possess surface appendages called pili (singular pilus) that are coded by the F factor. At least 12 *tra* genes are required for the modification and assembly of pili that resemble the hair in structure. The conjugation starts from the top of pilus that binds at the surface of the recipient cell (figure 9.6).

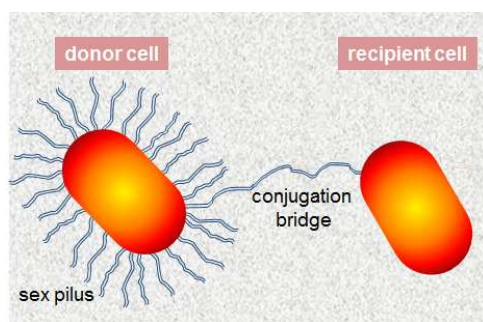


Figure (9.6); bacterial conjugation according to electron microscope.

The ability of bacteria to be the donor is the result of the existence of additional DNA in the cell, which is called F factor or sex factor. The F factor is a circular piece of DNA that can replicate autonomously in the cell. Therefore, it is considered as an independent replicon. The F factor has several genes that are necessary for its replication and for its ability to transfer the DNA into the recipient cell. One of the F factor genes codes for its ability to produce the sex pilus or F pilus on the surface of bacteria. This pilus is very important in the conjugation. The F factor is not the only plasmid that mediates the conjugation, but it's used as an example to clarify this process. The ability of the recipient to act as a recipient for the genetic information is the result of its loss of F factor.

There are three physical stages that the donor cell can exist, which must be studied to understand the different mechanisms of this process;

1. Autonomous fertility (F^+) factor; in this case, fertility factor carries only the genes necessary for its replication and for the transfer of DNA. There are no chromosomal genes bound with fertility factor in the F^+ strains. In the $F^+ \times F^-$ crossing, the F^- converts to F^+ , while F^+ remains F^+ . Thus, the fertility factor is

described as contiguous. Moreover, there is only low level of the transfer between the chromosomal genes.

2. Integrated host factor (Hfr); in this case, the fertility factor integrates into the bacterial chromosome through recombination as shown in the figure (9.10). In the Hfr X F^- crossing, it's rarely for the fertility factor to be converted into Hfr, while Hfr remains Hfr. Moreover, there is a high frequency of the transfer for the donor chromosomal genes.
3. Autonomous fertility factor with chromosomal genes (F'); in this case, the fertility factor is described as autonomous, but it carries some of the chromosomal genes. The F' factors are resulted from the excision of F factor from the Hfr bacteria as shown in the figure (9.7). Sometimes, when the fertility factor is excised from the Hfr chromosome, it's possible for the donor genes that located on the either sides of the fertility factor to be excised with the fertility factor to generate the autonomous fertility factor with the chromosomal genes (F'). In the crossing $F' \times F^-$, the F^- converts to F' , while the F' remains F' . Moreover, there is high frequency of the transfer of the other donor chromosomal genes.

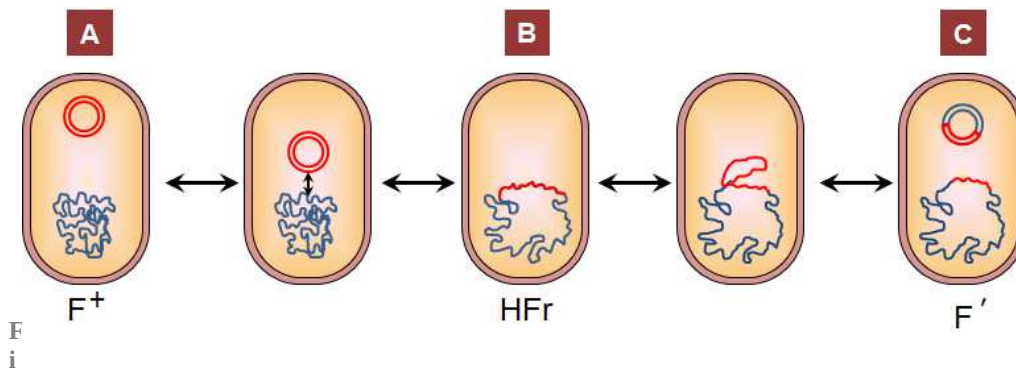


Figure (9.7); the three physiological states of the fertility factor; (A) the autonomous (F^+) state, (B) the chromosomal integrated (Hfr) state, and (C) the autonomous (F') that carries some chromosomal genes.

Mechanisms of Conjugation

As long as the donor cell can exist in three different states, so, there are three different mechanisms that the donor cell determines on the recipient counterpart, and as follows;

1. $F^+ \times F^-$ mechanism (figure 9.8); which includes two steps. The first one is the pilus formation, in which the top of sex pilus is connected with the recipient, and the conjugation bridge is formed between the donor and the recipient cell. It's thought that this bridge is the passage through which the DNA passes from donor to recipient. Thus, DNA is protected from the nucleic acids digestive enzymes that exist in environment. It's possible for conjugation to be hampered, since the mating pair is separated by mechanical forces. Thus, the conjugation continues until the connection between the conjugated cells is broken. It takes ~100 minutes to transfer

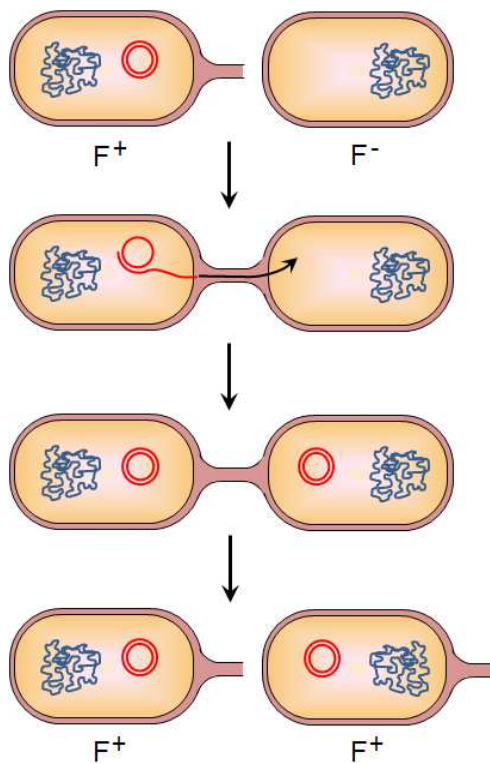
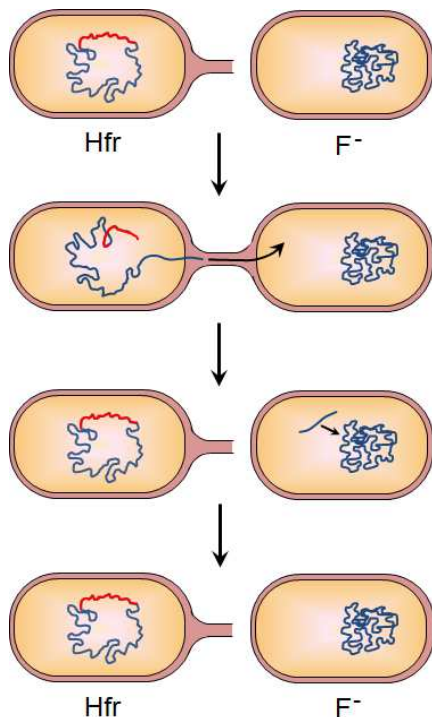


Figure (9.8); the $F^+ \times F^-$ crossing mechanism.

the entire bacterial chromosome, and under standard conditions, contact is often broken before the completion of transfer. Consequently, the mating pair doesn't stay connected for long period of time. The next steps includes the DNA transfer, in which the plasmid DNA is cleaved at a specific site called origin of replication or *oriT*, then, its replicated by rolling circle mode (see chapter three).

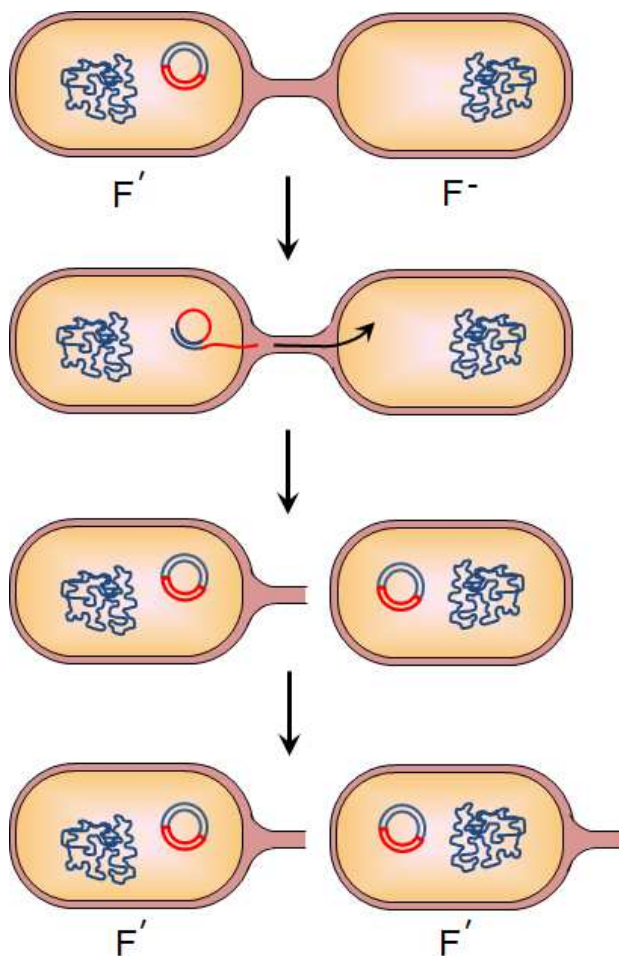
Single copy of DNA is passed through the conjugation bridge and enters into the recipient cell, in which the first strand is replicated to form its other complementary partner. This process explains the features of $F^+ \times F^-$ crossing type. Here, the recipient converts into F^+ , while the donor remains F^+ , and there is a low frequency of transfer of donor chromosomal genes. In the practice, there is a low level of transfer of donor chromosomal genes in such cases.



2. The $Hfr \times F^-$ mechanism (figure 9.9); the *E. coli* strain that has integrated fertility factor to assist high frequencies of recombination are called Hfr strains. This type of crossing involves three steps; 1) pilus formation (see above), the DNA is cleaved at the origin of transfer, and the plasmid is replicated by rolling circle mechanism. But the first DNA to be transferred is the chromosome. Depending on the position and on the direction at which the fertility factor is integrated into the chromosome, different chromosomal genes with different times are transferred.

Figure (9.9); the $Hfr \times F^-$ crossing mechanism.

In general, the transfer system and the distance between genes remain constant. The only case at which the fertility factor is transferred is the transfer of the whole bacterial chromosome. Thus, the recipient cell doesn't uptake F factor in the case of Hfr X F^- crossing. The third step, homologous recombination, in recombination occurs between the transferred DNA and the chromosome that results in the exchange of the genetic material between the donor and recipient. This mechanism demonstrates the features that Hfr X F^- crossing enjoys, in which the recipient cell as it is (F^-), and the donor remains (Hfr) as well. There is high frequency of transfer of the donor chromosomal genes.



3. The $F' \times F^-$ mechanism (figure 9.10); it includes three steps; 1) pilus formation, 2) the transfer of DNA, which resembles to $F^+ \times F^-$ crossing. However, since the F' factor carries some bacterial chromosomes, the later are transferred with it, 3) homologous recombination, which is not necessary despite of its occurring possibility. This mechanism explains the features that $F' \times F^-$ enjoys, where the F^- converted into F' , while F' remains without change. Moreover, there is a high frequency of the transfer of other donor chromosomal genes.

Figure (9.10); the $F' \times F^-$ crossing mechanism.

The Importance of Conjugation

The conjugation method is the essential mechanism to transfer the genes between Gram negative bacteria the transfer can be done between different types of bacteria.

The transfer of several resistances for antibiotics can be done by conjugation, which becomes a crucial problem in the treatment of certain bacterial maladies. Since the recipient cell become donor after the transfer of the plasmid, it's easy to see the reason of the conversion of the sensitive colony into a resistant counterpart after the entry of the antibiotics resistance gene that carried on the plasmid.

Second: Transformation

Transformation is the genetic process that results from the taking up of the naked DNA by the recipient cell. Some bacteria can take up fragments of DNA from the external medium. The source of the DNA can be other cells of the same species or cells of other species. In some cases the DNA has been released from dead cells; in other cases the DNA has been secreted from live bacterial cells. The DNA taken up integrates into the recipient's chromosome. If this DNA is of a different genotype from the recipient, the genotype of the recipient can become permanently changed, a process aptly termed transformation. As it shown in the chapter one, Frederick Griffith was discovered transformation in the bacterium *Streptococcus pneumoniae* in 1928. It's noteworthy that the way by which the DNA is incorporated into the bacterial chromosome in transformation is analogous to the production of recombinant exconjugants in Hfr X F⁻ crosses. Note, however, that in *conjugation* DNA is transferred from one living cell to another through close contact, whereas in *transformation* isolated pieces of external DNA are taken up by a cell through the cell wall and plasma membrane. Figure (9.11) shows one way in which this process can occur.



Frederick Griffith

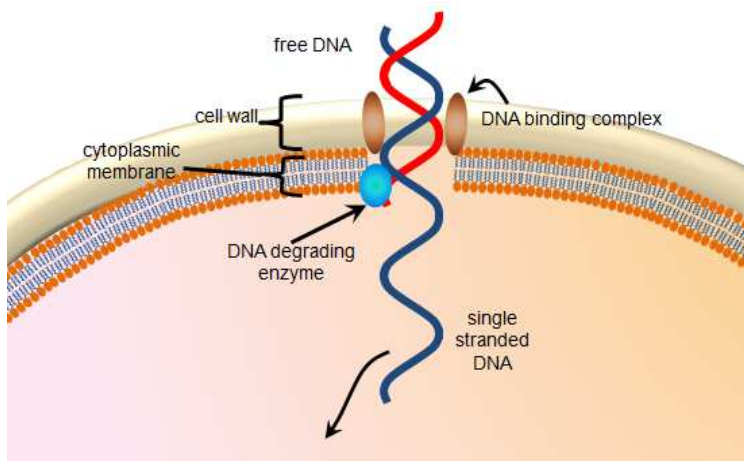


Figure (9.11); the mechanism of transformation in Gram positive bacterium. Bacterium undergoing transformation (a) picks up free DNA released from a dead bacterial cell. As DNA-binding complexes on the bacterial surface take up the DNA (inset), enzymes break down one strand into nucleotides; a derivative of the other strand may integrate into the bacterium's chromosome (b).

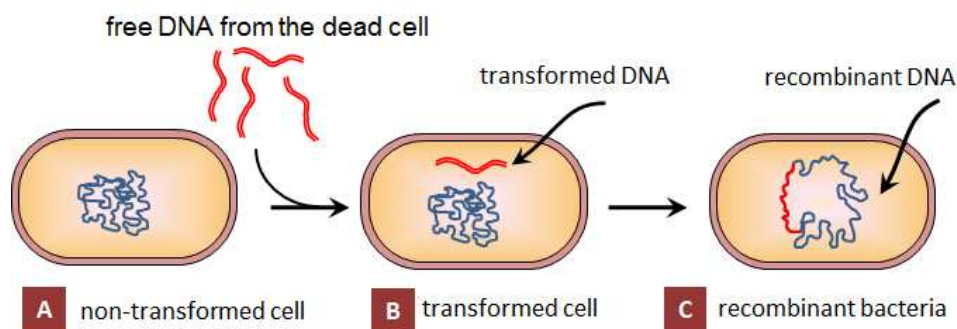
There are two essential factors affecting on transformation; the first one, is the size and state of DNA, the size of the typical DNA double helix to perform transformation is 5×10^5 daltons, and it's necessary for this foreign DNA to contain

certain features to be inserted into the host chromosome (see the appendix of this chapter). The transformation is characterized by its sensitivity for nucleases in the environment. The second factor is represented by the efficiency of the recipient cell or the readiness of the recipient cell for transformation. Some bacteria characterize with their ability to uptake DNA naturally. However, these bacteria take up DNA only in a limited period in their growth cycle. This time represented by their ability to produce protein called competence factor. At this stage, it's possible for these bacteria to be described as "competent". In other bacterial types, it's impossible for the DNA to be taken up naturally. However, in these bacteria, it can induce bacteria in vitro by some chemical materials (such as calcium chloride " CaCl_2 ").

The Mechanism of Transformation

In general, the steps of bacterial transformation can be summarized in two steps; the first one is the uptake of the foreign DNA, and the second step is its integration into the chromosome if the recipient cell (figure 9.12).

1. DNA uptake; the way of DNA uptake differs in Gram positive bacteria compared with its Gram negative counterpart. The DNA in Gram positive bacteria is taken as single stranded molecule, whereas its complementary strand is synthesized in the recipient cell. In contrast, Gram negative bacteria uptake the foreign DNA as a double stranded form.



12); Transformation Under the right conditions, bacteria can take up pieces of naked DNA from the external environment. The fragment of DNA may pass through the outer cell layers without the aid of a protein or virus. Once inside the bacteria, the fragment of DNA must recombine with the chromosome to prevent degradation by exonucleases or restriction enzymes.

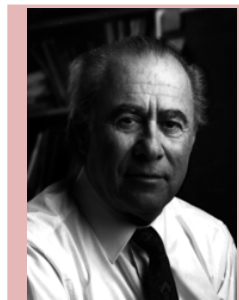
2. Homologous recombination; after the uptake of DNA, reciprocal recombination is taken place between the chromosome and the donor DNA. This recombination requires the presence of homologous regions between the donor DNA and chromosome, and results in a reciprocal of DNA between the donor and recipient as shown in the figure (9.12). then, the DNA can transfer successfully between the very closely related bacterial cells because the existence of homologous regions between the donor and recipient sequences.

The Importance of Transformation

The researchers demonstrated that the transformation has an importance which increased by the increasing of virulence. Transformation is used as an essential tool in many bacterial researches because of the ability to deliberately transform the genotype of the bacterial strains very specifically through transformation. For example, the researchers used transformation widely in the experiments of genetic engineering and gene cloning since their early developmental stages (see chapter eleven).

Third: Transduction

Some phages are able to pick up bacterial genes and carry them from one bacterial cell to another: a process known as transduction. Thus, transduction joins the battery of modes of transfer of genomic material between bacteria - along with Hfr chromosome transfer, F' plasmid transfer, and transformation. In 1951, Joshua Lederberg and Norton Zinder were testing for recombination in the bacterium *Salmonella typhimurium*. They performed the same way followed in recombination, since they used two different auxotrophic strains each of which can't grow when it's plated into the minimal medium. However, after the two strains were mixed, wild-type prototrophs appeared at a frequency of about 1 in 10^5 (thus far, the situation seems similar to that for recombination in *E. coli*). However, in this case, the researchers also recovered recombinants from a U-tube experiment (Davis experiment), in which conjugation was prevented by a filter separating the two arms. By varying the size of the pores in the filter, they found that the agent responsible for gene transfer was the same size as a known phage of *Salmonella*, called phage P22. Furthermore, the filterable agent and P22 were identical in sensitivity to antiserum and immunity to hydrolytic enzymes. Thus, Lederberg and Zinder had discovered a new type of gene transfer, mediated by a virus. They were the first to call this process transduction. As a rarity in the lytic cycle, virus particles sometimes pick up bacterial genes and transfer them when they infect another host. Transduction has subsequently been demonstrated in many bacteria. To understand the processes of transduction we need to distinguish know some of the important terms in bacteriophages genetics. The term bacteriophage is used to describe the viruses that exclusively infect bacteria. The bacteriophages have two pathways after their infection to bacteria; they either immediately lyse and kill the host (virulent phages) or they are integrated into the bacterial chromosome by site-specific recombination, and remain within the host cell for a period without killing it (non-virulent or temperate phages). A phage integrated into the bacterial genome is called a prophage; it's called so because it behaves like a member of the chromosomal family. A bacterium harboring a quiescent phage is called lysogenic, since the phage sometimes and under the suitable circumstances it's released from the bacterial chromosome, and the bacterial cell is converted into the lytic stage. It's possible for the bacterial cell to be induced from the lysogenic to lytic either by UV, X-rays, or by some chemical materials.



Norton Zinder

The Mechanism of Transduction

The ability of the phage to intermediate transduction is related with the cell cycle of the phage itself. This is reflected on the mechanism through which it's possible for this phage to transfer the genetic material.

1. Generalized Transduction; it's the process by which it's possibly for "any gene" to be transferred from the donor to the recipient through some of the bacteriophages. This type of transfer requires a virulent phages lyse the host cell. The mechanism of generalized transduction is shown in the figure (9.13).

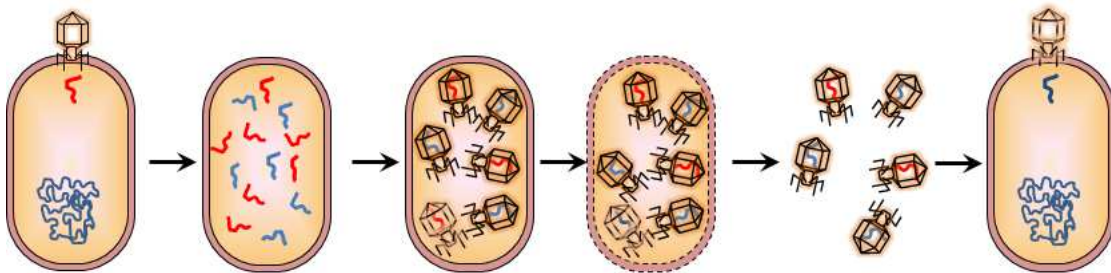


Figure (9.13); the mechanism of generalized transduction for the bacterial genes by phages.

During replication of these phages, DNA is usually exposed to degradation into smaller pieces. Some of these pieces packaged sometimes into the head of phage itself instead of the phage genes itself by a mechanism called "head-full mechanism". Sometimes, one of the host DNA pieces is randomly packaged into the head of the phage. Thus, it's possible for any donor gene to be transferred but it has to constitute enough DNA in order to fit the head of the phage. If the recipient cells are infected by phages that contain the donor DNA, it's possible for the donor DNA to get inside the recipient cells. Inside these cells, it's possible for homologous recombination to occur that replace the donor DNA with the recipient DNA (see figure 9.13). Thus, the phage that carries the bacterial DNA infects another bacterial cell. And it's possible for this bacterial DNA to integrate into the chromosome of the recipient cell by homologous recombination. Many phages depend on this mechanism of transfer but P22 phage is a typical example for this mechanism. It's possible for everyone who pursue the mode of action of the generalized transduction to realize that this process is rarely happened since it requires three conditions for the phage in order to be able to do so; it should be able on destroying the host chromosome, a strict mechanism for packaging of all of its own DNA in the head of phage in order to capacitate the host DNA, and the phage transferred genes should be able to recombine with the chromosome of the recipient cell.

2. Specialized Transduction; it's the process by which it's possible for "only the flanking bacterial genes of the phage insertion" to be transferred into the recipient cells. This type of transfer requires to a non-virulent or temporal phages to integrate into the host chromosome. The genes that are destined to be

transferred depend on the place at which the prophage insert itself on the host chromosome, but the specialized transduction mechanism that illustrated in the figure (9.14) of λ phage is considered as a typical example for this mechanism. This is taken place only when λ phage want to excise itself to get out of the bacterial chromosome once its induced, since an abnormal excision is taken place in the bacterial chromosome. This results in the fact that the resulting phage carries its flanking genes and leaves some of its genes because of its limited capacity for the genetic material. This leads to the production of a defective phage since it leaves some of its genes behind in the host chromosome. But, in the same time, its earned bacterial genes such as *gal* or *bio* genes. These phages called *dgal* phages or *dbio* phages. And it's possible for this abnormal DNA that carries the flanking genes to be packaged into the head of the phage, and to infect another bacterial cell. By this way, it's possible for the donor cell genes to be transferred into the recipient cell. And it's possible for the homologous recombination to occur between the donor and recipient genes.

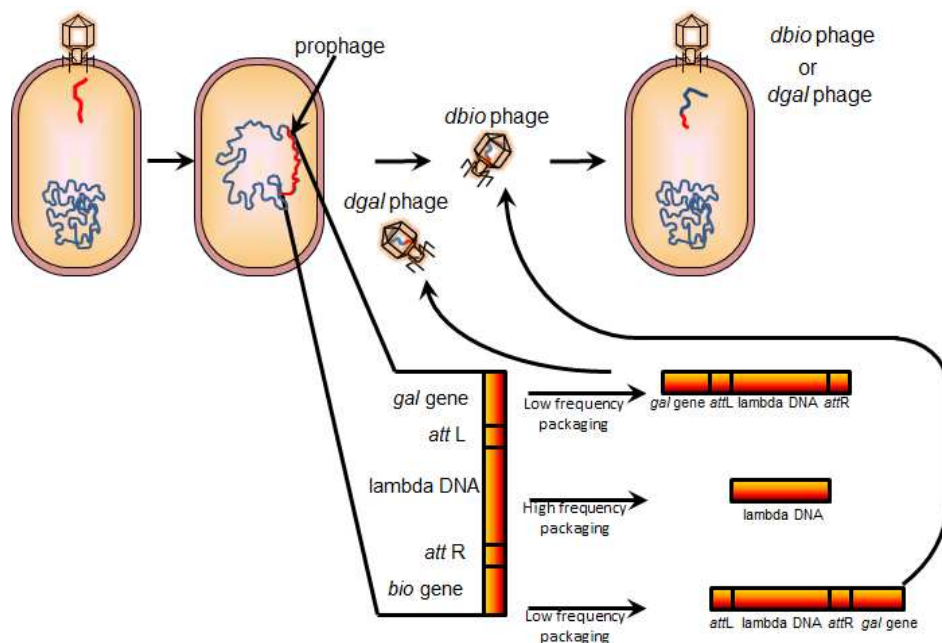


Figure (9.14); the specialized transduction mechanism; When lambda phage enters its lytic cycle and makes phage particles, it usually packages the lambda DNA between the *attL* and *attR* sites. Occasionally, a mistake will occur, and part of the bacterial chromosome DNA will be packaged. Since lambda DNA normally integrates between the *gal* and *bio* genes of the *E. coli* chromosome, the defective lambda particles will most likely contain one or other of these genes.

The Importance of Transduction

The transduction has a significant importance represented by the conversion of the lysogenic phages in the nature, and they considered as a source for the virulent bacterial strains. The phages are used in the experiments of genetic engineering as

vehicles, in which the foreign DNA that is isolated from any organism is integrated (see chapter eleven). The ability of these phages to carry small pieces of DNA has been exploited and used as a submissive tool in identification of the bacterial genes maps and in the bacterial markers.

Chapter Nine Appendix

The Destiny of Foreign DNA in Bacterial System

Irrespective of the mechanism the DNA relies on in its entry into the bacterial cell, it has three different routes; it can be kept as an autonomous DNA molecule, it can be destroyed completely, or only part of it is survive after it's integration or recombination with the host chromosome before the degradation of its other parts (figure 9.15). In order to stay inside the bacterial cell as an autonomously replicating molecule, DNA must have the origin of replication (*oriC*), and must loss the exposed termini, which mean, it must be circular in most bacterial types. The plasmids are a typical example of such cases since they have *oriC* in addition to their circular form. Hence, the recombination is not necessary for the plasmid survival. Therefore, this phenomenon is exploited in genetic engineering experiments, where the gene of interest is inserted in the plasmid molecule (see chapter eleven).

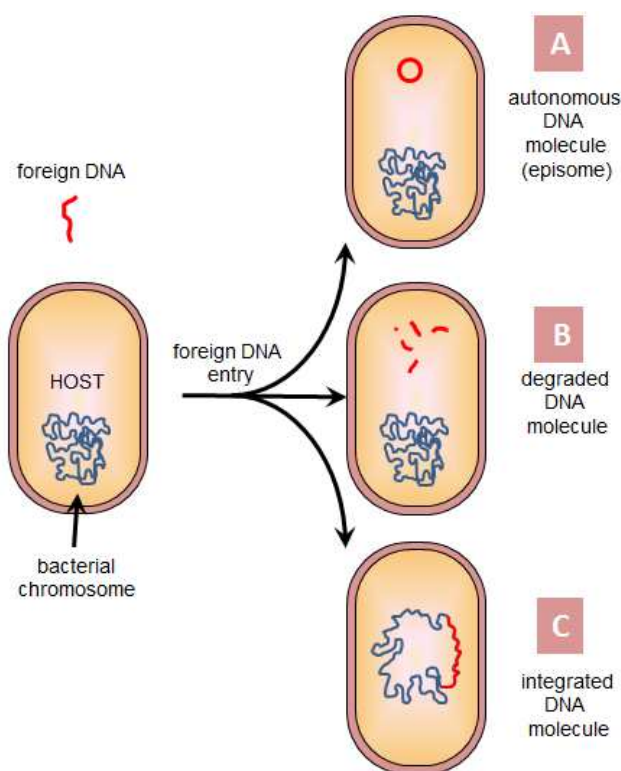


Figure (9.15); the destiny that the foreign DNA faces after their entry to the bacterial cells. In case (A), the foreign DNA behaves autonomously since it doesn't possess any free termini. In case (B), the foreign DNA is degraded by digestive enzymes. Or in case (C) the foreign DNA is integrated into the host chromosome by homologous recombination.

If the molecule doesn't have these features, such as it has linear form, it's usually degraded by exonucleases that attack the exposed termini. Thus, it must survive to be inserted into the chromosome of the recipient cell be recombination (see chapter nine). In order to perform recombination, the foreign DNA must contain regions of homologous sequences with the chromosome of the recipient bacterium to recombine with it.

Consequently, the genes that are originated from the foreign DNA are integrated and replaced the original genes of the recipient cell. But, the process of homologous recombination doesn't occur except with the closely related DNA molecules, such as the DNA of two strains that belong to only one bacterial species. And it's possible for the distant related DNA to integrate also by recombination if it's flanked by

related sequences in the recipient cell. If the foreign DNA contains transposable elements, these elements can leave it and transpose into the chromosome of the new host.

In addition to the aggression of exonucleases that the DNA molecules exposed to, it also exposed to the degradation by another class of digestive enzymes such as endonucleases that also called restriction enzymes (see chapter eleven). Allah Almighty designed this protective mechanism for the incoming foreign DNA degradation. The bacteria perform these severe since they previously postulate that each foreign DNA is a DNA belongs to certain enemy such as viruses. Thus, the DNA must be destroyed into small pieces by these restriction enzymes. But, it's possible for some phages to survive from this rigorous action by the methylation of their recipient sequences. Therefore, the bacteria accept these phages and consider them as a part of its genetic material. This phenomenon is also exploited in genetic engineering in the protection of the foreign DNA after its entry into the host cell.

It's so obvious to us that the studying of microbiology, particularly the tracking of the destiny of the entered foreign DNA by conjugation, transformation, or transduction is provided a big service especially in establishing the essential fundamentals of the gene cloning experiments as it's shown in the next chapter.

Chapter Nine Revision Questions

Q1- Justify the following:

1. The transfer of the genetic material in bacteria is unidirectional?
2. Instead of calling the conjugated bacterial cells as male and female, they are called donor and recipient?
3. It's impossible (practically) for the whole chromosome to be transferred completely during the bacteria conjugation?
4. Sometimes, when the patient administers the antibiotics, the later doesn't perform its scheduled task in the killing of its sensitive bacteria?
5. The bacteria that have latent phage are called lysogenic?
6. It's impossible for any gene to be transferred through the generalized transduction, instead, particular genes that can be transferred through the specialized transduction?
7. Some phages successfully infect some bacterial species despite of the severe treatment of the host against any foreign aggression?
8. The foreign DNA may succeed after its entry into the chromosome despite of its being doesn't possess any homologous sequences with that chromosome?

Q2- Choose the correct answer:

1. When there is a direct contact, the transferred DNA is protected from the donor into the recipient cell through.....
a. DNA binding proteins b. conjugation bridge c. cell membrane of competent cells d. cell wall of competent cells
2. the method is one of the means by which the bacterial cells diverse their genetic content.
a. horizontal gene transfer b. vertical gene transfer c. only "a" and "b" d. cell division
3. when the conjugation occurs, the fertility plasmid transferred by a replicative mechanism from the donor to the recipient cell.
a. unidirectional theta b. bidirectional theta c. rolling circle d. D loop
4. it's possible for the bacterial cell to receive the DNA that released from the dead cell in a process called
a. transformation b. conjugation c. transduction d. non
5. the integrated phage into the bacterial chromosome is called since its behave like a member of the chromosomal family.

- a. virulent phage b. non-virulent phage c. temperate phage d. prophage
6. when a circular DNA molecules having *oriC* enters, it usually by the bacterial host.
a. treated like episome b. degraded by exonuclease c. degraded by restriction enzyme d. only "b" and "c"
7. the only way by which the DNA integrates into the bacterial chromosome in the transformation is homologous to the crossing of type.
a. $F^+ \times F^-$ b. $Hfr \times F^-$ c. $F' \times F^-$ d. non
8. the bacterial cells that have the ability to receive the foreign DNA are called
a. natural cells b. competent cells c. transformed cells d. transfected cells
9. when the bacterium exposed to the generalized transduction by certain phage, it must follows the pathway.
a. only lytic b. only lysogenic c. both lytic and lysogenic d. non

Q3- Define the following:

prototrophs, episome, *tra* region, pili, Hfr, F' , *oriT*, full head mechanism, *dbio* phage

Q4- Connect terms exist in right with its suitable counterparts in left

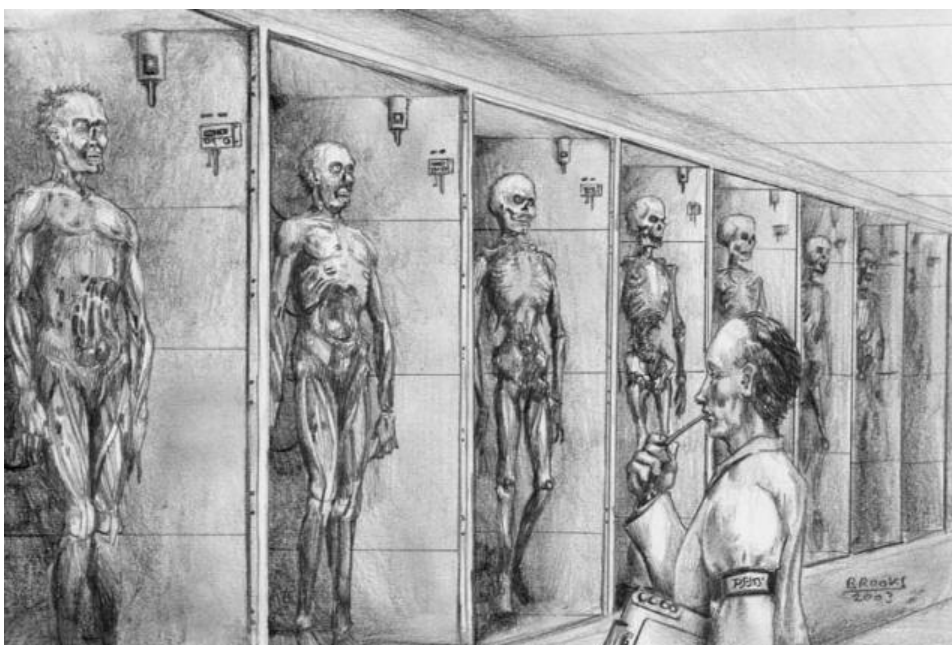
Bacteriophage	$F^+ \times F^-$
Direct cell contact	$Hfr \times F^-$
F^- converted to F'	$F' \times F^-$
High frequency transfer of chromosomes	Generalized transduction
Error during excision	Specialized transduction
Error during packaging	Transformation
F^- converted to F^+	Conjugation
Naked DNA transfer	Transduction

Further Readings

- Alberts** B., Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Clark** D. Molecular biology: understanding the genetics revolution. Elsevier, 2010.
- Griffiths** A., Wessler S., Lewontin R., Gelbart W., Suzuki D., Miller J. An Introduction to Genetic Analysis. Eighth edition, 2004.
- Guttman** B., Griffiths A., Suzuki D., Cullis T. Genetics: a beginner guide. Oneworld Oxford, 2004.
- Hanahan**, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557- 580.
- Kayser** F., Bienz K., Eckert J., Zinkernagel R. Medical microbiology. 2005, Thieme.
- Kresina** T., F. (2001). An introduction to molecular medicine and gene therapy. Wiley-Liss, Inc.
- Miller**, R. V. 1998. Bacterial gene swapping in nature. *Scientific American* 278(1):66–71.
- Murray** et al. medical Microbiology. Third edition, 2005.
- Novick**, R. P. 1980. Plasmids. *Scientific American* 243(6):103–124.
- Pace**, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–740.
- Passarge** E. Color atlas of Genetics .Second edition, Thieme, 2007.
- Pierce** B. Genetics: A conceptual approach. 2002
- Syvanen** M., Kado C. Horizontal gene transfer. Second edition, AP academic press, 2002.
- Summers** D. The biology of Plasmid. Blackwell science, 1996.
- Talaro**, K. and Talaro, A. Foundation in microbiology. Fourth edition. The McGraw–Hill, 2002.
- Verma** P., and Agarwal V. molecular biology. First Edition Schand group, India (2004).
- Wollman**, E. L., F. Jacob, and W. Hayes. 1962. Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harbor Symposium on Quantitative Biology* 21:141–162.

Chapter Ten

Principles of Gene Cloning



is scene will be right at certain day to be routine? Is that possible to reach to "offspring selection" where we choose what we want of genes and eliminate what we don't want?

Introduction

Genetic engineering or the genetic manipulation is defined as the *in vitro* synthesis for new forms or new arrangements of DNA in such away as to allow the incorporation or propagation of altered genetic condition to continue in nature. The *in vitro* manipulation of DNA is called recombinant DNA technology. Broadly, genetic engineering means the manipulation of genes under controlled laboratory conditions. The importance of genetic engineering as a modern technique stems from many reasons, such as it allows the isolation and studying of individual genes in large quantities and in complete purity. These genes are possible to be expressed on and it's possible to reused in the cells whether they are of one type or different types.

The experiments of genetic engineering achieved many goals that a lot of fields of biology were unable to achieve, such as the induction of bacteria to synthesize hormones and the cost effective and rare biological products. The best know example for such precious compounds is insulin, human growth hormone, and rennin. All of these three compounds were extracted before the discovery of genetic engineering from animals, such as cows, pigs and calves, or from human bodies. The hope of the workers in genetic engineering is to use these experiments in the future to correct some disorders that originated from mutations such as sickle cell and thalasemia (gene therapy). These manipulations also aim to provide new features such as in herbicides,

the frost resistance for plants, or to produce bigger and best farm animals. Most of these huge achievements for this science are originated from the enormous development in gene cloning successfully in the recipient cells. Therefore, in this chapter, we will review the fundamental tools to show us how to clone certain gene step by step until we reach to the successful expression of this gene into the host cells.

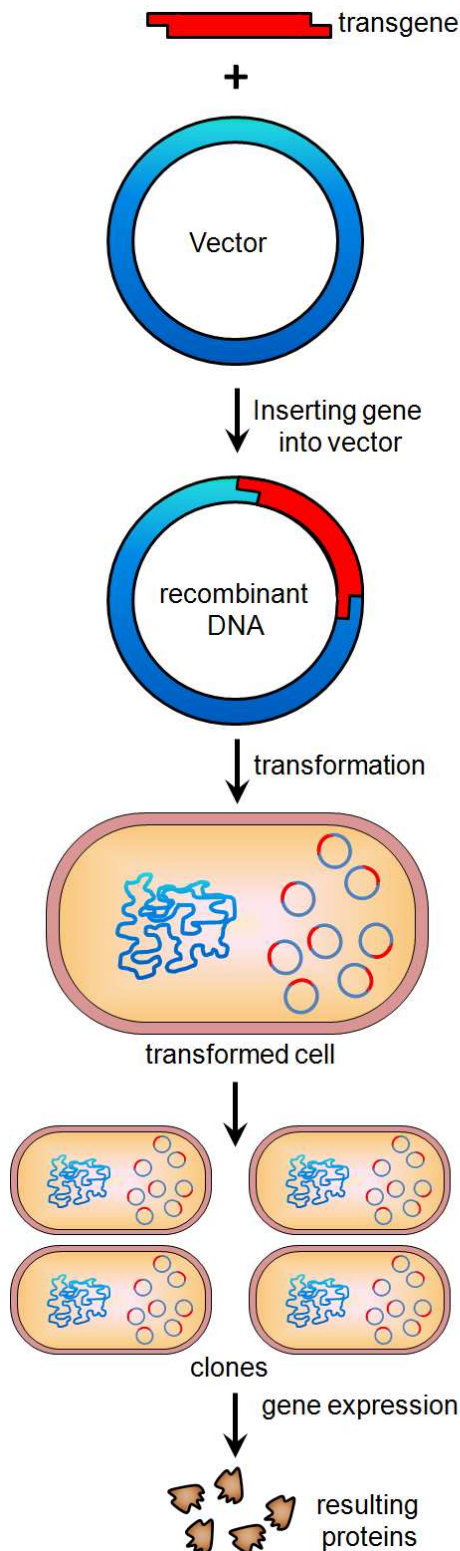
The Successive Tools for Gene Cloning

Many biological steps are used to achieve the manipulation of the genetic material to perform to clone certain gene. But before getting to them we have to know that many DNA manipulation experiments require enormous quantities (copies) of a particular DNA piece. These steps are called "gene cloning", because of the production of homologous copies of the original DNA fragment. To perform this practically, it's usually to insert the gene into a certain vector to form the recombinant DNA molecule. The vector acts as a vehicle to present the gene into the host cell, to direct its replication in its required form, and to enable it to express about itself in some cases to produce the hybrid protein (figure 10.1). The host cell carrying the gene-containing vector produces progeny all of which contain the inserted gene. These identical cells are called "clones". In the transformed host cell and its clones, the inserted gene is expressed about itself in some cases in such away it's transcribed and translated into the recombinant proteins.

Classically, to clone certain gene, certain tools should be provided through which the genetic material that have this gene is isolated. Now, several tools are provided by which the researcher can easily isolate the genetic material. Then, this gene must be separated from the rest of the genes in the isolated genetic material. To do so, many restriction endonucleases are available which cut the gene of interest of its either sides. Then, the necessary tools that connect this gene with the suitable gene vector must be provided (such as ligases – see below). This is done after cutting the vector with the same restriction enzyme that has been usually used in the separation of the gene of interest. After formation of the recombinant molecule, other tools must be provided that introduce this recombinant molecule into the host cells. Finally, the tools that enable the researcher to know or select the clones that have these recombinant molecules must be provided. These tools constitute the essential basics in gene cloning that are not difficult to be understood by the inexperienced reader. Therefore, we will review these tools sequentially to complete the picture to us about the cloning of the gene.

First; Tools of Separation of DNA containing the Desired Gene

Before we clone any gene in the experiments of genetic engineering we have to obtain it in satisfied quantities. This is done by three methods differ with the difference of characteristics of the desired piece of DNA and the purpose of its cloning. In the first method, it's possible to isolate the organism's entire DNA, then, its fragments are digested into several pieces, the desired fragment of interest isolated, and eventually cloned. Or, it's possible to clone all the fragments by a suitable vector, than each clone



is tested for the presence of the desired gene. Or the desired DNA fragment can be directly synthesized in order to be cloned afterward (see chapter eleven).

Figure (10.1); general scheme for gene cloning. After isolating the gene of the DNA fragment to be cloned, this fragment is ligated with the suitable gene vehicle to form the recombinant DNA molecule. This molecule is introduced into the recipient cells by transformation. The recipient cells express about the products of the cloned gene.

Second; Tools of cutting and Sealing DNA Fragments

After obtaining the genetic material that has the desired gene with high purity, we need now to isolate this gene from the rest of the genetic material of the donor cell. To do so, we have to provide a tool that "scissor" this gene from its edges. Perhaps just to think about the presence of certain enzymes that can act as "cellular scissors" for the desired gene fragment is a fictitious thing.

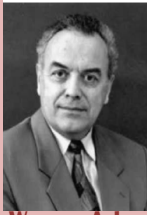
This fiction was begun to convert into reality at 1950s with the observation that *E. coli* can resist infection somehow by "restricting" (i.e., inhibiting) replication of the invading phage. By 1962, Werner Arber and his research group had found an enzyme system in bacteria that could inhibit phage replication by cleaving the bacteriophage DNA once it entered the cell. The team isolated the DNA-cleaving enzyme from *E. coli* cells and called it an endonuclease, because this endonuclease enzymes that occur naturally and function in the restriction and modification system in bacteria and quickly became known as restriction enzymes.

It must be known that there are several types of nucleases; RNases attack

RNA, while DNases attack DNA. There is a type of specificity usually in the action of these digestive enzymes. Since some of these enzymes attack only single stranded nucleic acids and other enzymes attack double stranded nucleic acids, while only a small ratio attack the either types of nucleic acids. Exonucleases attack the nucleic acid molecule from its termini. It deserves to be noticed that any of exonuclease types attack one of the two termini of the nucleic acid, either 3' or 5' end, but not both. Despite the enzyme that discovered by Arber and his colleagues is also a digestive enzyme, it doesn't digest nucleic acid from its termini, instead, it "cleaves" the nucleic acid internally. Consequently, this enzyme is considered as endonuclease, since this enzyme cleaves the nucleic acids, and of course this enzyme cleaves the nucleic acids, and the "cleavage" occurs within the DNA molecule.

In case of bacteriophage infection, these *E. coli* endonucleases cleave the invading DNA but never cleave the bacterial host. This "restriction" is part of what so called now the bacterial restriction – modification system, that protect the bacteria from infection by bacteriophages through the recognition between the host DNA and the foreign DNA. This means, that all the bacterial DNA, including plasmid DNA, is DNA usually protected by methylation. The reason for calling these enzymes restriction enzymes is the presence of these enzymes in bacteria restrict the growth of bacteriophages. Thus, their function that because of which they are present in nature is represented by the protection of the bacterial host from the DNA of the foreign organisms (such as infective phages), therefore they considered as defensive enzymes. Hence, we have to ask the question; if the restriction – modification system, that contains methylases coupled with restriction endonucleases, faces two un-methylated DNA strands, one of them is cellular and the other one is foreigner, how it can differentiate between them in such away it modifies the first and cleaves the second?

To answer this question, we have to know that this system, restriction – modification system, produces additional complications on DNA replication that for simplicity were never mentioned in chapter three. The cellular double stranded DNA contains methyl groups on either strands of DNA in the recognition sites of restriction endonuclease. When replication synthesizes two new strands, hence, one of the strands in each daughter double helix losses methyl groups at the beginning. This hemi-methylated DNA must not be recognized from the cell as "foreign DNA" and cleaved. Instead, it's recognized as "self DNA" and modified by the restriction – modification system by addition of methyl (figure 10.2).



Werner Arber



Hamilton Smith



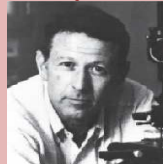
Daniel Nathans



Herbert Boyer



Stanley Cohen



Paul Berg

Therefore, the restriction – modification system acts in a certain way to distinguish three different cases of the recognition sequences methylation and takes one of three different actions. If the sequence is not methylated, the restriction endonuclease in the restriction system cuts it. If the DNA is only methylated in one of its strands (hemi-methylated), the modifying (methylase) enzyme in the modification system adds methyl groups into the other unmethylated strand (figure 10.2). If the DNA is methylated on either strands, the restriction – modification system don't act anything. By this way, the methyl tag is regarded as a distinguishing point between host DNA and invading DNA. Since the DNA of *E. coli* is methylated DNA, restriction enzymes distinguish it as an original chromosome, and protect it from the cleavage by the restriction enzymes that are available in the cell. The phage DNA is unmethylated, therefore, it's regarded as a target for restriction enzymes to restrict (cleave) it to prevent by this way the replication of the invading phage inside the bacterial host cell.

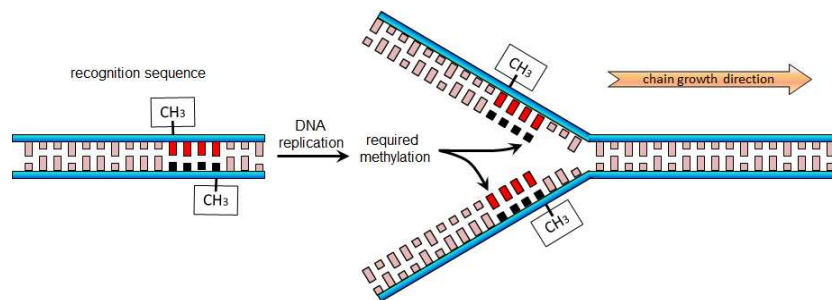


Figure (10.2); Methylation of an asymmetrical sequence necessitates recognition and methylation of two different sequences on the daughter strands of DNA replication.

At the ends of 1970s, the golden era of molecular biology development, In 1970, Hamilton Smith and his colleagues isolated a new restriction enzyme called *Hind* III from the bacterium *Haemophilus influenzae*. Daniel Nathans used the *Hind* III restriction enzyme to cut viral DNA isolated from simian virus capsids (SV40). Smith discovered that the *Hind* III restriction enzyme cut each SV40 DNA genome at only one site, even when an excess of enzyme was present. The scientists were amazed to learn that the *Hind* III enzyme cut the SV40 DNA in exactly the same location along the DNA in every SV40 DNA molecule that was exposed to the enzyme. For their work, Arber, Smith, and Nathans were awarded the Nobel Prize in physiology or medicine in 1978. Moreover, there are two great achievements in genetic engineering; the first one was done for the

first time by Herbert Boyer research group that isolated the *Eco* RI restriction enzyme from *E. coli* bacteria and discovered that when *Eco* RI cuts DNA it produces 5' sticky DNA ends. This discovery was widely accepted in the scientific media since the sticky ends that resulted from the cutting of this enzyme are recognized by their ease of connecting with each other since they are complementary to each others in the sequence. And this facilitates the connecting with each others. Add to that, Stanley Cohen provided a lot of information about plasmids that enabled the researchers to use them as cloning vehicles in the future. At 1980, Paul Berg got Nobel Prize because of his synthesis of the first recombinant molecule, which is a bacteriophage containing a piece of DNA derived from SV40.

After this era, scientists gained more hands-on experience with restriction enzymes in the lab, they learned that restriction enzymes could cut any DNA molecule on demand, regardless of the source of the DNA (except when the DNA was protected by chemical modification). For example, the *Hind*III restriction enzyme will readily cleave any DNA molecule that contains its specific *Hind*III recognition DNA sequence, regardless of the source of the DNA molecule: virus, plant, animal, or human. It is easy to appreciate why restriction enzymes quickly became essential tools for any type of gene research.

With accumulating years, the utilization of these enzymes in the experiments of genetic engineering becomes popular, in such away the researcher that is working in this field in the late 1990s could choose from thousands of commercially available highly purified restriction enzymes for cut it molecule at any desired site. But, before the number get so high, researcher found a successful way to avoid confusion about restriction enzymes nomination. These nominations were derived from the name of the bacterium from which they were isolated. For example, the *Eco* RI restriction enzyme was named from *Escherichia coli*, *Bam*HI was named from *Bacillus amyloliquefaciens* (table 10.1). the three letters of the name of the restriction enzyme consists of; The first letter of the enzyme name is the first letter of the bacterium's genus name (such as *E*), followed by the first two letters of its species (such as *co*), then a letter representing the bacterial strain, and finally a Roman numeral that signifies the order in which this enzyme was discovered. This is applied in organisms that have more than one restriction enzyme (such as *Eco*RI or *Eco*RII). Not that the first three letter of the name of enzyme are written in italic style.

Each restriction enzyme cuts (or nicks) DNA at a specific base sequence, called the recognition sequence (table 10.1), which is can be varied from 4 into 16 or 20 base pairs. As the number of base pairs of the recognition sequence increased, as the number of cuts made in DNA usually decreased. If we postulate that the DNA is distributed randomly in the DNA molecule, hence, we can calculate how much the enzyme cut the DNA molecule.

A significant notion must be known, which is that for each position of DNA molecule, there are four possibilities, (A, C, G, or T). Therefore, the restriction enzymes that recognize four nucleotides sequence, they will cut once per each 256 bps (or 4^4). Whereas, other enzymes recognize six nucleotide sequences, so, they cut once per each 4096 bps (or 4^6) (figure 10.3).

Table (10.1); the restriction enzyme nomination system as it shown in many examples (the symbol / refers to the position at which the restriction enzyme cut the molecule, the letter N refers to any base, and the letter Y refers to any pyrimidine)

Enzyme	Source organism	Recognition sequence
<i>HpaII</i>	<i>Haemophilus parainfluenzae</i>	C/CGG GGC/C
<i>MboI</i>	<i>Moraxella bovis</i>	/GATC CTAG/
<i>NdeII</i>	<i>Neisseria denitrificans</i>	/GATC CTAG/
<i>EcoRI</i>	<i>Escherichia coli</i> RY13	G/AATTC CTTAA/G
<i>EcoRII</i>	<i>Escherichia coli</i> RY13	/CC(A or T)GG GG(T or A)CC/
<i>EcoRV</i>	<i>Escherichia coli</i> J62/Pgl74	GAT/ATC CTA/TAG
<i>BamHI</i>	<i>Bacillus amyloliquefaciens</i>	GAT/ATC CCTAG/G
<i>SauI</i>	<i>Staphylococcus aureus</i>	CC/TNAGG GGANT/CC
<i>BglI</i>	<i>Bacillus globigii</i>	GCCNNNN/NGGC CGGN/NNNNCCG
<i>NotI</i>	<i>Nocardia atitidis-caviarum</i>	GC/GGCCGC CGCCGG/CG

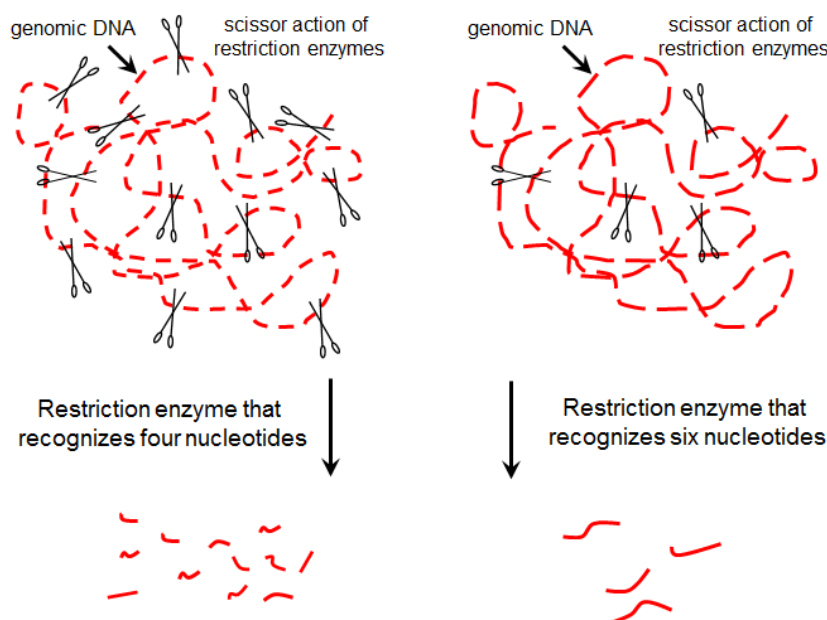


Figure (10.3). The reverse relationship between the frequency of cleavage of DNA strands by restriction enzymes with the number of recognized nucleotides at the site of cleavage. The restriction enzyme that recognize four nucleotides sequences cuts the DNA with more frequency and produces smaller fragments compared with the restriction enzyme that recognizes six nucleotides sequences.

After the identity of restriction enzymes has been known, it must be know how restriction enzymes identify the recognition sequences that available at DNA in order to cut

them. So how do enzymes find their recognition sequences? Scientists have found that restriction enzymes adopt specific three dimensional shapes that progress along nonspecific sequences in a DNA helix, “ scanning ” until reaching their specific recognition sequence. Most restriction enzymes recognize and cut at a palindromic DNA sequences. These sequences represent the substrates exposed to attack by restriction endonucleases. In the English language, a palindrome is when the letters of a sentence spell the same words whether they are read forward or backward; for example, “Madam, I’m Adam” or “Ma is a nun as I am”. As for the DNA, the palindromic site is a sequence of bases in the DNA duplex which has the same reading either from forward or from reverse direction (from 5’ to 3’ or the opposite). For example, the sequence GAATTC is “palindromic” since both sequences of the either strands have the same reading whether from G end or from C end (figure 10.4).

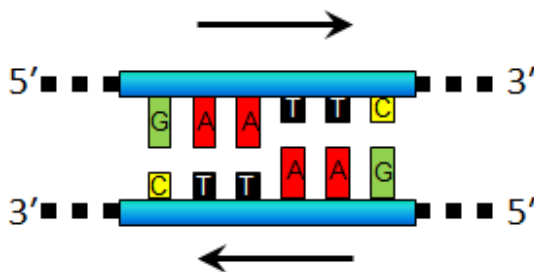


Figure (10.4); an example about the palindromic sequence (the sequence that is cut by *EcoRI*).

The most known example for restriction enzymes is *EcoRI*, which is synthesized by the bacterium *E. coli* from the strain RY13. This enzyme attacks the nucleotide sequence GAATTCCTTAAG. It can be noticed that this palindrome is symmetrical about its center. The restriction enzyme *EcoRI* makes single stranded cuts called “nicks”, between A and G of both strands and this opens the circular DNA molecule.

There are different known types of restriction endonucleases, which nicks different regions of the DNA molecule. And on the basis of mode of nicking of the specialized DNA sequence, there are two types of restriction enzymes; the first group nicks straightly across the DNA duplex, to produce blunt end DNA. The other type of restriction enzymes nicks the DNA from the center of the palindromic sequence but between the same nitrogen bases on the opposite strands. This leaves one or more overhang base from each strand. These ends called sticky ends (figure 10.5). They are called in this name because their formation for hydrogen bonds with their complementary bases that are nicked by the same enzyme. The sticky ends are especially important in the construction of the recombinant DNA molecules.

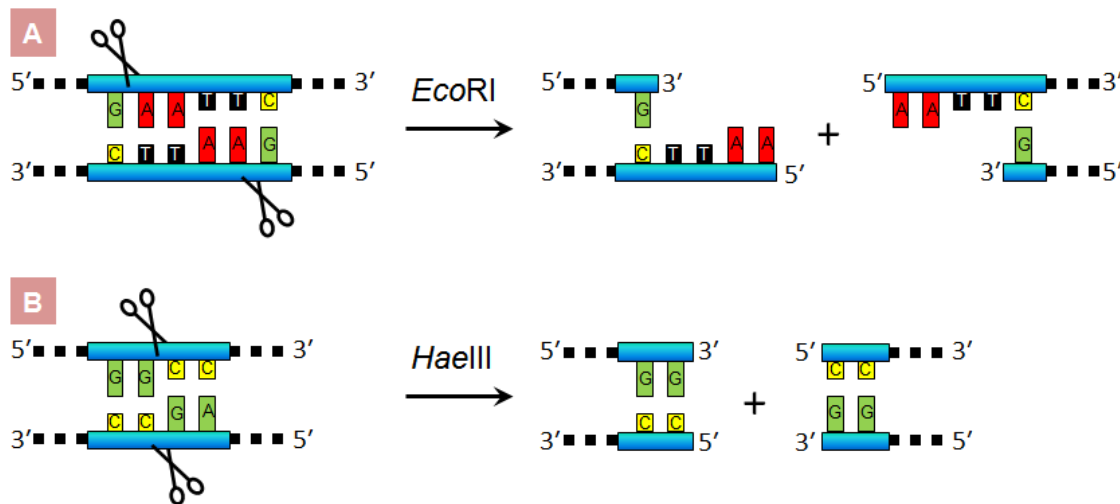


Figure (10.5); types of restriction enzymes according to the mode of nicking. (A) An example about sticky ends restriction enzymes. (B) An example about blunt ends restriction enzymes.

Since the single stranded ends that generated in the nicking site are complementary, they can be corrected with each other. Thus, the generated DNA fragments by the same endonuclease can be joined back easily together by forming base pairs (figure 10.6).

After nicking the foreign DNA molecules by suitable restriction enzymes, they usually joined with the suitable gene vehicles. This is done by ligation enzymes or DNA ligase that is discovered at 1967 by Martin F. Gellert. The joining of the two strands is done a sealing enzyme that is also called "gluing agent" to join the separated nucleic acids together. This ligase joins the DNA fragments covalently. In intracellular level, the DNA ligase acts during DNA replication, which joins Okazaki fragments together to construct the lagging strands (see chapter three). Two requirements must be met for DNA ligase to do it task. First, the molecules must be the correct substrates, that is, they must possess 3'-hydroxyl and 5'-phosphate groups. Second, the groups on the molecules to be joined must be properly positioned with respect to one another. Simply, if the DNA ligase found two DNA fragments annealed together from their termini, it seals them together. In the experiments of genetic engineering, the sticky ended DNA fragments tend to be aligned together for long period of time. Consequently, DNA ligase joins them efficiently.

Since the blunt ended DNA fragments don't have any way to be joined together, they are far away from each other for most of time. Consequently, the sealing of blunt ends is very slow and requires a high concentration of DNA ligase, in addition to the high concentration of DNA fragments to be sealed. It's noteworthy that the bacterial DNA ligases are unable to join the blunt ends. Therefore, the T₄ DNA ligase that is originally derived from the bacteriophage T₄ is used in the ligation of the blunt ended DNA fragments (figure 10.6).



Martin F. Gellert

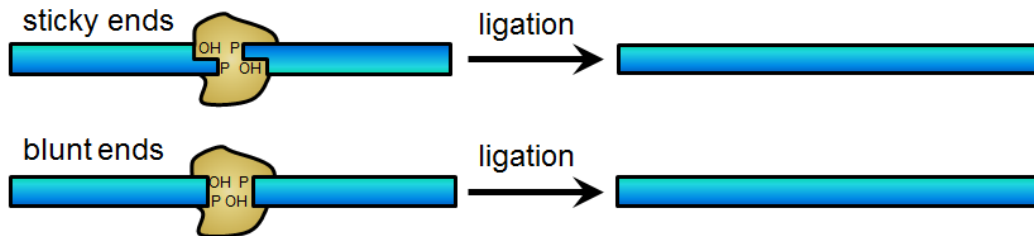
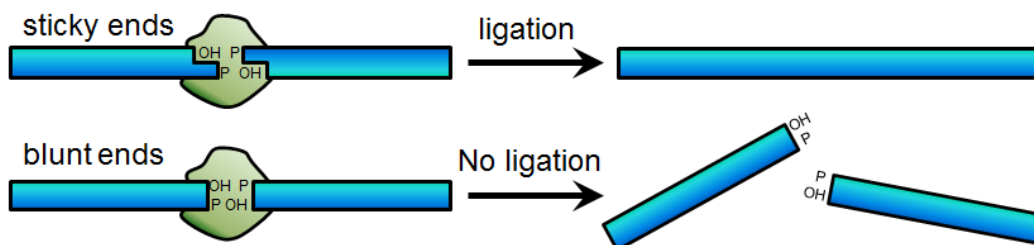
A T4 DNA ligase**B** *E. coli* DNA ligase

Figure (10.6); ligation of DNA fragments by DNA ligase in an energy required reaction. (A) the success of the sealing of blunt or sticky ended DNA termini by T4 DNA ligase, (B) the success of the sealing of DNA termini by bacterial DNA ligase and the failure of sealing of blunt ended DNA termini by the same enzyme.

A second method can be used to create recombinant DNA with the DNA fragments and the vectors that lost the sticky ends because of their cutting by a blunt ended restriction enzyme. After cutting the vector (such as the plasmid) and the donor DNA, the polyA can be added, for example, to the 3' end for the plasmid DNA using an enzyme called terminal transferase (figure 10.7). Similarly, polyT is added to 3' end for the donor DNA fragments. Hence, the two fragments can then be hybridized together by virtue of their self-complementary ends and ligated together. If the tails are long enough, the complex can be directly introduced into cells, where the gaps and nicks will be filled and sealed by the cellular enzymes. This procedure is called homopolymer tailing. Although this technique is more difficult than the ligation of the sticky ends, it has an advantage represented by joining of any pair of termini. But there are disadvantages of this that represented by the loss of control on the direction of insertion of joined molecules, add to that, there is no easy way for the recovery of inserted DNA after its insertion into the vector.

The so called linkers can also be used to generate sticky ended DNA molecules that are complementary to each other. Linkers are blunt ended short DNA molecules that have blunt ends containing recognition sequences for a restriction enzyme that generates sticky ends. The ligation of the linkers with the DNA fragments is very efficient, since the feasibility to obtain high concentration of linkers. After joining of linkers with DNA fragment, the mixture is digested with a restriction enzyme that cuts the linkers and

generates the sticky ends. By this way, the blunt ended molecule is converted into sticky ended molecule that can be joined with other DNA molecules.

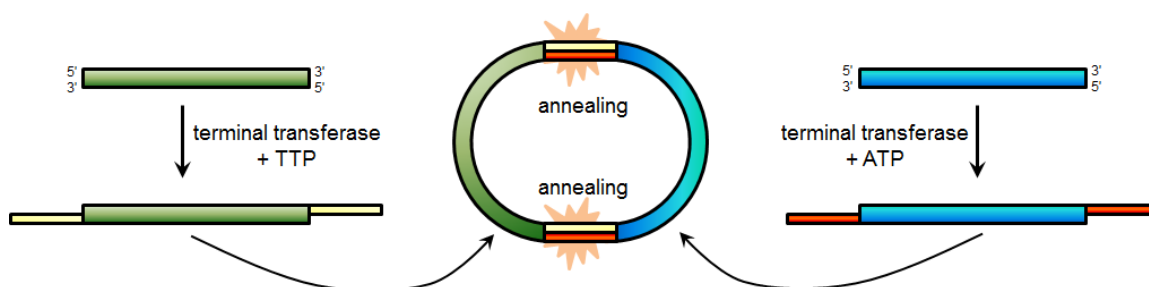
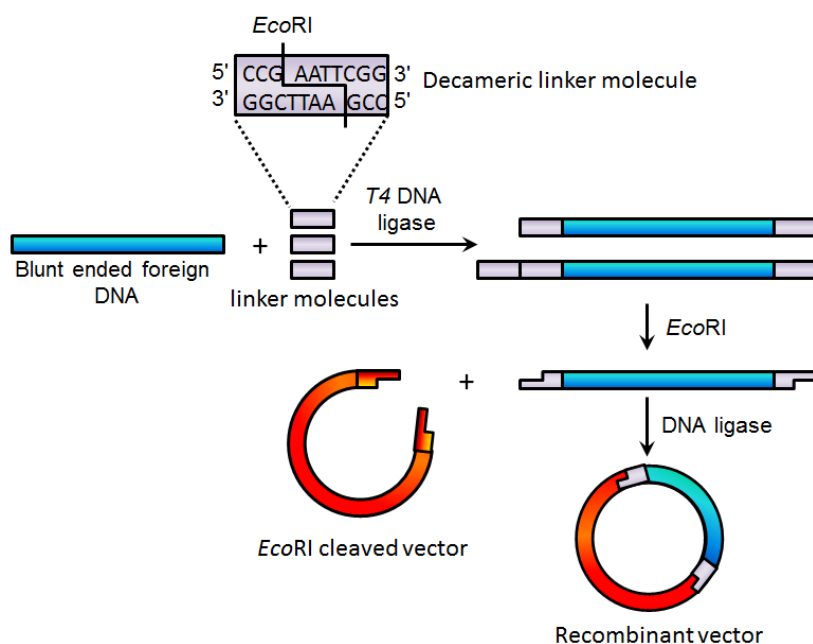


Figure (10.7); homopolymer tailing. The technique of homotailing by poly-dA and poly-dT tails can be used to construct sticky ends of the desired DNA molecule and to generate recombinant DNA molecule.



The above account should not be interpreted to mean that gene cloning is an easy technique. In fact, the production and identification of a bacterial cell which has incorporated a functional copy of the desired gene is a formidable task.

Figure (10.8); ligation of linkers that contain EcoRI site by T_4 DNA ligase with the blunt ended foreign DNA molecules. Consequently, the resulting DNA molecules have cohesive ends by the restriction enzyme EcoRI. This DNA is joined with the vector that is treated with the same restriction enzyme.

Third; Tools of DNA Transfer (Cloning Vehicles)

Gene cloning is an indispensable part of any genetic manipulation and involves the joining of a desired DNA fragment to a vector molecule which serves to propagate that DNA segment in bacteria.

Each individual vector characterizes with features which might not be exist in the other vectors. It deserve to be noted that each type of cloning vectors has certain advantages distinguish it from other vectors. However, plasmids are the simplest

molecules to deal with. Bacteriophages and other viruses can be stored properly for long period of times. Cosmids and artificial chromosomes can carry larger DNA fragments for cloning purposes. It's noteworthy that all vectors are shared with multiple general features; they are typically small, the gene transfer from the donor cell into the recipient counterpart is relatively easy, they have known DNA sequences, have at least one origin of replication and they can replicate in the suitable host even when they carry the foreign DNA. Finally, these vectors codes for a phenotype that can be used to identify their presence and the parental vectors (the vectors that don't have foreign DNA) can be identified from the recombinant vectors (the vectors that have the foreign DNA). The most important types of these cloning vectors are:

- a. **Plasmid vectors;** plasmids have many features made them exceptionally advantageous as a cloning vehicles, since they are found in multiple copies within the bacterium, and they can replicate autonomously with respect to the bacterial DNA. The complete sequence of plasmids has been identified. Hence, the locations of the restriction enzymes cleavage sites have been identified as well. Moreover, the plasmids are smaller than the host chromosome. Therefore, they can easily be separated from the later. Add to that, the desired DNA is easily to be isolated by cutting the plasmid with an enzyme specialized to a restriction site at which the original DNA fragment has been inserted. Plasmids were the first cloning vectors. They are easy to isolate and purify, and they can be reintroduced into a bacterium by transformation. Plasmids often bear antibiotic resistance genes, which are used to select their bacterial hosts. A recombinant plasmid containing foreign DNA often is called a chimera, after the Greek mythological monster that had the head of a lion, the tail of a dragon, and the body of a goat. One of the most widely used plasmids is pBR322. This plasmid contains the resistant genes for the both antibiotics, ampicillin and tetracycline and many of restriction sites that exist only once in the plasmid and located within the antibiotic resistance gene (figure 10.9). This arrangement helps in the detection of the recombinant plasmids after performing the transformation. For example, if the foreign DNA is inserted into the ampicillin resistant gene, the plasmid is unable to confer resistance for ampicillin after this insertion. Thus, the transformants that lost resistance to ampicillin contain "chimeric" or recombinant plasmid. As it shown previously (see chapter nine), the plasmid DNA don't usually code for any essential function, and the plasmids free cells can divide normally. Plasmids usually exist in multicopy form, and they can constitute a means to amplify the number of any essential single-copy gene that becomes inserted into their genome. For example, insertion on a plasmid can lead to a rapid amplification of genes needed to achieve high-level of resistance to antibiotics. The significance of what is mentioned above in relation to genetic engineering is that any gene that small plasmids carry is present in large numbers, thereby making it possible that large numbers of the corresponding protein product (of the gene) can be made. Such plasmids thus provide the ideal vehicles to carry the many multiple genes needed for high-level antibiotic resistance, where detoxifying enzymes are required in huge amounts.

Most of the plasmid vectors are multicopy, and one of the earliest plasmid vectors was pBR322, which is one of the most known plasmid vectors since it carries two antibiotic resistance genes, *ampR* and *tetA* that act as markers to select and characterize the recombinants containing these plasmids. In addition to its containing to unique regions for many restriction enzymes that are available to insert DNA fragments. For instance, let us see what is going on when *Pst* I fragment is inserted into *Pst* I restriction site on pBR322 vector, which is located within *ampR* gene. Such insertion will never usually destroy the function of the corresponding gene. Consequently, an active β -lactamase is never further been synthesized. Therefore, the recombinant plasmids can be detected by their ability to confer resistance to tetracycline and not to ampicillin.

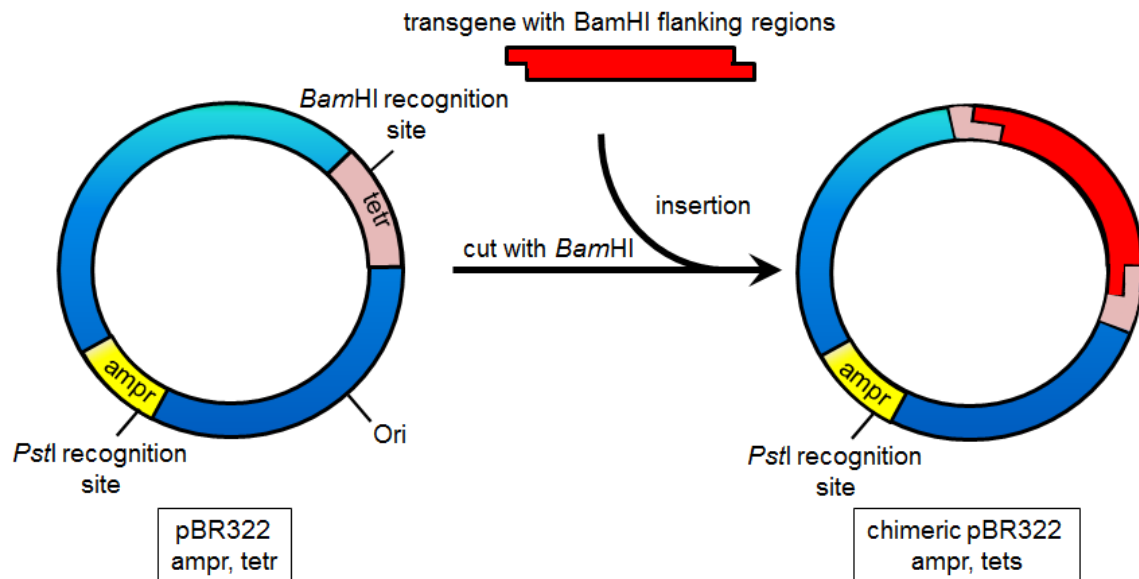


Figure (10.9); structure of pBR322 plasmid, and the mechanism of insertion of the gene of interest in it.

As for *Bam* HI, the same consequences are met, in such away when the foreign DNA fragment is inserted in *Bam* HI restriction site, the gene that confers resistance to tetracycline for the host bacterium is destroyed. After the transformation of the bacterial host with recombinant plasmid, the bacterium is placed into a growing medium containing ampicillin. Other copy is grown is grown on a medium containing tetracycline to test their sensitivity to tetracycline. This facilitates the obtaining of the gene of interest through eliminating these bacteria that don't have plasmid at all, or these that contain original plasmid without addition.

- b. **Phage Vectors;** phages have linear DNA molecules, in which the DNA can be inserted through many restriction sites. The recombinant DNA is accumulated after the phage proceeds through its lytic cycle and produces the mature phage particles. The essential advantage for phage transport on the plasmid vehicles is their ability to transfer longer DNA molecules, when the plasmid accommodate DNA fragment. However, when the

plasmid vectors accept a 6 – 9 kb DNA fragment, the phage vectors can accept a DNA fragment of about 9 – 20 kb.

Bacteriophage lambda, which infects *E. coli*, has been widely used as a cloning vector. Lambda is a well-characterized virus with both lytic and lysogenic alternatives to its life cycle (see chapter six). Although lambda DNA circularizes for replication and insertion into the *E. coli* chromosome, the DNA inside the phage particle is linear (figure 10.10). At each end are complementary 12 bp long overhangs known as *cos* sequences (cohesive ends). Once inside the *E. coli* host cell, these pair up and the cohesive ends are ligated together by host enzymes forming the circular version of the lambda genome.

Only DNA molecules of between 37 and 52 kb can be stably packaged into the head of the lambda particle. Small fragments of extra DNA may be inserted into the lambda genome without preventing packaging. However, to accommodate longer inserts it is necessary to remove some of the lambda genome. The left hand region has essential genes for the structural proteins and the right hand region has genes for replication and lysis.

The middle region of the lambda genome is non-essential and may be replaced with approximately 23 kb of foreign DNA. Since the middle region of lambda has the genes for integration and recombination, such lambda replacement vectors cannot integrate into the host chromosome and form lysogens by themselves.

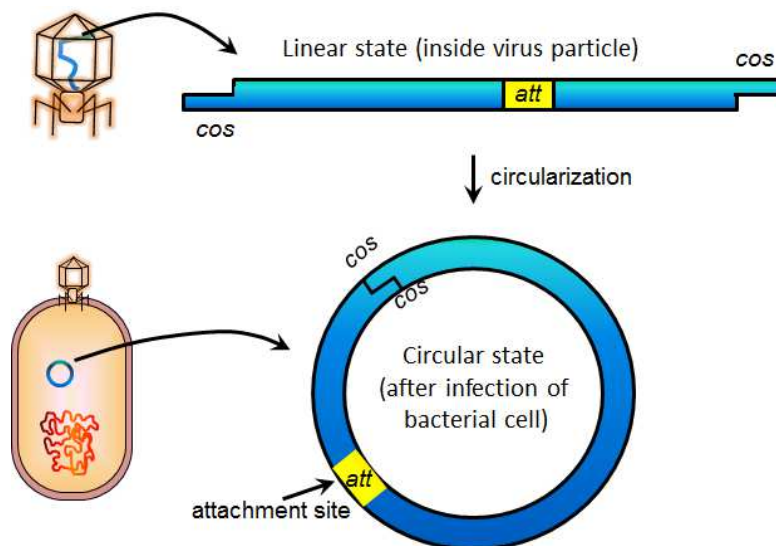


Figure (10.10); the genetic material of lambda phage, which is represented by a linear DNA molecule with two *cos* sequences at each end. After the phage injects its DNA into the bacterial host, the DNA circularizes. The two cohesive ends base pair and are ligated together by bacterial enzymes so forming a circle.

The formation of a recombinant molecule between the foreign DNA and lambda phage is much more complex than that taken place in the plasmid vectors. It's not restricted on the integration of the foreign DNA in the middle of the phage to form the linear recombinant DNA of two cohesive ends. Instead, this molecule has to be packaged into the head of the phage efficiently to form a recombinant phage able

to infect the bacterial cells. Hence, the *E. coli* cell requires to what is known *in vitro* packaging (figure 10.11).

In this technique, a mixture of lambda proteins is mixed with the recombinant lambda DNA *in vitro* to form phage particles. Infecting two separate *E. coli* cultures with two different defective lambda mutants generates the necessary lambda proteins. Each of the two mutants lacks an essential head protein and cannot form particles containing its own DNA. A mixture of the two lysates gives a full set of lambda proteins and when mixed with lambda DNA can generate infectious phage particles capable to infect new *E. coli* cells (figure 10.11).

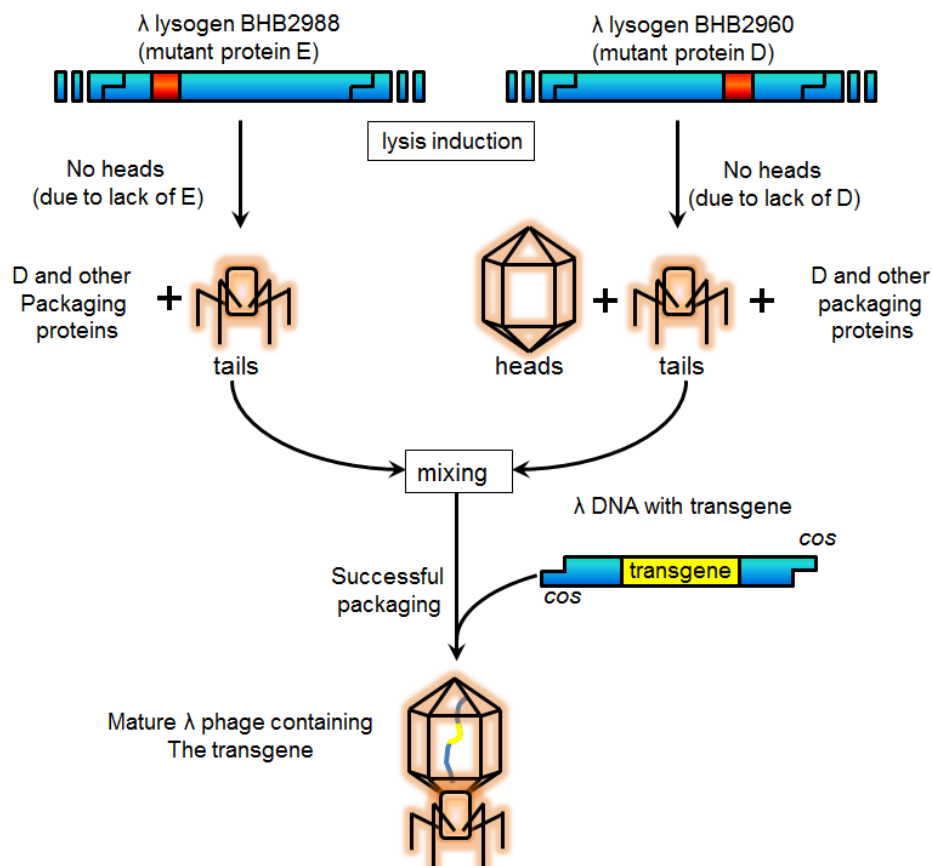


Figure (10.11); In Vitro Packaging of Lambda Replacement Vector A lambda cloning vector containing cloned DNA must be packaged in a phage head before it can infect *E. coli*. Before the DNA can be packaged, the phage head proteins must be isolated. To do this, a culture of *E. coli*, is infected with a mutant lambda which lacks the gene for one of the head proteins called E. A different culture of *E. coli* is infected with a different lambda mutant, which lacks phage head protein D. Both *E. coli* cultures are grown with the mutant lambdas and the viruses are induced to enter the lytic cycle. Although the *E. coli* are lysed by the phage, they cannot form complete heads. Instead a soluble mixture of phage proteins is isolated. Each lysate contains phage tails, assembly proteins, and components of the heads, except either D or E. These two lysates are mixed along with the lambda vector containing the cloned DNA. Although mixing is done *in vitro*, the components can self-assemble into a functional phage that can infect *E. coli*.

These phages have often been used to generate the genomic libraries. As for the single stranded phages (such as M13), they are used in the determination of DNA sequences (see chapter eleven).

- c. **Cosmid Vectors;** Cosmids are small multicopy plasmids that carry *cos* sites that derived from lambda phages and can be packaged into the heads of phages (figure 10.12 A). The lambda genome contains a recognition sequence called a *cos* site (or cohesive end) at each end. When the genome is to be packaged in the head, it is cleaved at one *cos* site and the linear DNA is inserted into the head until the second *cos* site has entered. Thus any DNA inserted between the *cos* sites is packaged. Cosmids typically contain several restriction sites and antibiotic resistance genes.

The only DNA requirements for *in vitro* packaging into λ phage are the presence of two *cos* sites that are separated by 37–51 kb of intervening sequence. Cosmids were developed in light of this observation, and are simply plasmids that contain a λ phage *cos* site. As plasmids, cosmids contain an origin of replication and a selectable marker. Cosmids also possess a unique restriction enzyme recognition site into which DNA fragments can be ligated. After the packaging reaction has occurred, the newly formed λ particles are used to infect *E. coli* cells. The DNA is injected into the bacterium like normal λ DNA and circularizes through complementation of the *cos* ends. The lack of other λ sequences means, however, that λ infection will not proceed beyond this stage. The circularized DNA will, however, be maintained in the *E. coli* cell as a plasmid. Therefore selection of transformants is made on the basis of antibiotic resistance and bacterial colonies (rather than plaques) will form that contain the recombinant cosmid. Since λ phage particles can accept between 37 and 51 kb of DNA, and most cosmids are about 5 kb in size, between 32 and 47 kb of DNA can be cloned into these vectors. This represents considerably more than could be cloned into a λ vector itself. Therefore, using a small cosmid, of say 4 kb, allows inserts of up to about 45 kb to be cloned. Cosmids, like plasmids, are very stable, but the insertion of large DNA fragments can mean that recombinant cosmids are difficult to maintain in a bacterial cell. Repeat DNA sequences are common in eukaryotic DNA, and DNA rearrangements can occur *via* recombination of the repeats present on the DNA inserted into the cosmid. The major difficulty in working with cosmids is, however, the production of linear, ligated DNA fragments in which the cosmid and insert are concatamerized together (figure 10.12 B).

Phage or cosmid vectors are used in the construction of genomic libraries. Usually, the DNA library is a group of cloned DNA fragments in the cloning vector that can be used to search the DNA of interest. There are two types of these libraries are used to obtain the DNA of interest. The first type is called genomic DNA libraries. This library is made from the genome of the organism of interest. For example, the genomic mouse library is made by the partial digestion of the nuclear mouse DNA by restriction enzyme to produce large number of different DNA, but all of which have homologous cohesive ends (see below).

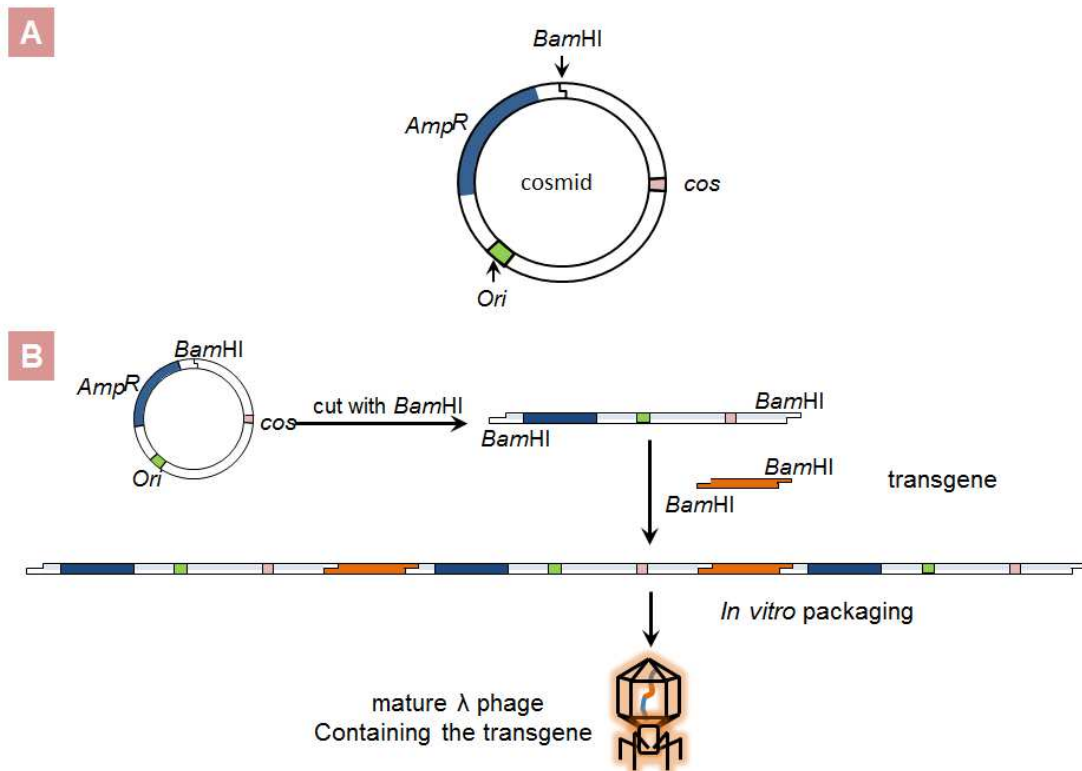


Figure (10.12); DNA cloning in cosmid vector. To clone large pieces of DNA into cosmid vectors, both must have compatible sticky ends. The cosmid vector is first linearized so that each end has a *cos* site. Then the linear cosmid is cut with *Bam*HI, which generates sticky ends with the overhang sequence GATC. The genomic DNA from the source of interest is also digested. Instead of *Bam*HI, this DNA is partially digested with *Mbo*I, which also generates a GATC overhang. Partial digestion leaves some sites uncut and allows large segments of a genome to be isolated. These segments are mixed with the two halves of the cosmid and joined using ligase. The final constructs are packaged into lambda particles *in vitro* and are used to infect *E. coli*.

Then, the DNA fragments are ligated by phages that derived from lambda phages or by cosmid vectors that are ligated by the same enzyme. This library contains all the mouse nuclear sequences and any desired gene in the mouse can be looked for. Its noteworthy to say to not each clone contains a whole gene since – in many cases – restriction enzymes make their cuts within the gene. The second type of DNA libraries is called cDNA libraries. The cDNA library can be made by using reverse transcriptase that is derived from retroviruses (see below).

- d. **Artificial Chromosomes;** In addition to the three major types of vectors, which are plasmids, bacteriophages, and cosmids, artificial chromosomes may considered as an important type for cloning vectors. The artificial chromosomes are common tools in gene transfer since they contain large quantities of genetic material. Yeast artificial chromosome or YAC is one of the widest artificial chromosomes used. The YACs are

DNA extensions contain all the necessary elements to spread chromosomes in the yeasts, which are the origin of replication, centromere, and telomeres (see chapter two). They also have sites for restriction enzymes and genetic markers where they can be tracked and selected. The YAC is cleaved by suitable restriction enzyme, which cleaves it and allows a fragment of foreign DNA to be inserted between the centromere and telomere. Thus, the YACs contain DNA fragments between 100kb into 200kb in size, and can be placed in *Saccharomyces cerevisiae*, and can be replicated side by side with the real chromosomes.

The bacterial artificial chromosomes are alternative cloning vehicles, their utilization is increasing proportionally. The BAC are cloning vehicles based on F plasmids of *E. coli* (see chapter nine; F factors), and contain suitable sites for restriction enzymes and selectable marker, such as the resistance for chloroform. The modified plasmid is cleaved in the restriction site, and the foreign DNA fragment is ligated that estimated to be more than 300kb using the DNA ligase. The BACs proliferate in *E. coli* after its entry by the electrical pulses or electroporator (see below). It's possible for this vehicle to be reproduced and modified easily, and doesn't suffer from any recombination as it's noticed in YACs. The recombination being don't occurring has an advantage that held in the fact that the integrated DNA can't usually be rearranged in such away it keep its own sequence without being changed. Since the BACs vectors can carry large fragments of DNA, the artificial chromosomes are particularly important in the identifying the genomic sequences.

Fourth; Tools of Foreign Genes Entry into the Cells

After constructing the recombinant DNA molecule, which usually consists of the DNA of interest inserted into a suitable cloning vehicle, it must introduce this vector into a suitable host to produce large quantities of it. This is done in order to be studied by the genetic engineers. Generally, the hosts are divided into two major types; prokaryotic hosts and eukaryotic hosts.

a. Genes entry into prokaryotic cells

Actually, it's possible to introduce both prokaryotic and eukaryotic genes in bacteria, but the way that is used in each of them differs from each other. In prokaryotes, such as bacteria, after isolating the desired DNA fragment, its ligation into the vector by suitable restriction enzyme and its forming the recombinant transfer molecule, this molecule is introduced into a suitable host usually by transformation (see chapter nine). The actively growing cells for the bacteria of interest, such as *E. coli*, in diluted solution of calcium chloride (CaCl_2) in an ice surrounded medium, which relatively increases the ability of bacterial cells to uptake foreign DNA (figure 10.13). The exposure to CaCl_2 is usually for 30 min. the incubation of this medium for a short period allow cells to uptake the foreign DNA and to form the transformants.

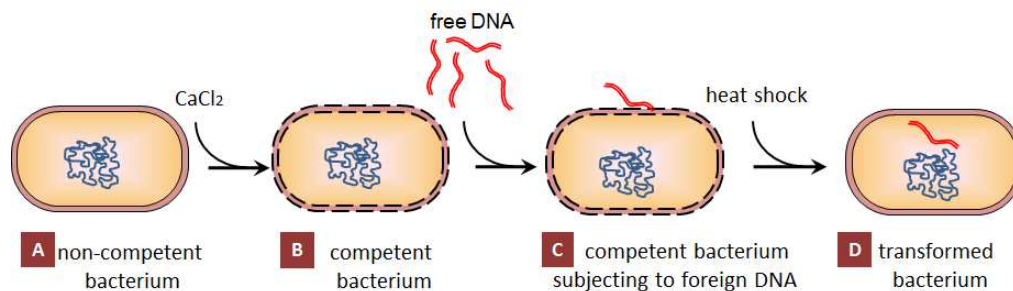


Figure (10.13); the chemical transformation of bacterial cells. The treatment of cells with calcium ions can make these cells competent to uptake DNA. The DNA may be adhere on the surface of the bacterial cell, and by a heat shock, the entry of foreign DNA is taken place.

After the entry of the recombinant vector molecules into the recipient cells, we must refer to the fact that small ratios of the cells are transformed, as for others, they fail in uptaking the foreign DNA (the new plasmid). Hence, the necessity to the presence of some ways to enable us to select the truly transformed cells is emerged. Therefore, the vectors (such as plasmids) that are used in DNA inserting must contain some selectable markers, such as the resistance for some antibiotics.

After isolating the gene of interest, its ligation into the suitable cloning vector, and its entry into the host cell by certain way, a final step remains, which is the preparation of the suitable circumstances for this gene to express about itself, and to produce the desired protein. it must refer that the cloned gene is not usually expressed in the host cell without the presence of certain modifications ensure the expression of this gene in the host cell. To be expressed, the foreign gene that is inserted in the recombinant molecule must contain promoter that must be recognized by the host RNA polymerase. The translation of the mRNA copy depends on the presence of leader sequences, and on the suitable mRNA modifications to ensure its binding with ribosomes. As its discussed previously (see chapter five), these diverse processes in prokaryotes are widely different from these present in eukaryotes. When the foreign DNA is introduced in prokaryotic cells, such as bacteria, it must be supplied with the leader sequence in order to assist it in protein synthesis in bacteria. Eventually, introns must be removed from eukaryotic genes since the prokaryotic host doesn't excise the introns after mRNA transcription. Consequently, the eukaryotic protein cannot be active without removing its introns before translation.

The problems of foreign genes expression have been largely conquered in the host cells by the help of special vectors called expression vectors. Usually, these vectors are derived from pBR322, and contain the necessary transcription and translation signals. They have useful restriction sites for these sequences, where the foreign DNA can insert in. some expression vectors contain portions of *lac* operon (see chapter six), which can efficiently regulate the expression of the cloned genes in the same mode of operon. As for to the entry of eukaryotic genes into bacteria, since the beginning of genetic engineering, many scientists were excited in this

respect particularly after the success of some prokaryotic genes to express about themselves in the bacterial cells. After the discovery of the universality of the genetic code, in which the genetic code gives the same amino acid whether in prokaryotes or eukaryotes (see chapter five). They tried, for this reason, to introduce the eukaryotic genes in bacteria in the same way used in the introduction of prokaryotic genes. These trials were taken place for the hope to succeed these genes to express about themselves in bacteria, but nothing was obtained. These failure trials were continued until the fact of introns was discovered in the eukaryotic genes (see chapter two). Since, there is no system in bacteria that can remove introns and connect exons together (see chapter four), therefore, a tool that enable the researchers to remove the unwanted introns before the entry of eukaryotic genes into bacterial host must be found.

In eukaryotic cells, many genes are transcribed into only mRNA in specialized cellular types. For example, the mRNA molecules that code for globin proteins are exist in the reticulocytes only. Similarly, the albumin coding mRNA, produces the essential protein in the serum, is only produced from the liver cells during albumin synthesis. The specialized DNA sequences that code for mRNA in the particular cell can be cloned by synthesis of DNA copies for the isolated mRNA from that type of cells, then, these DNA copies are inserted in the plasmid vectors, or in the other suitable cloning vectors.

The DNA copies of mRNA molecules are called complementary DNA (cDNA). In addition to its representation for the sequences that code for mRNA in a particular cell type, the cDNA losses the noncoding introns that present in the DNA genomic sequences. Thus, the amino acid sequences of the protein can directly be determined from the nucleotides sequence for its corresponding cDNA. Many of genes of higher organisms are considered very large and they cannot be inserted into plasmids and even in lambda phages, because of their long introns that intervening the exons of these genes. This in turn, made the researchers in this field to construct cDNA for the genes to be cloned (see chapter eleven).

b. Genes entry into eukaryotic cells

The researchers in the field of genetic engineering were devoted all of their every effort in order to develop the techniques for genes entry into the eukaryotic cells. All of these trials were performed because of what of this process have in applicable importance, where these cells of higher organisms enjoy with a unique sort of processing that could not be provided in the prokaryotic cells. Many of these techniques were used successfully in transforming many of the prokaryotic cells. These require to special treatments each of its type. For example, some of yeasts, fungi, and plant cells have cell walls that are required to be digested to produce the protoplast (cell that has had its cell wall removed) before the DNA is being uptake, while the animal cells don't contain any thick cell wall to surround them. This in turn, facilitates the entry of genes form this aspect. Hence, the variability of treating these multiple origin eukaryotic cells is emerged according to their physical nature.

There are many methods to introduce the foreign DNA into the eukaryotic cells; the most common are summarized by the following:

1. **DNA uptake (co-precipitation) technique;** the general method to introduce the foreign DNA into the mammalian cells involves co-precipitation of the DNA with calcium phosphate, then, the mixture is presented to the cells of the cultured medium (figure 10.14). The individual DNA usually inserts as multiply copies in the cellular genome.

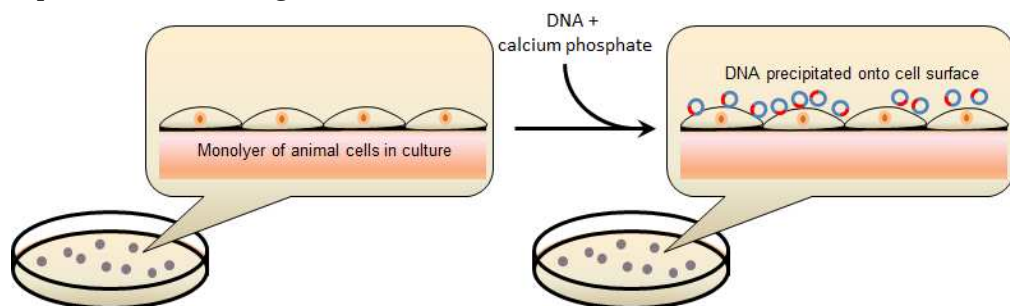


Figure (10.14); the strategy of foreign DNA entry through its precipitation on to the animal cells.

2. **Electroporation technique;** to increase the efficiency of DNA uptake, the electroporation is frequently used. This procedure is usually applied in yeasts, fungi and in plant cells, and less frequently in animal cells. In this procedure, the cells are subjected to a brief electrical pulse by a pulser device, which causes a localized transient disorganization and breakdown of the cell wall, making it permeable for diffusion of recombinant DNA molecules (figure 10.15). there are many "noticeable" advantages of this physical procedure perhaps one of the most prominent advantages is its high velocity, its low cost compared with other microinjection techniques. In addition to its ability to produce high percentage of genetically engineered cells. Despite of its ability, these high power electrical fields have harmful physical effects on the survival of these subjected cells. Therefore, the researchers want to optimize the conditions of this technique as much as possible to avoid the damage of the cells that are intended to be genetically engineered.

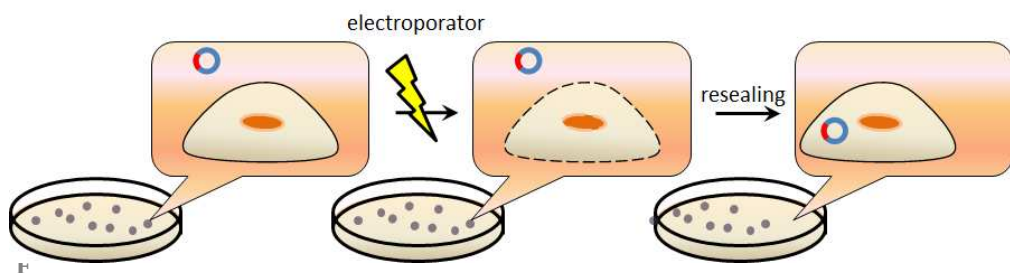


Figure (10.15); electroporation technique. The electrical pulser applies calculated electrical pulses on the foreign DNA subjected cells. This leads to the formation of transient pores in the cellular membranes. This may lead to the entry of this DNA into the cells to form the recombinant cells.

3. **Protoplast generation technique;** it's performed in the wall surrounded cells. In which, the walls of these cells are removed by the addition of number of available enzymes to generate what is called protoplast in the optimum conditions (figure 10.16). After this stage, the foreign DNA is added into protoplast. Then, the protoplast is regenerated into its normal state when the suitable conditions are available. This method suffers from the difficulty of reforming protoplast into the normal cellular state.

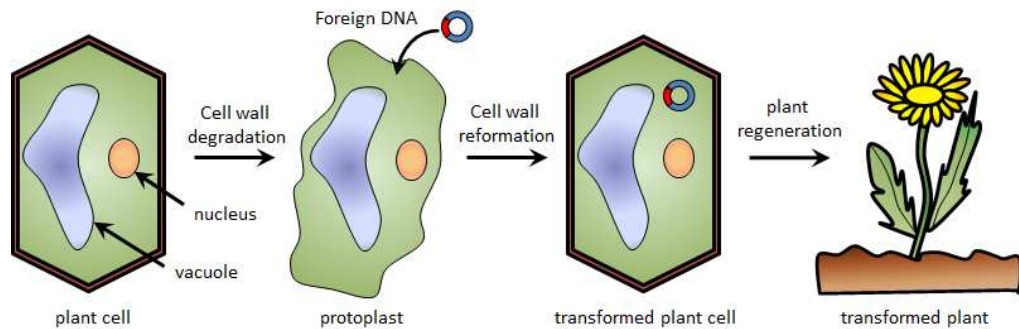


Figure (10.16); the entry of foreign DNA into the protoplast cells.

4. **Agrobacterium injection technique;** it which Ti plasmids are used for plant cells. One of the most used vectors in genetic engineering in plants is Ti plasmid, which means "tumor inducing plasmids" that exists in *Agrobacterium tumefaciens*. In dicotyledons, this plasmid produces tumor cells known as crown galls. It's possible for this plasmid to be altered, where it can carry a passenger DNA into plants without transforming cells into tumor. Within Ti plasmid, there are genes in a fragment called T DNA, which means "transfer DNA". This region contains the genes that code for the necessary enzymes for abnormal amino acids synthesis, opines, and other materials that induce abnormal cellular growth (i.e. tumor) (figure 10.17).

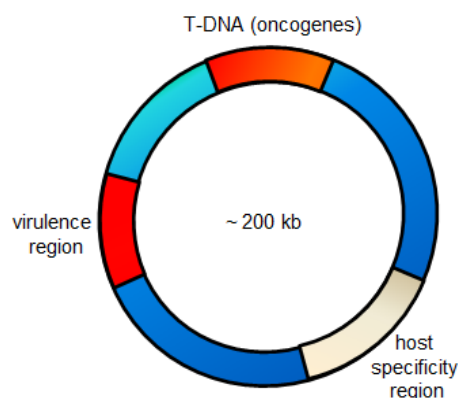


Figure (10.17); the structure of the normal Ti plasmid that exists in *Agrobacterium tumefaciens*.

This plasmid has been isolated from *Agrobacterium tumefaciens*, which is a bacterium that grows in the soil and infects plants to cause a tumorigenic disease called crown galls disease (figure 10.18). in the case of infection, a

small fragment (about 20kb) is transferred that is called T DNA, which is exist in Ti plasmid, and integrates on the plant chromosome. The transfer process is controlled by *vir* (virulence) gene that is located in Ti plasmid.

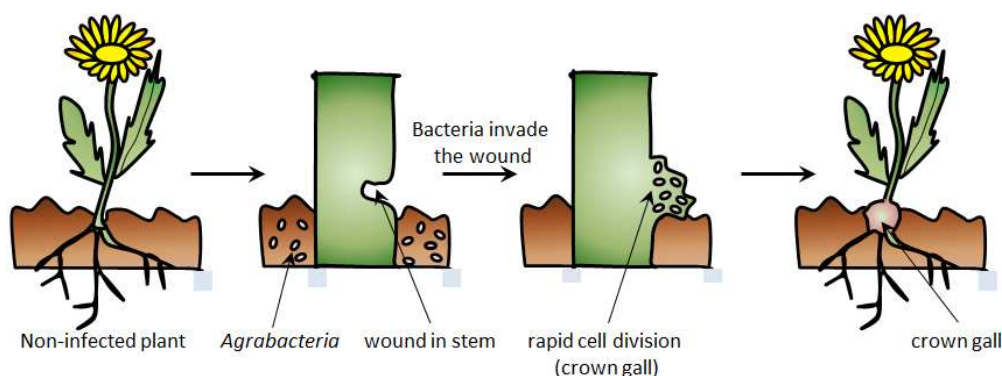


Figure (10.18); the method of plant cells infection by *Agrobacterium tumefaciens* and the correlation of this with the inflection of crown gall disease.

The size of Ti plasmid makes it not suitable to act as cloning vector. This is at least because of two reasons; (1) when plant cells infected with Ti plasmid, they converted into tumor cells that cannot be converted into normal plants. (2) The size of Ti plasmid is 150 – 200 kb make it so difficult to be genetically manipulated. Therefore, two strategies were developed to clone the genes by this plant vectors. The first strategy is known as co-integration vectors, while the second strategy is known as binary vectors. In the first one, the new DNA integration on Ti plasmid is produced from the recombination of small plasmid vector, such as *E. coli* vectors, with Ti plasmid of *Agrobacterium*. The recombination takes place between two homologous regions that localized in both plasmids (figure 10.19 A).

In the second strategy, a binary vector system is designed, which consists of helper plasmid and donor plasmid. The helper plasmid is a Ti plasmid, but in a disarmed form because of the deletion of T DNA fragment (tumorigenic genes carrier) entirely. Donor plasmid is a plasmid expresses in *E. coli*, and carries a brief fragment of T DNA. This means, that the two plasmids work together in a complementary way to each other, in such away the donor plasmid carries the integrated gene surrounded with the terminal sequences of T DNA that are specialized in excision, whereas helper plasmid provides the enzymes (that coded by *vir* genes) that are necessary to direct the transfer of the recombinant T DNA (that carried the inserted gene to be cloned). This is applied practically by carrying the *Agrobacterium* strain the disarmed helper plasmid, while the donor plasmid is injected in the bacterium, this leads to the formation of the binary cloning system (figure 10.19 B).

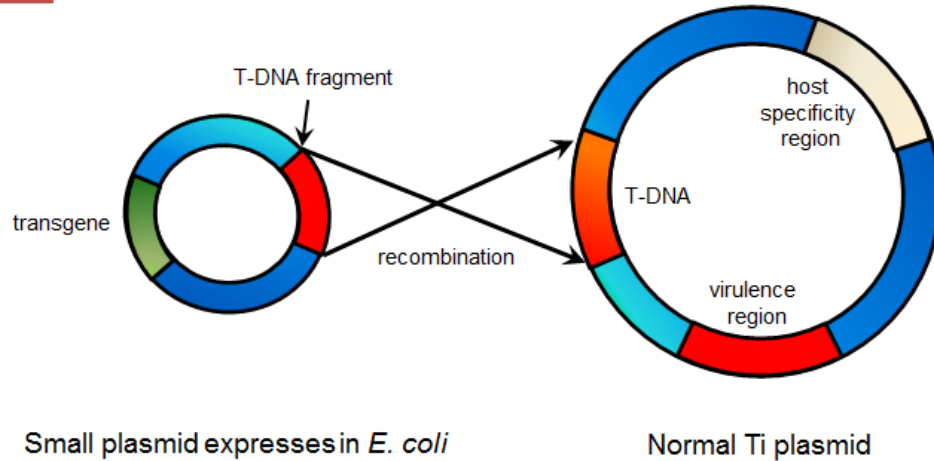
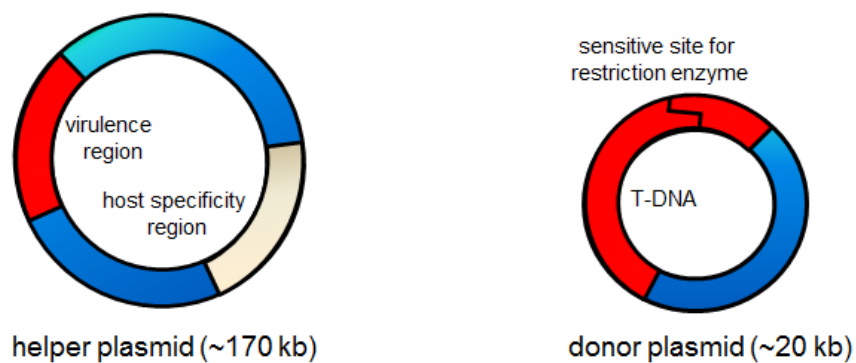
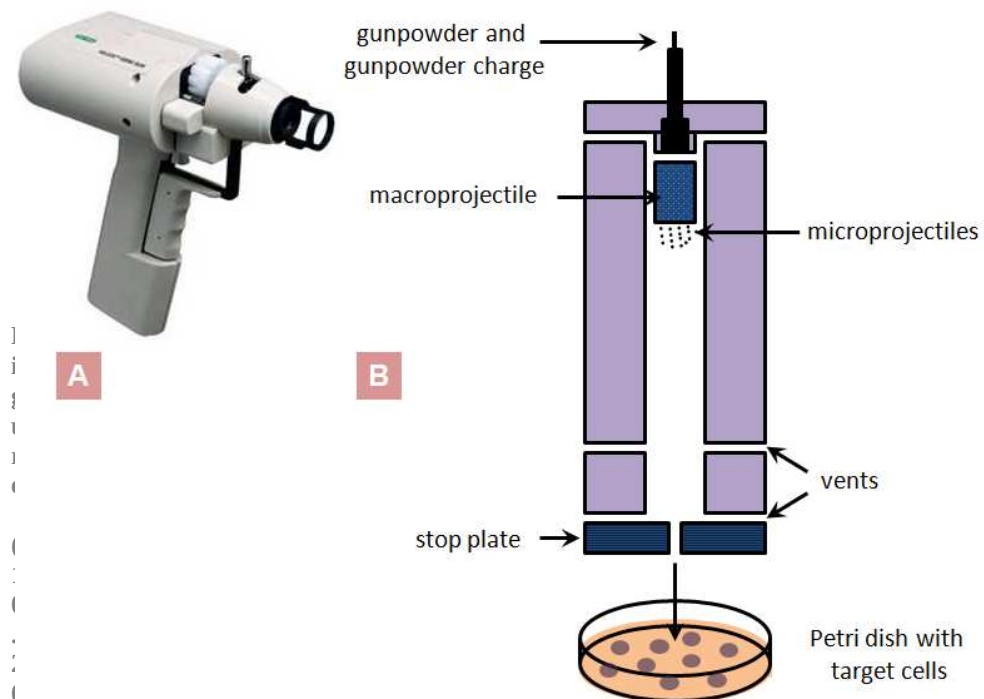
A The co-integration vectors system**B** The binary vectors system

Figure (10.19); cloning strategy using cloning vehicles that are derived from Ti plasmid. (A) co-integration vectors system, in which the small plasmid carry the gene to be cloned directly, and the gene is transferred by recombination with the normal Ti plasmid. (B) binary cloning system; the donor and helper plasmids complement each other when present together in the same *Agrobacterium tumefaciens* cell. The T DNA carried by donor plasmid is transferred into chromosomal DNA by proteins coded by genes carried by helper plasmid.

5. **Biolistic technique;** the name of this technique is came from the word "biological ballistic". The gene gun is a device which "literally" shoots recombinant DNA into the target cells. This technique is a direct physical method to genetically modify cells, in which a thin coat of DNA is deposited onto the surface of 0.5 -1.5 μm of tungston or gold microbeads. The DNA-coated beads are then loaded and fired from a gene gun by explosive, electric, or pressure charge. The DNA-coated beads are bombarded onto plant tissues, enter

the cells, and are integrated into the cell chromosomal DNA randomly (figure 10.20 a).

The gene gun is used particularly in infecting cells that are difficult to be infected with nucleic acids in other methods such as plant cells. Although this technique has been originally developed for plant cells, it can be applied in animal cells, tissues, organs, yeasts, and even in other bacterial and microbial organisms.



); the gene gun (biological ballistic firing device) that used to fire the coated DNA beads into plant cells. (a) Gene gun device (Bio-Rad corporation – USA), (b) an illustration shows the mode of biological ballistic firing by this device. The DNA is coated onto microprojectiles, which are accelerated by the macroprojectile on firing the gun. At the stop plate the macroprojectile is retained in the chamber and the microprojectiles carry on to the target tissue.

6. **Pronuclear microinjection technique;** this technique was regarded as one of the most effective method by which some transgenic animals can be generated. This technique is done by injecting male pronucleus of fertilized oocyte using a specialized micropipette for this purpose (figure 10.21).

The reason for injection of the transgene into the male pronucleus attributed to the fact that male pronucleus is larger than female counterpart. Off course, this facilitates the mission of the researcher who intends to inject the gene properly. For expression purposes, the gene of interest must be properly constructed with a promoter region and other control elements to direct tissue-specific production of the protein. The transformed zygote is implanted into a surrogate mother to give birth to transgenic offspring.

This technique is the most direct way in this field to generate transgenic animal directly as in zygotes and sometimes the foreign DNA is stably integrated into the host genome to generate the transgenic animal. Very fine glass micropipette is used in this type of transfer, through which the transgene is injected into the newly fertilized cell that stabilized on by a holding pipette.

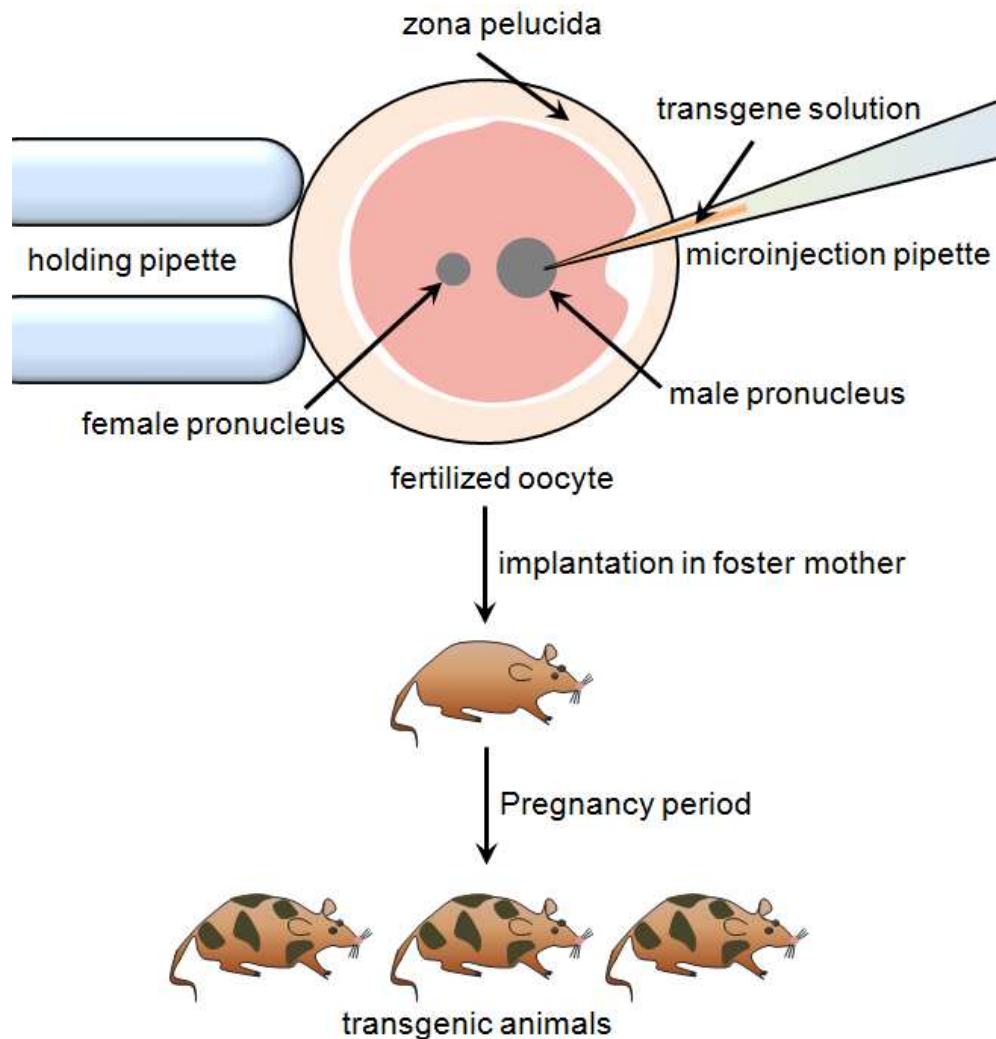


Figure (10.21). Illustration of pronuclear microinjection. the desired DNA is pipetted into the pronucleus by the micropipette that used to pierce oocyte. Following withdrawal of the micropipette, surviving oocytes are reimplanted into the oviducts of pseudopregnant foster females to may generate new born genetically engineered babies.

It must be noted that this technique is restricted on the vertical gene transfer exclusively. The meaning of vertical gene transfer is the transfer of the gene from one generation to other one (from parents to offspring). This doesn't

occur except by manipulating germ cells. When it succeeds, a new born transgenic animal. Since this technique is specialized in the injection of the foreign gene into the male pronucleus of oocytes, which of course are germ cells, it cannot be used in horizontal gene transfer, i.e., gene transfer through somatic cells that occurs either *in vitro* such as growing cells in Petri dish, or by injecting gene *in vivo* for gene therapeutic purposes, such as injection of viral coated gene, or injection of liposome mixed gene (see below).

7. **Viral infection (transfection) technique;** Viral-mediated transfer provides a convenient and efficient means of introducing eukaryotic genes into mammalian cells. This method involves the use of a number of viruses, such as simian virus 40 (SV40), bovine papilloma virus (BPV), Epstein-Barr virus (EBV), and retrovirus. Baculovirus is also included, although insect cells are used as the host in this system. The mammalian expression vectors for this purpose are derived from the regulatory sequence (promoter and enhancer) that belongs to viruses. Using this technique, the gene can be injected vertically (figure 10.22), or horizontally in the somatic cells directly for gene therapeutic purposes as it's shown above.

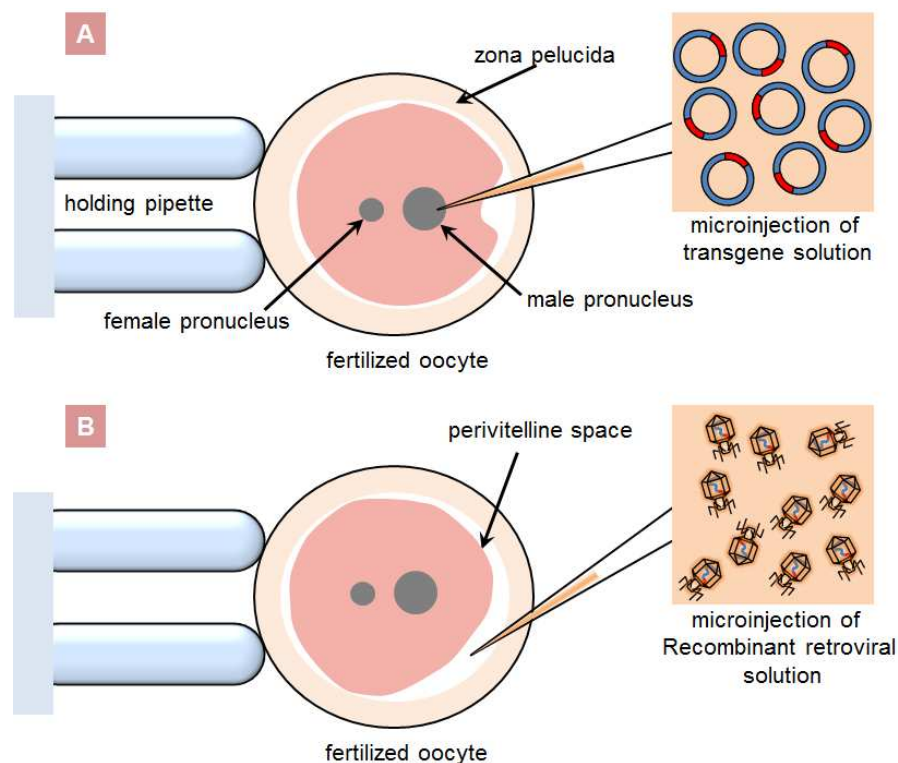


Figure (10.22); comparison between pronuclear microinjection (A) and retroviral mediated gene transfer (B), the retroviral particles have been injected in the perivitelline space that is located between zona pellucida, the embryo outer protection membrane that prevent the entry of recombinant viruses, and the oocyte membrane.

May be the most important problems that correlated with the gene transfer by this way are represented by the biohazard emerged from the practical dealing with these viruses. Although many researchers were did there every efforts to delete the hazardous biological genes, such as viral oncogenes, form these viral vectors, but the potential biological danger on the health of researchers and recipients for these therapeutics still exist. This needs to pursue the development of nonviral gene delivery techniques such as liposomes.

8. **Liposome mediated DNA transfer technique;** it's a pure chemical technique for gene transfer that depends on DNA binding with liposome. It can be used – as in viral vectors – in horizontal gene transfer. Liposomes are hollow spheres surrounded with membranes made up of fatlike molecules called phospholipids (figure 10.23).

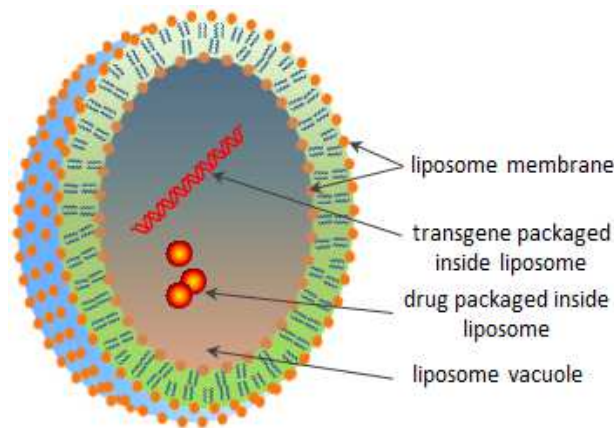
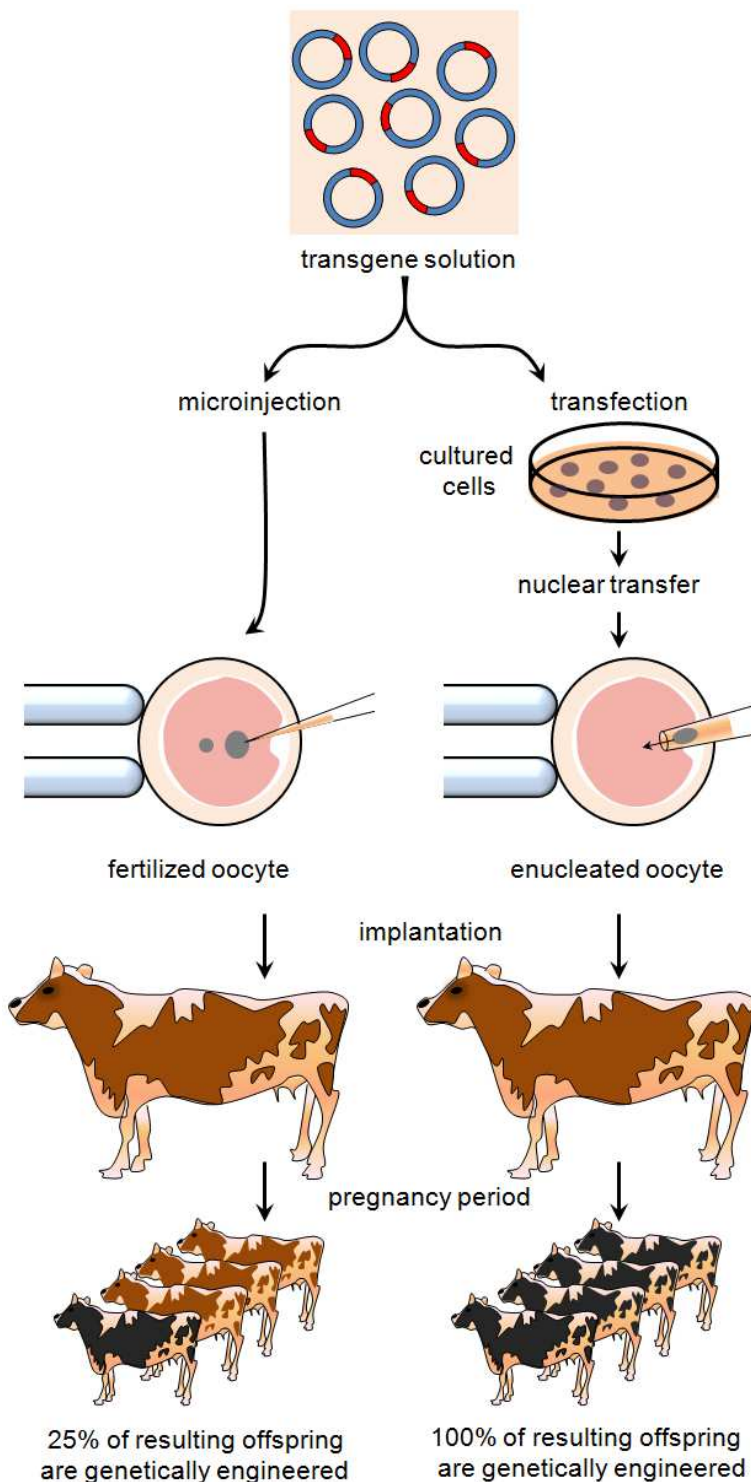


Figure (10.23); Liposomes can carry DNA genes. Like eukaryotic cells, liposomes are surrounded by phospholipid bilayer membranes. Liposomes carrying therapeutic DNA genes or drugs can fuse with the plasma membrane of a eukaryotic cell to deliver the contents inside the cell.

After mixing the genetically engineered DNA (the DNA that inserted into the suitable gene cloning vector) with these spheres, its introduced into the recipient (diseased cells) cells either by incubating these complexes with the diseased cells (*in vitro*), or by direct injection of these complexes into the organisms body (*in vivo*). When the liposome is fused with the cell membrane, it releases the DNA into the cytoplasm. By an unknown mechanism, the vector DNA is transported into the nucleus where the therapeutic gene is expressed. In addition to the efficiency of liposomes, they are not dangerous form the biological side. This, by this feature, may overcome the known viral vectors. Therefore, many types of liposomes are tested on many diseases, and they still under development till now.

9. **Somatic cell nuclear transfer (SCNT) technique;** which is known in its original meaning, the technique that can be used in somatic cell nuclear transfer (SCNT), in its original meaning, can be identified as a technique that can be used to create a genetically identical copy (clone) of an animal. In performing nuclear transfer, the nucleus is first removed from an unfertilized



egg (oocyte) taken from an animal soon after ovulation. This is accomplished by using a dedicated needle to pierce through the shell (zona pellucida) to draw out the nucleus under a high power microscope. The resulting cell, now devoid of genetic materials, is an enucleated oocyte. In the next step, the donor cell carrying its complete nucleus is fused with the enucleated oocyte. The fused cells then develop like a normal embryo, and finally implanted into the uterus of a surrogate mother to produce offspring (figure 10.24).

One of the most important examples of this technique that found huge advertisement impact is the cloning of Dolly the sheep, which is a copy of its donor mother animal (see the appendix of this chapter).

Figure (10.24); Schematic illustration shows a simple comparison between pronuclear microinjection (left) and somatic cell nuclear transfer (right). In the SCNT, all the generating offspring are transgenic since there is no nucleus except one which represents the donor nucleus. This nucleus is taken from foreign DNA infected somatic cell culture.

Some animals can be generated by SCNT and could be considered as “transgenic”, when only when the donor nuclei is originated from a cell that carries some genetic modification. Therefore, any individual animal generated by SCNT of genetically modified cells will also be defined as “transgenic” since they carry the initial modifications present in the somatic cells from which the donor nuclei have been taken. Simply, SCNT transgenic animals are generated by infecting cells with the foreign DNA fragment after taking it from the donor animal and before injecting its genetically modified nucleus in the enucleated oocyte (figure 10.24).

Fertilization of an egg by a sperm results in the formation of a zygote which ultimately gives rise to all the cells of the adult body – more than a hundred trillion cells of diverse structures and functions – through progressive developmental changes. The zygote begins the process of cleavage in which it undergoes rapid division from a single cell to 2 daughter cells, 4, 8, 16, and so on until the embryo reach a stage called the blastula. As the embryo enters into this stage, the cells become differentiated. Differentiation is a process whereby originally similar cells follow different developmental pathways into specialized cells, for example, nerve cells, muscle cells, etc. - that eventually make up the various tissues and organs of the body. And the process is all controlled by the collective actions of genes in a particular group of cells. The cells produced from the first few divisions after fertilization are undifferentiated, meaning that they can develop into any of the cell types. Undifferentiated early embryonic cells have been used as the source of choice for cloning using nuclear transfer techniques, prior to the cloning of Dolly from differentiated somatic cells (see the appendix of this chapter).

Chapter Ten Appendix

Cloning Dolly the Sheep

There was great developmental achievement in biology and medicine field at 1996 when scientists succeeded in Roslin institute in Scotland by leadership of Ian Wilmut in cloning animals from cultured cells taken from mature ewe. Dolly (1996 – 2003) is the first mammalian copy created by nuclear transfer from "mature" cell into unfertilized (enucleated) oocyte. Cloned organisms were produced from mice, cattle, sheep, goats, pigs and other animals.

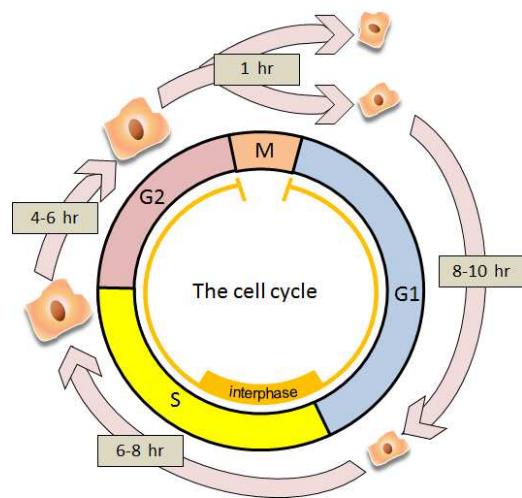
The technique of nuclear transfer was first applied to cloning frogs in 1952, but the cells never developed beyond the tadpole stage. In the mid-eighties, several research groups succeeded in producing sheep and cattle by nuclear transfer using early embryonic cells. In some later studies, cells from embryos that had advanced to 64- and 128-cell stages were used to produce calves. The major breakthrough that set the stage for creating Dolly came in 1995 when scientists at the Roslin Institute successfully produced lambs by nuclear transfer from cells taken from early embryos that had been cultured for several months in the laboratory. The experiment using cultured embryonic cells led to the cloning of Dolly using adult (differentiated) cells, which sets it apart from all previous cloning attempts of employing embryonic (undifferentiated) cells. The success of cloning adult cells



Ian Wilmut



Dolly the sheep



proves that cell differentiation is reversible, and the hands of time in a developmental process can be manipulated to reprogram its course.

The key to the success of cloning Dolly was the careful coordination of the cell cycle of the donor cell (figure 10.25). The cells used as donors for nuclear transfer were at a stage of the cell cycle called "quiescence" - the state at which the cell is arrested and stops dividing. For most cells, the life history can be represented as a repeating cycle of metaphase (M phase, see Section 1.5 on mitosis) and interphase. The cell's DNA replicates during a special portion of the interphase, called the S phase. The interphase also contains two time gaps: G1 between the end of mitosis and the start of the S phase, and G2 separating the S phase and onset of mitosis. No DNA is made during G1 and G2, however, protein synthesis occurs throughout the entire interphase.

Figure (10.25). the typical cell cycle in eukaryotic cells such as mammals.

For simplicity, one can view a cell cycle to consist of two major phases - one phase for nuclear division to form two daughter cells (mitosis), and another phase for DNA and protein synthesis. Cells that do not divide are usually arrested in the G1 phase. For the cloning of Dolly, the donor cells used were derived from the mammary gland of a 6-year old Finn Dorset ewe in the last trimester of pregnancy. The cells were made quiescent by growing in media with reduced concentrations of nutrients for 5 days. Unfertilized eggs were obtained from Scottish Blackface ewes between 28 to 33 hr after injection of gonadotropin-releasing hormone, and the nuclei were removed (figure 10.26).

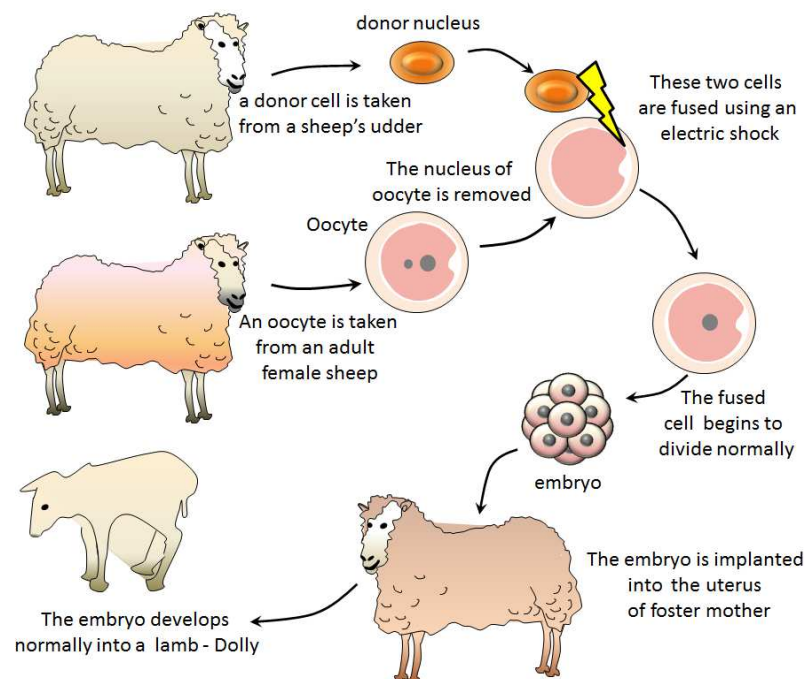


Figure (10.26). Cloning steps of Dolly the sheep. The donor nucleus was removed from the fibroblasts of mature ewe and introduced into the empty cytoplasm of the ewe enucleated oocyte to form the cloned embryo. The proteins of oocyte cytoplasm reprogram the donor mature ewe. Then, the reprogrammed genome directs the cell to be less specialized. The resulting clone (the blastocyst) is implanted into uterus of the surrogate mother, to produce "Dolly" the sheep.

The following step was to transfer the nucleus from a quiescent mammary donor cell into the enucleated oocyte. Electrical pulses were applied to induce fusion of the donor cells to the enucleated oocyte and to activate the cell to development. From a total of 277 fused cells obtained, 29 developed into embryos (at the morula or blastocysts stages). They were implanted into 13 ewes, resulting in 1 pregnancy, and one live lamb - Dolly. The success rate of the entire process was only 0.4%!

Chapter Ten Revision Questions

Q1- Justify the following:

1. The restriction enzymes represent the immunological system of bacteria?
2. It's possible for bacteria to distinguish between its own DNA and foreign DNA?
3. The ligation of sticky ends of DNA molecules is easier than the ligation of blunt ends of the same molecule?
4. In some applications, adaptors should be added when blunt ended DNA fragments are intended to be cloned?
5. Homopolymer tailing is less common than other methods in the ligation of DNA molecules termini to each other?
6. Lambda phage can successfully be modified specially through deleting its middle region?
7. The using of lambda derived gene vehicles is much more complicated than the using of plasmid vehicles in gene transfer?
8. Plasmids are usually doesn't used in the construction of genomic library for higher organisms?
9. When genomic library is intended to be constructed, it's preferred to partially digest DNA fragments with restriction enzyme?
10. In some applications, the higher organisms are favored to be used compared with *E. coli* cells as a host for foreign DNA?
11. When an animal gene is cleaved by a restriction enzyme and ligated with suitable cloning vector and this vector is introduced in *E. coli*, it fails to express on this gene?
12. The normal Ti plasmid is unsuitable to be used in cloning experiments?
13. Despite the efficiency of the gene transfer method through the genetically engineered viruses, the scientific society favors to transfer genes by a viral independent method?
14. The cloning of Dolly earned a great advertisement impact despite that Dolly is not the first clone organism in the world?

Q2- Choose the correct answer:

1. The EcoRI is one of enzymes.
a. endonucleases b. RNases c. DNases d. polymerases
2. The modification system in bacteria acts on..... DNA strand.
a. methylated b. non-methylated c. hemimethylated d. none
3. The reason behind the presence of restriction enzymes in bacteria is to participate in
a. infection b. protection c. metabolism d. endocytosis
4. If we suppose that the octanucleotide sensitive sequences that NotI distinguishes is spreaded evenly in the genome, it's possible for this enzyme to cleave one site per each bps.
a. 256 b. 4026 c. 65536 d. 1048576
5. The is a non-palindromic sequence.
a. 5'-GAATTC-3' b. 5'-GAATTCCTTAAG-3' c. 5'-GGCC-3' d. 5'-GCTACT-3'
6. The DNA ligases doesn't participate in intracellular reactions.
a. lagging strand synthesis b. excision nucleotide repair c. photoreactivation d. non-homologous end joining
7. The terminal transferase requires to to perform its mission successfully.
a. primer only b. primer + template c. template + dNTPs d. primer + dNTPs
8. The feature is not usually common in gene vehicles.
a. low copy number b. completely sequenced c. presence of one origin of replication d. easy mode of gene transfer

9. If a certain DNA fragment is inserted in ampicillin resistance plasmid, the resulting recombinant plasmid is for ampicillin.
 a. highly resistant b. moderately resistant c. not resistant d. none
10. are used in the DNA sequences determination.
 a. M13 b. λ phages c. cosmids d. pBR322
11. If a DNA fragment contains *oriC*, with centromere and telomere, then, it's possible to be called.....
 a. plasmid b. λ phages c. cosmids d. artificial chromosome
12. The method is a direct physical technique to modify cells genetically.
 a. viral transfection b. liposome mediated transfer c. biolistic transfer d. none
13. the technique is restricted exclusively on vertical gene transfer.
 a. biolistic transfer b. nuclear microinjection c. electroporation d. viral transfection

Q3- Define the following:

Cloning, clone, palindromic DNA, *cos* sequences, *in vitro* packaging, cosmids, DNA library, T DNA, chimeric DNA, liposomes, quiescence stage of the cell cycle

Q4- Connect terms exist in right with its suitable counterparts in left

λ phages	DNA transfer of 5 kb capacity
Plasmids	DNA transfer of 20 kb capacity
Bacterial artificial chromosomes	DNA transfer of 50 kb capacity
Yeast artificial chromosomes	DNA transfer of 200 kb capacity
Cosmids	DNA transfer of 350 kb capacity

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science, USA, 2008.
- Al-Shuhaib M**, Al-Saadi A., Mahanem N. Study of rabbit's transgenesis efficiency by sperm mediated gene transfer technique. PhD thesis presented to college of science, university of Babylon, 2011.
- Berk A.**, Zipursky S., Baltimore D., Darnell J., and Lodish H. Molecular Biology. Fourth edition. Paul Matuataira, USA, 1998.
- Brown T. A.** gene cloning and DNA analysis. Sixth edition, Wiley-Blackwell, 2010.
- Clark D.** Molecular biology: understanding the genetics revolution. Elsevier, 2010.
- Dale J.**, and Schantz M. From Genes to Genomes; concepts and applications of DNA technology. John Wiley and Sons, 2002.
- Dunn, L. C.** *A Short History of Genetics*. New York: McGraw-Hill. 1965.
- Fitzgerald-Hayes M.** and Reichsman F. DNA and Biotechnology. Third edition, Elsevier, 2010.
- Hanson B.**, and Jorde L. Kaplan USML Step 1 lecture notes. Kaplan Incorporation, 2001.
- Houdebine L.** Animal transgenesis and cloning. 2003, John Wiley & Sons Ltd.
- Klotzko A.** The cloning source book. Oxford University press/2001.
- Kumar H. D.** Molecular genetics and biotechnology. Second edition, Vikas Publishing House, India, 1997.
- Koolman J.** and Roehm K. Color Atlas of biochemistry. Second edition, Thieme, Stuttgart-New York – 2005.
- Lanza, R. P.**, Dresser, B. L., and Damiani, P. 2000. Cloning Noah's Ark. *Sci. Am.* 283(5), 84-89.
- Matthew, J. E.**, Gutter, C., Loike, J. D., Wilmut, 1., Schnieke, A. E., and Schon, E. A. 1999. Mitochondrial DNA genotypes in nuclear transfer- derived cloned sheep. *Nature Genetics* 23, 90-93.
- Nicholl D. S. D.** An introduction to genetic engineering. Third edition, Cambridge university press, 2008.
- Paterson, L. A.**, Wells, D. N., and Young, L. E. 2002. Somatic cell nuclear transfer. *Nature* A19, 583-586.
- Primrose S.B.**, Twyman R.M., Old R.W. Principles of gene manipulation. Blackwell science. 2001.
- Prescott L.M.** Microbiology. Fifth edition, The McGraw-Hill Companies 2002.
- Reece R.J.** Analysis of genes and genomes. John Wiley & Sons, 2004.
- Robinson, R.** Genetics. Volume 3. Thomson™ Gale/USA/2003.
- Schleif, R.** Genetics and molecular biology. Second edition, Addison-Wesley Publishing Company, 1993.
- Tamarin.** Principles of genetics. Seventh edition. The McGraw-Hill. 2001.
- Verma P.**, and Agarwal V. molecular biology. First Edition Schand group, India 2004.
- Wilmut, 1.** 1998. Cloning for medicine. *Scientific America* 279(6), 58-63.
- Wilmut, 1.**, Beaujean, N., de Sousa, P. A., Dinnyes, A., King, T. J.,
- Wong D.W.S.** The ABCs of gene cloning. Second edition, Springer, 2006. Somatic cell nuclear transfer. *Nature* A19, 583-586.

Chapter Eleven

How to Deal with the Gene



DNA samples digested by restriction enzymes and electrophoresed on the agarose gel and submitted under a source of UV to visualize the fluorescent bands due to their staining with ethidium bromide (Fitzgerald-Hayes and Reichsman,, 2010).

Introduction

In the last chapter, the essential steps of cloning and the mode of the gene entry were presented whether in prokaryotic or eukaryotic organisms. But, we didn't view the way of dealing with this genes in the practical mode. We didn't focus on the fate of the foreign DNA or the gene after its entry into these cells and how to discover this fate, since this discovery demanded different techniques, in such away a certain technique may detect only the presence of absence of the gene, whereas other technique identifies the possibility of this gene integration into the genome. Meanwhile, other technique detects the ability of this gene to gives RNA copies of itself and the amounts of these copies. Furthermore, certain techniques are used to detect the ability of this gene to express about itself and to give genetically engineered protein. it must be noted that this chapter is the most important one in this textbook on the practical site, since it shed the light on different sites of genetic engineering. Moreover, it shows that this field is not restricted on the cloning of certain gene only, but it focuses on other sites of gene technology that must be enrolled for the time being and considered as a modern basis put in the reader's eyes how the researchers practically deal with genes. This may be obvious when we make an overview about the most important techniques used in this aspect nowadays.

How to Obtain the Gene (or the Desired DNA Fragment)

The most important things in gene cloning correlates with how to find the desired sequence of DNA, since there are millions or billions of base pairs of DNA in the cell. Accordingly, there are several ways to obtain the DNA or the gene of interest.

First method (Isolation of Genetic Material and Digestion of Restriction Enzymes); after isolating genetic material, the gene of interest is separated from the rest genetic material by a suitable restriction enzyme. This method, i.e., DNA isolation, is not the only way to obtain the gene but it's the oldest way in this aspect.

DNA molecules are long, linear, and easily to be broken when the DNA suspension is passed through an narrow-gauge syringe needle. Therefore, a suitable way to obtain genetic material that contains the desired gene must be provided. After making sure form the presence of this gene in the genetic material of certain organism, the genetic material for this organism is extracted. After obtaining the DNA as a whole, the desired gene is usually isolated by suitable restriction enzyme, then, its ligated with it's the suitable cloning vector (see chapter ten).

The obtaining the DNA whether if it is chromosomal or non-chromosomal (such as plasmids) is considered as the starting point for many genetic engineering experiments. The DNA was extracted and purified through the classical techniques which concerned with the bursting the cellular extracts with the presence of detergents, and through the using of phenol (an organic material don't mixed well with water) in the removal of proteins, in which the cellular proteins are bound with and denatured, whereas nucleic acids remains in the aqueous solution in order to be withdrawn with a micropipette. As for RNA, it can be eliminated by RNase. This old method suffers from an essential problem represented by using phenol, which is a material has severe biohazard on human health, and it may cause cancer after its accumulation in the human body. The thing that brought researchers to use these dangerous materials is the need of genetic engineering experiments – as in restriction enzyme digestion and cloning – into a highly purified DNA in order to avoid nucleases, and to get rid of the possible inhibition of restriction enzymes due to contamination of the extracted DNA samples. There are other technique purifies DNA completely by using density gradient centrifugation (figure 11.1).

Despite the high precision of this technique in extracting DNA, it demands the presence of high speed centrifuge, which is very cost effect and it's difficult to be provided for this reason in any routine lab. Add to that, this process demands continuous centrifugation for at least 48 hr to separate bands sufficiently. Therefore, this technique is tedious if it's compared with other efficient low cost techniques that used for the time being largely in the separation and purification of DNA. Nowadays, many kits that involved in the extraction and purification of DNA are available. These kits provide a inexpensive, simple, and rapid tool to isolate DNA from different sources whether if it's from cells belong to microorganism, fungi, plants, or mammals. The procedure of these kits were simplified last years, in such away special spin columns with several features were exploited to isolate any desired DNA and to eliminate any other contaminants. The efficiency of these low cost columns was ensured their effectiveness in giving high pure DNA yield. For example, the separation columns play a big role in purifying plasmid DNA,

where these columns solve great problem in cloning experiments represented by contamination of plasmid DNA samples with the chromosomal DNA.

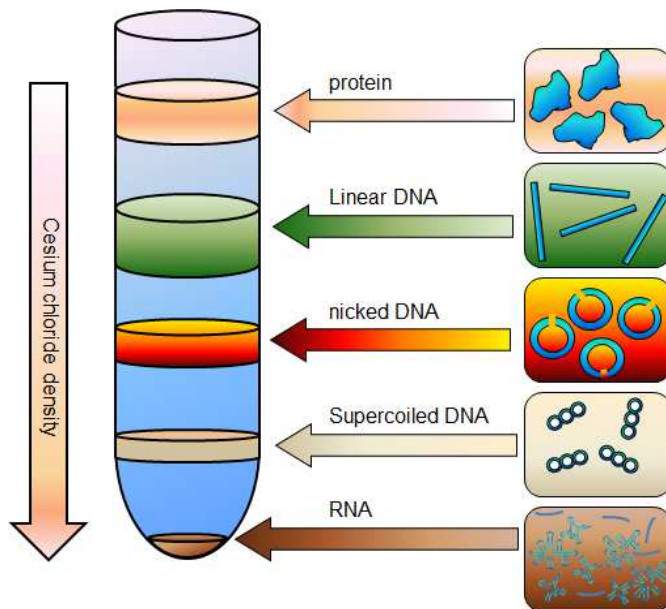
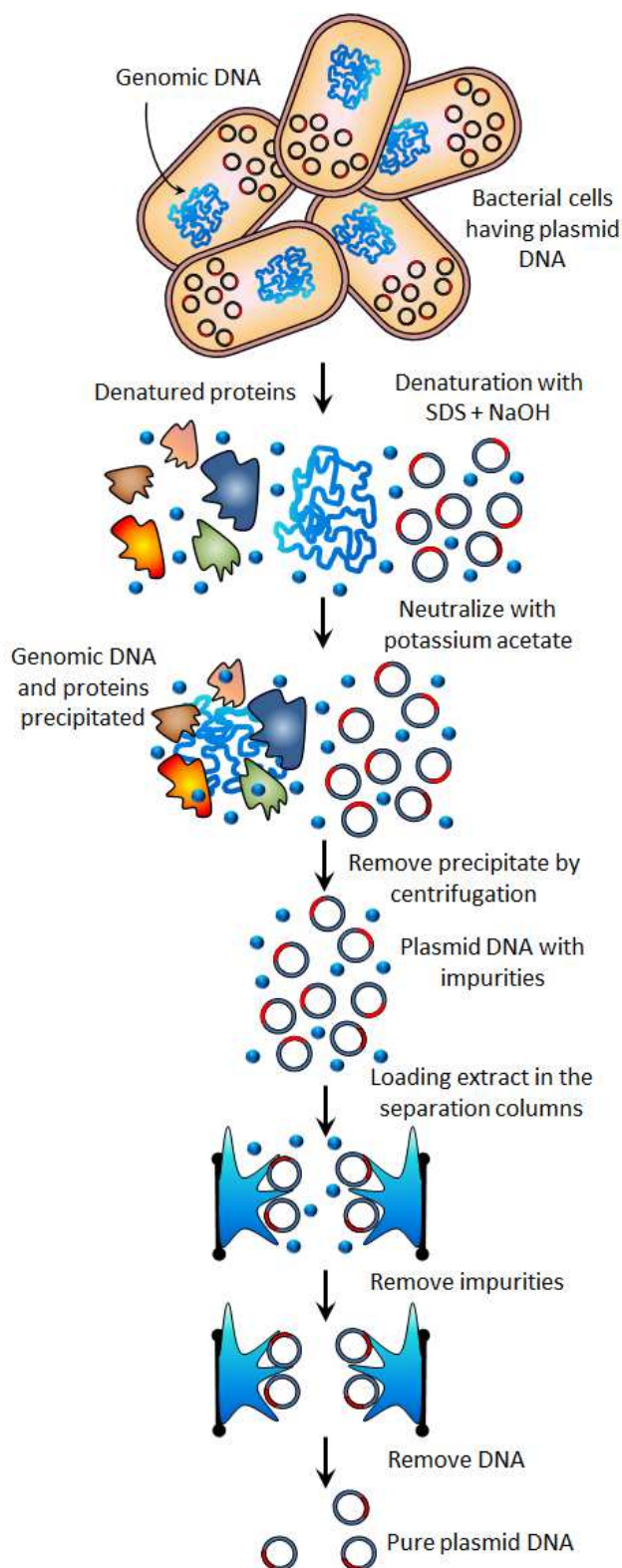


Figure (11.1). Isopycnic centrifugation of DNA molecules through a cesium chloride–ethidium bromide gradient. The DNA sample is mixed with cesium chloride and ethidium bromide and centrifuged at high speed for approximately 48 h. Each component bands at its own density, which, for the DNA, is influenced by the extent to which the molecule can be unwound by the ethidium bromide. The relaxed molecule has greater capacity to be stained with ethidium bromide due to the high ability of this stain to intercalate in the relaxed molecule compared with the supercoiled molecule. This reflects with its fluorescence proportionally.

The essential basis that most of these kits rely on in the purification of plasmid DNA is the same basis of Brinboim and Dolly at 1979, which is called alkaline lysis method, but there are different modifications simplified this process by using the separation columns. This method, termed the alkaline lysis procedure, takes advantage of the fact that at alkali pH (between 12.0 and 12.5) linear DNA, but not closed-circular DNA, will become denatured. These procedures inevitably result in shearing of the circular *E. coli* genome to generate large linear DNA fragments. Although *E. coli* bacterial chromosome is a circular chromosome, but it big in size compared with the small circular plasmids. So, the simple degradation of bacterial cell walls leads to fragmentation of their chromosome and its conversion into linear forms. Since the linear forms don't return into their normal state after their denaturation with alkaline solution compared with the circular forms (plasmids), this phenomenon is exploited in the separation of plasmid DNA from its chromosomal partner (figure 11.3). After separating plasmids, they are purified by commonly available separation columns that have the suitable matrices to adhere with the desired DNA.



Whatever the technique used in DNA extraction, there are several methods by which the extracted DNA amount can be determined, such as the reading of the of DNA absorbance at 280nm. Moreover, the purity of the extracted NDA can be identified by the dividing of the reading result at 280nm on the reading result of the same specimen at 260nm. Consequently, if the result is around 1.8, the extracted DNA enjoys with a high degree of purity. While, if it's lower, the extracted DNA is usually contaminated with proteins, and if it's higher, the extracted DNA is usually contaminated with RNA. As for the size of the extracted DNA, agarose gel electrophoresis is usually used to identify its size. It's also possible to use this technique for some expert s to estimate the purity, amount and size of the electrophoresed DNA.

Figure (11.2). Plasmid DNA purification. Bacterial cells harboring plasmid DNA are broken open using SDS. The presence of sodium hydroxide ensures the denaturation of genomic DNA. Upon neutralization, the genomic DNA aggregates and forms a precipitate together with the SDS–protein complexes. These can be removed by centrifugation. The plasmid DNA is then bound to a matrix to which contaminants are unable to adhere. The pure plasmid DNA can then be eluted from the matrix.

In agarose gel electrophoresis technique, the huge molecules are separated according to their size, heavy molecules move slowly while light molecules move rapidly, and the size of these molecules can be determined in the

gel by a suitable size marker of ladder for this purpose (figure 11.3 a).

Not all heavy molecules can be separated by agarose, since this gel is suitable to separate the nucleic acid molecules (DNA and RNA), as for proteins, they can be separated by polyacrylamide gel electrophoresis (PAGE) after their denaturation with sodium dodecyl sulphate (SDS). The reason behind using this method in separating proteins is the inability to use the electrical charge in proteins separation, where some proteins have positive or negative "net" charge in certain pH and loss this charge during changing pH, other proteins characterized with neutral charge. Therefore, Laemmli at 1970 was abolished the relying on protein charge and relied exclusively on the MW for the separated proteins by denaturing proteins with SDS. This technique earned great popularity due to their great advantages which represents by the high power of resolution that also called SDS-PAGE (figure 11.3 b).

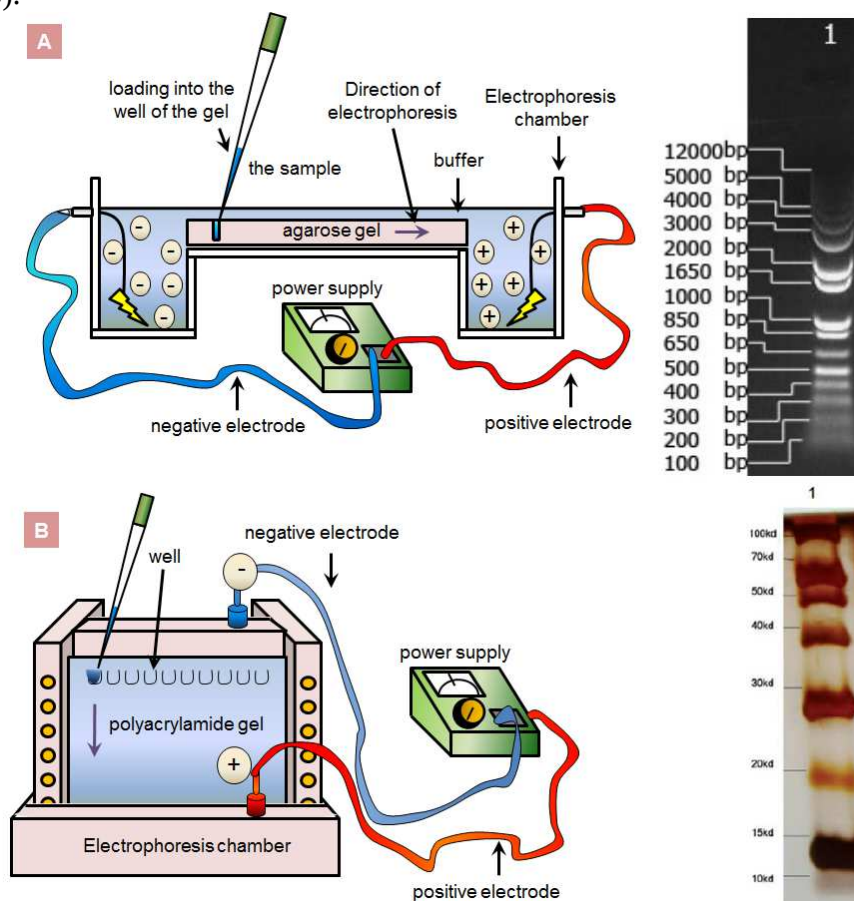
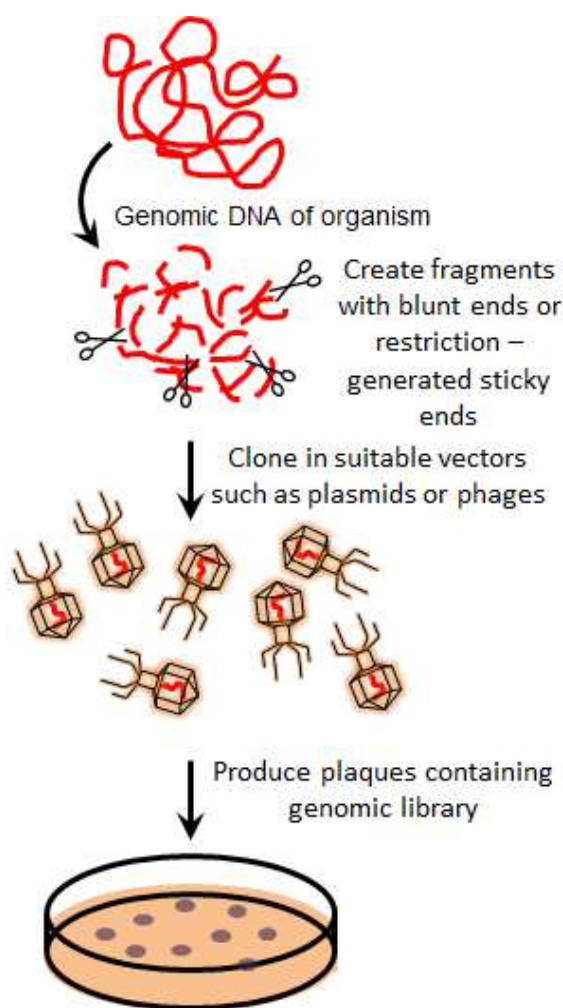


Figure (11.3). Illustrated electrophoresis. (A) Horizontal electrophoresis in agarose gel and DNA bands separated in ethidium bromide stained agarose gel. Lane No. 1 represents the DNA size marker. (B) Vertical electrophoresis in polyacrylamide gel and proteins bands separated in denatured polyacrylamide gel (SDS-PAGE) and silver nitrate stained. Lane No. 1 represents the protein size marker.

The second method (gene library construction); despite the ability of the researcher to isolate the selected DNA fragments as it discussed above, its favorable sometimes to degrade all the genome and to clone all the segments using this vector. Afterword, it's favored to select each desired clone. To make sure of that each genome is represented in these groups of clones, the genomic library, it has to keep more than thousand bacterial strains, in which as long the genome is large as long the need for more

clones is more as well. The sum of clones containing all the DNA segments from only one source is called the DNA library. For instance, when genomic DNA is isolated from human cells, and digested into many segments, hence, these segments called human genome library, which contain all the sequences except the DNA exist in the human genome.



Genomic DNA library; to construct genomic library, the cells are collected, degraded, the DNA is digested into smaller pieces by partial digestion with a restriction enzyme. The enzymatic digestion is a partial due to the DNA is partially subjected to the restriction enzyme for limited period of time, in which "some" sensitive regions of each individual DNA is degraded. Since the degradation of these regions is random, the different DNA molecules are digested in different locations. This of course leads to the production of a group of overlapped DNA molecules (figure 11.4).

After ligating these fragments with a suitable gene cloning vector, such as phages and cosmids, which are transported into bacteria, this process produces a group of bacterial cells that contain overlapped genomic DNA fragments. Thus, a few clones contain the fully desired gene. The genomic library must contain a lot of clones to ensure representation of all DNA sequences available in the genome.

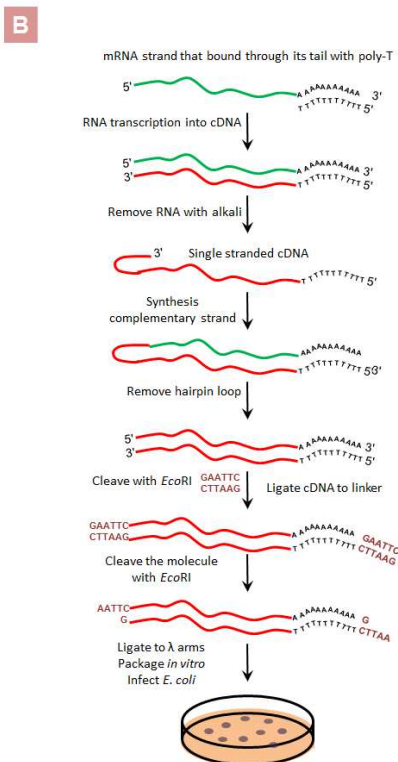
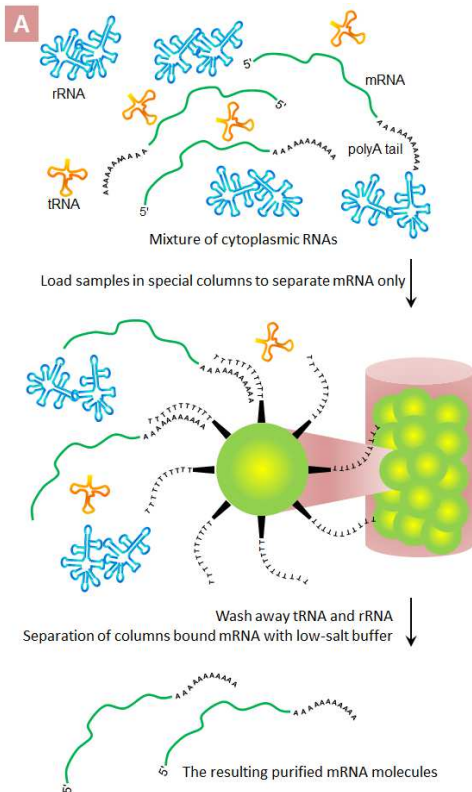
Figure (11.4). Creating a genomic library. First, the genome is fragmented. The fragments are then cloned randomly in vectors. The collection of these vectors is referred to as a genomic library.

Since the genomic libraries were prepared from random fragments of the genome, these libraries might not be efficient in finding the gene, especially in the eukaryotes when many of DNA sequences are noncoding. In contrast, alternative gene libraries that called cDNA libraries characterize in their efficiency in the cloning of the desired gene.

cDNA library; the gene libraries doesn't created only by constructing genomic libraries, instead, there are other means represented by constructing a library consists of

DNA sequences have been transcribed into mRNA. These libraries are called cDNA libraries since the entire DNA in this library is "complementary" to mRNA molecule. Since most of the DNA in eukaryotes consists of repetitive sequences called non-genetic sequences due to their inability to make a copy of themselves through the conversion into mRNA (see chapter two), these sequences cannot be represented in such type of gene libraries. Therefore, the cDNA library has two advantages, the first one; it's a rich library with transcriptionally active genes. So, its represent genes that give high ratio of mature mRNA in that particular location in the body. This made researchers to target tissues that have the desired gene with high transcriptional activity (see below). The second; in cDNA libraries, introns are incapable to interact with the gene sequences. That's mean; such type of libraries doesn't the introns of the genes. This point has been exploited in enabling many eukaryotic genes to express into proteins in bacteria after many researchers were unable to do so since bacteria doesn't have intron removing system (see chapter four). As for the disadvantage of cDNA library, it has the sequences available only in mature mRNA (since they have poly A tail) that is added after removing introns), where such type of libraries clone the same gene and not its flanking region. Hence, introns and many other sequences are disappeared, such as promoter and enhancers, that doesn't also transcribed into RNA. This removes the gene its entire identity. It deserves to note that this library represents only these gene sequences that not expressed in tissues that from which RNA removed. Add to that, the frequency of particular genes in cDNA library depends on the availability of this particular mRNA in the tissue from which it's removed. In contrast, many genes are localized in the same frequency in the genomic DNA library irrespective of their ability in gene expression.

The cDNA library can be constructed in many steps. The first step in cDNA preparation is isolation of total mRNA molecules from the cell or from the tissue of interest. The nature of eukaryotic mRNA molecules facilitates their isolation from the rest molecules. The 3' end of each individual eukaryotic mRNA consists of about 50 to 250 adenine residues that known as poly (A) tail. Because of this tail, it's easy to isolate mRNA molecules from the common rRNA or tRNA molecules present in the cellular extract by using a column containing beads have short strings poly T extensions (TTTTTT....etc) (figure 11.5 a). When the cellular extract is passed through this column, poly (A) tails are base paired with beads linked poly T. by this way, mRNA molecules linked with the column. Since tRNA, rRNA and other molecules don't bind with the column, they can be eliminated. The mRNA molecules are recovered by washing with low salt buffer. Since the cDNA is derived from mRNA, it contains mRNA expressing sequences in the utilized tissue source (cDNA contains continuous expressing sequences of exons without introns). Then, reverse transcriptase is used to synthesize a complementary DNA strand for mRNA molecule. This enzyme can polymerize deoxynucleoside triphosphates (dNTPs) into cDNA strand using the original mRNA molecule as a template for polymerization in the usual direction (5' to 3'). As in the case of other DNA polymerases, reverse transcriptase adds nucleotides only into 3' end of the previously existing primer that base paired with the template. The added poly T termini accomplish this task by their pairing with poly A that present in the 3' end for each mRNA template.



After synthesizing cDNA copies from the isolated mRNA molecules, later molecules are removed by their treatment with alkaline solution (NaOH solution), that hydrolyses RNA and not DNA. Then, the single stranded cDNA is converted into double stranded DNA. Hence, the role of DNA polymerase is emerged since its exploit the fact that reverse transcriptase leaves a bent end resembles hairpin loop when it finish the synthesis of DNA at 3' end.

Figure (11.5). cDNA library; (a) mRNA isolation of the desired gene by using poly T beads, (b) steps of *in vitro* cDNA synthesis using reverse transcriptase.

This end is used as primer by DNA polymerase I to synthesize a DNA complementary to the first strand that synthesized by reverse transcriptase. The double stranded DNA strand is treated by S1 nuclease, which removes the hairpin (single stranded DNA), to generate blunt ended double stranded DNA molecule (figure 11.5 b). To prepare cDNA for cloning purposes, linkers are used for both ends, with one condition; these linkers must have recognition sites for restriction endonucleases. When these resulting cDNA molecules that containing these recognition sequences on the attached linkers at both ends are treated with restriction endonucleases special for these linkers, sticky ended cDNA molecules are generated at either ends. The last step in cDNA construction is ligating sticky ended double stranded cDNA fragments with lambda vectors generated by cleaving with the same restriction endonuclease. Then, the recombinant cloning vector is introduced into *E. coli* cells, in which each cell contain a cDNA derived from a single mRNA. Each single clone is called cDNA clone. In contrast of genomic library, which contains all nuclear DNA sequences of the organism, the cDNA library contains only the sequences that give mRNA. Thus, many tissues in the animal body, which are known to express some different

genes, give different cDNA libraries. The cloned genes libraries can usually be generated by using lambda or cosmid as a vector (see below). The clone containing the desired gene in this library is then screened by radio-labeled or non radio-labeled way by using replica plating on nitrocellulose papers that contain this labeled marker, where the resulting plaques represent the clones presenting in the original plate that has the desired DNA fragment. If the gene to be detected is able to express about itself, the protein product of this gene is screened by a suitable immunological way such as Western blot (see below).

Third Method (direct oligonucleotides synthesis); Although it's not necessary for the reader in any textbook of molecular genetics to recruit organic chemistry, there is a tight relationship between the two specialties in relation to the chemical synthesis of oligonucleotides. Therefore, it's so necessary for the reader to know – even briefly – how the desired polynucleotide is synthesized. It can fairly be said that oligonucleotides are the fuel that drives the engine of molecular biology. It might not be an exaggeration to say that any technique used today in molecular genetics employs these chemically synthesized nucleotides. Now, it's possible to use the oligonucleotides synthesizer machine in the synthesis of the desired sequences from these polynucleotides (figure 11.6). Thus, once its sequence is identified, it's initially possible to synthesize the gene in an organo-chemical mode. The gene sequence can be known through identifying the sequence of its protein product. For example, if the is responsible about the production of a certain polypeptide chain of known sequence, hence, using the genetic code dictionary, the gene sequence can easily be deduced.



Figure (11.6). DNA synthesis machine. It's possible through this machine to synthesize different DNA sequences according to the purpose of research, such as PCR primers, genes ... etc. (this device is from Applied Biosystems – USA).

Nucleic acid synthesis is passed through several historical stages (can be noticed in chapter five), when researchers wanted to interpret the genetic code of nucleic acids until they found a chemical way to synthesize nucleic acids to accomplish this task. The first achievements in the chemical synthesis of nucleic acids were came from Holley and Khorana works, in which tRNA molecules are used because they are very small in size (about 76 nucleotides), where its s difficult at the beginning of nucleic acids synthesis in the

world to synthesize whole genes since they are very long to be chemically synthesized, where the average of what every gene has of nucleic acids is about 1500 bps. In spite of the successes of researchers like Khorana, chemical synthesis of oligonucleotides remained labor intensive and inefficient until 1983, at which the efficiency of this process is enhanced. Hence, the new synthesis process was based upon the use of phosphoramidite monomers and the use of tetrazole catalysis.

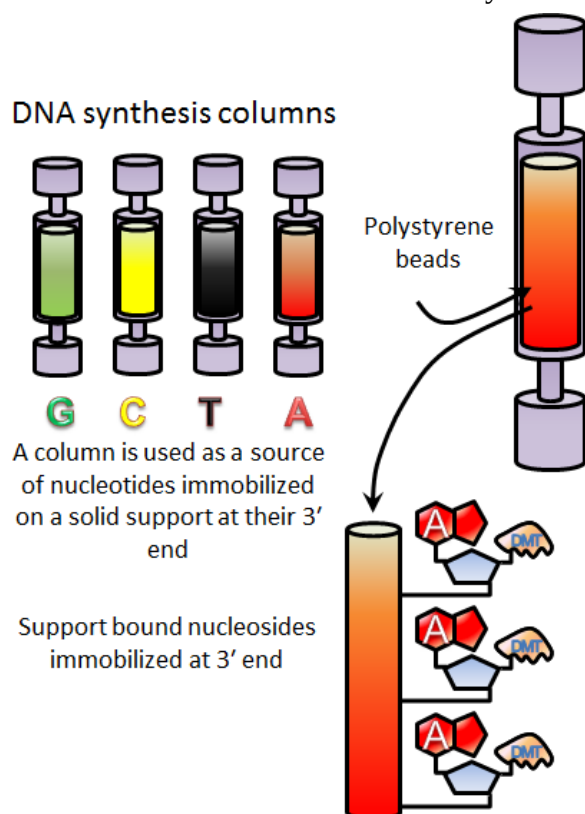
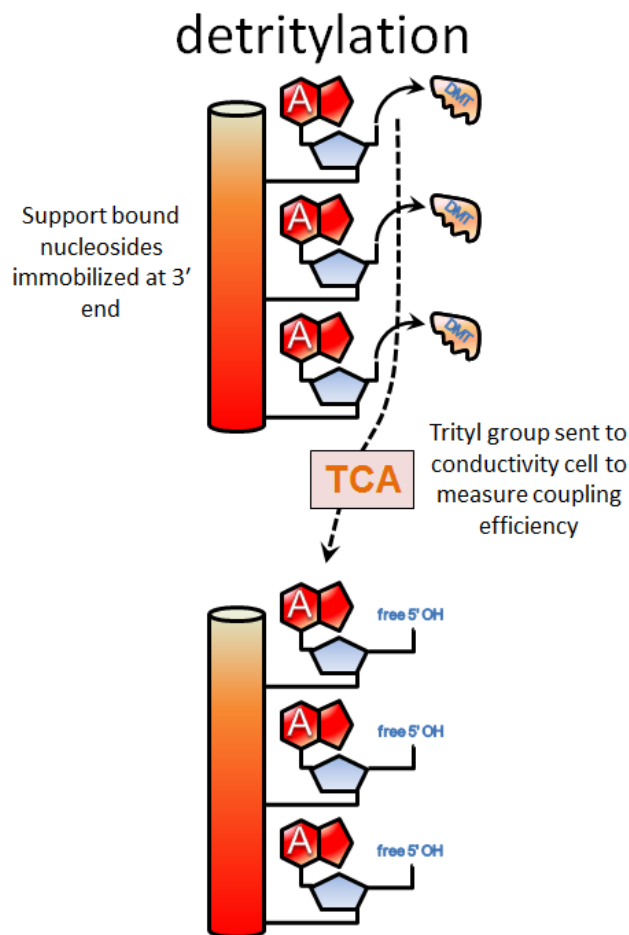


Figure (11.7). Columns used in the chemical synthesis of nucleic acids. Nucleotides are stabilized from their 3' end on a solid support consist of polystyrene that placed on suitable columns for this purpose. Each color refers to one of the nitrogen bases.

A phosphoramidite monomer is a unique synthesis unit that contributed significantly in the development in this respect. First of all, the trityl group is used as a removable protective group. The link to the solid support is made through the 3' carbon and synthesis proceeds 3' to 5' rather than the 5' to 3' synthesis used naturally. The solid support used is a controlled pore glass bead (CPG). This bead has a surface with holes and channels and it is in these that the protected nucleotide is attached.

The phosphoramidite method in oligonucleotides synthesis is the optimum method for most laboratories in obtaining nucleotides sequences because of the efficiency and velocity of nucleotides binding in this way and because of the stability of the starting material on which the nucleotides are added sequentially. Most nucleotides are added depending on the starting material, which is a solid chemical material, in which the excess of severe chemical indicators are removed by filtration. Therefore, there are no purification steps between the successive cycles. The starting material or the supporting material consists of silica or polystyrene beads (figure 11.7). the size of particles and pores is designed with finest precision in such away liquids motion are freely allowed meanwhile they have enough mechanical power to preserve them not to collapse in the subsequent synthesis steps.



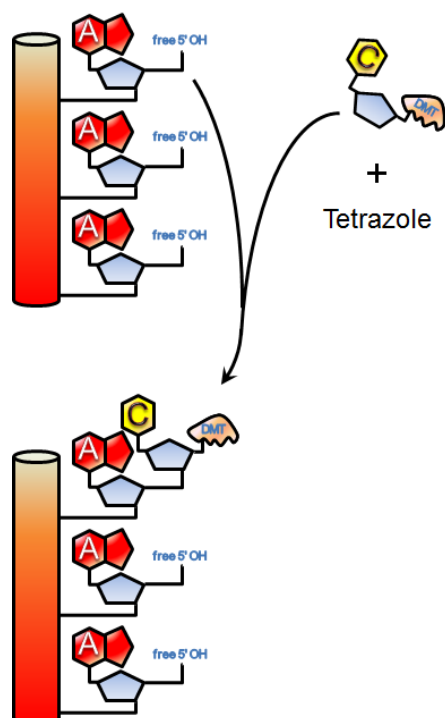
Phosphoramidite synthesis begins with the 3'-most nucleotide and proceeds through a series of cycles composed of four steps that are repeated until the 5'-most nucleotide is attached. These steps are deprotection, coupling, capping, and stabilization, and as follows.

Deprotection step: it also called detritylation step, which is the first step in the chemical synthesis of nucleic acids. In the deprotection step, the trityl group attached to the 5' carbon of the pentose sugar of the recipient nucleotide is removed by trichloroacetic acid (TCA) leaving a reactive hydroxyl group. Since the nucleoside links by its 3' end with the solid support through a linker attached to this end. Therefore, the direction of nucleotide addition is from 5' end instead of 3' end (figure 11.8). This means, the proceeding of DNA synthesis is opposite to the nature (see chapter three).

Figure (11.8). First step in nucleic acids synthesis, in which trichloroacetic acid (TCA) plays a crucial role in this step. TCA removes the acid-labile, dimethoxytrityl (DMT) protecting group on the 5'-hydroxyl of the support bound nucleoside.

Coupling step: at this step, the next phosphoramidite monomer is added with tetrazole. This weak acid adds proton groups for nitrogen atom in phosphoramidite molecule, which makes it exposed to a nucleophilic attack to form a tetrazole phosphoramidite intermediate. This intermediate structure is very reactive in such away it finish this process in a matter of 30 seconds. In which, this structure reacts with the hydroxyl group of the recipient and the 5' to 3' linkage is formed (figure 11.9). The added phosphoramidite molecule is added or protected at its 5' end by adding DMT group as well. It noteworthy that the use of tetrazole increased coupling efficiency to greater than 99% and, with this, opened the way for longer and longer oligonucleotides to be synthesized.

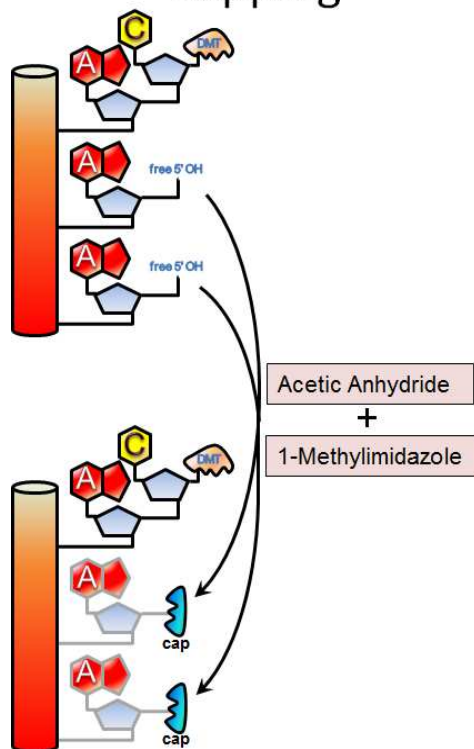
coupling



If this were to remain freely reactive, it would be able to couple in the next round and the result would be a missing base in the synthesis. Thus a creative step must be found to treat this problem.

Figure (11.9). The second step in nucleic acids synthesis, in which tetrazole plays a crucial role in the increasing the efficiency of this step

capping



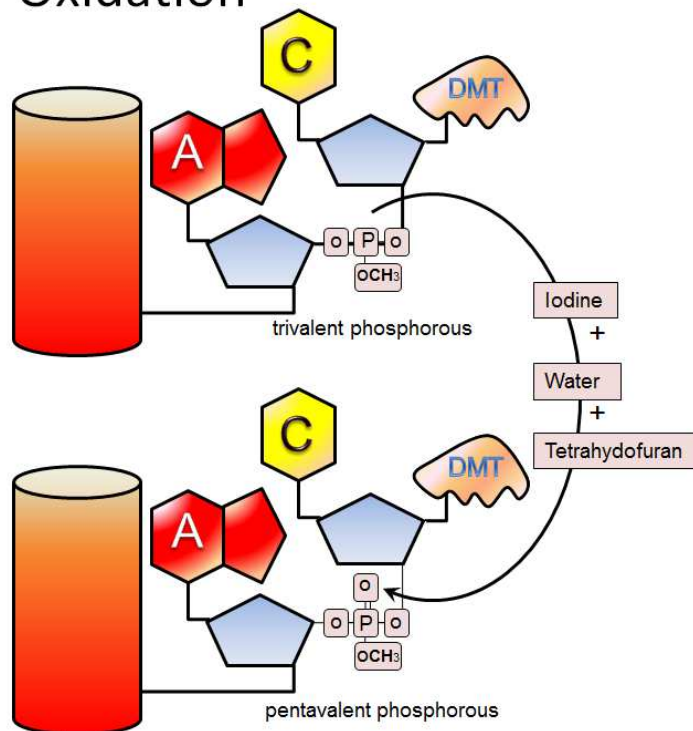
Capping step; This step is designed remove the failure occurring in the coupling reaction in order not to participate furthermore in the synthesis process. Since the non-reacted groups have 5' hydroxyl group, it can be terminated or covered by adding acetyl groups. This is classically accomplished by adding an acetylating reagent composed of acetic anhydride and N-methyl imidazole. This reagent reacts only with free hydroxyl groups to irreversibly cap the oligonucleotides in which coupling failed. These ends are called "failure products". The covering (capping) process is performed by adding acetic anhydride and 1-methylimidazole (figure 11.10).

Since the chains that reacted with phosphoramidites in the previous step is still blocked by DMT group, they cannot be effected by this way. Although the capping step is a step doesn't require DNA synthesis, it is a very desired step since it reduces the length of impurities and consequently it facilitates the identification and purification of the product.

Figure (11.10). the third step in chemical synthesis of nucleic acids, which is designed to overcome the problem of don't reacting of less than 2% of 5' hydroxyl groups with the added phosphoramidite group. This leads to a missing of nitrogen base. These failure sequences are never allowed to grow. Consequently, they must be covered. An equal size of acetic anhydride and 1-methylimidazole is added to act as acetylating reagents of superpower activity. After adding these reagents, the free 5' hydroxyl groups are converted into inactive forms.

Stabilization step: once the capping step is accomplished the last step in the cycle that called stabilization step or oxidation step is started, which stabilizes the phosphate linkage between the growing oligonucleotide chain and the most recently added base. Again, in the classic phosphoramidite synthesis, this step is carried out in the presence of Iodine as a mild oxidant in tetrahydrofuran (THF) and water (figure 11.11).

Oxidation



The Water acts as the oxygen donor and the iodine forms an adduct with the phosphorus linkage. The adduct is decomposed by the water leaving the phosphotriester bond stabilized. Following this final step the cycle is repeated for each nucleotide in the sequence (figure 11.12).

Figure (11.11). the fourth step in the chemical synthesis of nucleic acids. The inter-nucleotide linkage is triphosphate phosphodiester bond is non-stable. Therefore, it must be stabilized by their oxidation. Hence, phosphate acts as oxidative agent with water, in which iodine acts here as an oxygen acceptor.

At the end of the synthesis the oligonucleotide exists as, say, a 25-mer with the 3' end such as PCR primers (see below), are still attached to the substrate and the 5' end protected with a trityl group. In addition, there are protecting groups on G, C, A except T since it needs no protecting group. The completed synthesis is cleaved off the substrate and then detritylated leaving a hydroxyl on both the 3' and 5' ends. At this point the oligonucleotide is deprotected and exists as a functional single stranded DNA molecule. However, deprotection removes the protecting groups but they remain with the oligonucleotide as organic salts that must be removed. The process of removing these contaminants is called desalting.

Oligonucleotide elongation

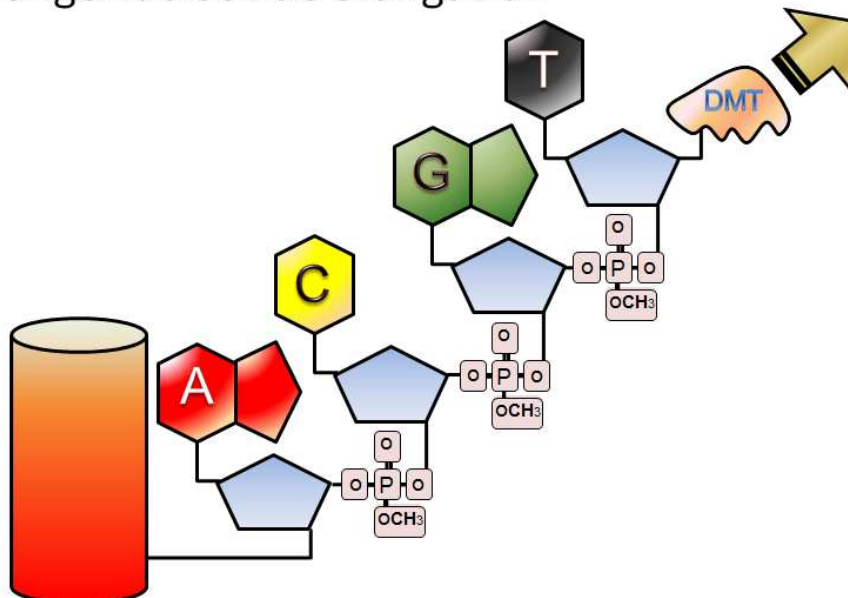


Figure (11.12). Elongation of nucleic acids.

After completing stabilization step, of each cycle, the DNA synthesis continues by removing DMT group and repeats another cycle of base addition. These cycles continue until the full desired number of the desired length of polynucleotides is completed.

In case the researcher desires to obtain longer nucleotides or modified or tagged nucleotides, desalting is not enough in this respect. So, the chemically synthesized polynucleotides are subjected into a further purification either by polyacrylamide gel electrophoresis (PAGE) or by high performance liquid chromatography (HPLC).

There are many applications for polynucleotides as it shown below. These synthesized nucleotides used as primers in PCR, or as probes to detect the gene through hybridizing with it. In the second case, it must tag these probes with a suitable marker. There are several ways to tag these probes must be used probes to undergoes their scheduled tasks in this field. For example, if any researcher desires to identify the location of any gene within the restriction enzymes digested DNA fragments, he must use a specific probe. The probe is nucleic acids of exactly complementary sequences for the target DNA, since this complementation is necessary to anneal with this target by hybridization. These probes are tagged where they can be identified afterword with autoradiography or by other sensitive chemilumnescent techniques.

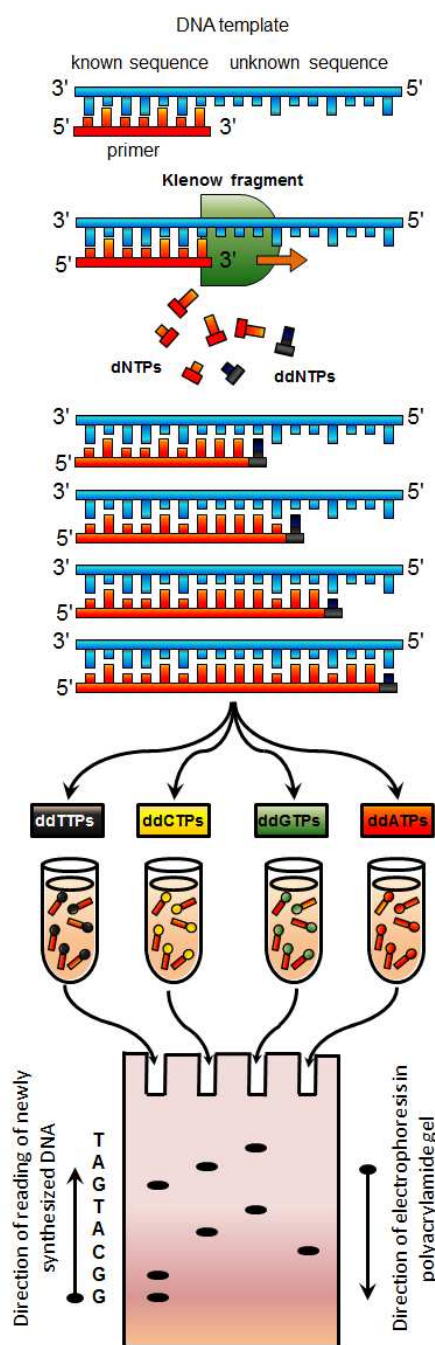
How to Identify the Gene Sequence

After isolating the desired gene by any method, it must know the complete sequence of this gene in case limited information is available about it. As it has been shown above, when amino acids sequence of the gene product is known, the final fragment of the isolated gene must be analyzed.

After all of this, it's so necessary to make sure that the isolated gene has the expected sequence. Moreover, the 5' end sequence must be known for the precise construction of gene, and to insert the gene in a suitable gene cloning vector. To know the



Frederick Sanger



sequence of the gene of the desired DNA fragment, it must be converted into a single stranded DNA. This is done by connecting it with special cloning vectors that released from the host cell in a single stranded form, transferring the desired DNA fragment by the same form as well. These vectors are known as M13 vectors. Nowadays, the dideoxy chain termination method or Sanger and Coulson method is the most widely adopted sequencing procedure. They invent this protocol in the mid 1970s. However, there is another method known as the chemical method or Maxam and Gilbert method, but it's less frequently used. Therefore, we exclude our discussion only on the first method.

The method begins with the synthesis of a complementary strand of the ssDNA using the enzyme, *E. coli* polymerase I Klenow fragment (or more frequently modified T7 DNA polymerase) and a mixture of dNTP (dATP, dTTP, dGTP, dCTP). A DNA primer is also needed for the enzyme to act. This primer is previously designed to bind upstream of the insertion of the gene to be sequenced on the M13 vectors. However, if 2',3'-dideoxyribonucleoside triphosphate (whether ddATP, ddTTP, ddGTP or ddCTP) which lacks the free OH group at the carbon 3' position is present in the reaction mixture, polymerization will terminate if a ddNTP is picked up (instead of dNTP) at any time during the synthesis (figure 11.13).

Figure (11.13). chain termination (Sanger) method in identifying DNA sequence. Note that bands exist at the upper part of the gel represent the 5' end. Thus, bands exist at the lower parts of the gel represent the 3' end.

In DNA synthesis, polymerization of the DNA strands occurs at different speed. Consequently, a heterogeneous mixture of DNA strands of various lengths is produced that is representative of all possible sequences. Depending on which ddNTP is used, the termination nucleotide will be either A, or T, or G, or C. In fact, all four reactions are

carried out separately, generating 4 groups of sequences, one group with all endings in ddATP, one group with all endings in ddTTP, one group with endings in ddGTP, and one group with endings in ddCTP.

DNA segments in each group can be separated by polyacrylamide gel electrophoresis. However, the resolved bands also need to be visualized or detected. A simple way to do this is to incorporate radiolabeled nucleotides, such as [α - ^{32}P] dATP, into the reaction mixture, so that all the fragments are radiolabeled. Now the resolved bands can be detected by X-ray autoradiography.

The bands are read from bottom up in a 5' to 3' direction. Usually the bands at the bottom are too faint and the bands at the top are too compressed to read. There are only a limited number of bands (~200 bases) that can be read with certainty. In order to determine a complete sequence of a gene, the procedure has to be repeated a number of times by moving the primer along the sequence.

In the last decade, manual sequencing described above has largely been replaced by automated sequencing. A major change in the protocol is the use of fluorescent labels instead of isotopes. Four fluorescent dyes are used, with one for each ddNTP. Chains terminated with A are labeled with one fluorescent dye, chains terminated with T are labeled with a second dye, chains terminated with G is labeled with a third dye, and chains terminated with C is labeled with a fourth dye. Using four different fluorescent labels, it is now possible to carrying out the four sequencing reactions as in the manual protocol in a single tube, and to load the one reaction mixture into one lane of the polyacrylamide gel. The fluorescence detector scans the separated bands, discriminates between the four different fluorescent labels, and translates the results into a readable chromatogram of peaks distinguished by four colors: four example; green for A, red for T, black for G, and blue for C (figure 11.14).

Automated sequencing in ordinary lab research can read up to about 600 bp in a single run. However, the sequence reading modern system are designed to generate more than half million nitrogen bases per day using sequencing gel having several capillary tubes instead of the usual slab format. It's noteworthy that the preparation and loading of specimens in the capillary gel can be automated by a "robot" while the results are automatically analyzed by the same way.

However, despite many modifications were performed in the identifying the nucleotide sequences of nucleic acids, they relies on the same principles invented by Sanger. But, in the last years, another method for analyzing DNA sequences is emerged, this method has high throughput but it doesn't depend on Sanger principles in analyzing DNA sequences. This method is called pyrosequencing method. It allows a rapid analysis of the entire sequence of the genome. It's called pyrosequencing due to the release of pyrophosphate in every time a nucleotide is added into the growing DNA chain (see chapter three). This method exploits the release of pyrophosphate as a marker for nucleotide addition into the growing DNA chain. In the nearest future, the popularity of this technique is expected to be increased.

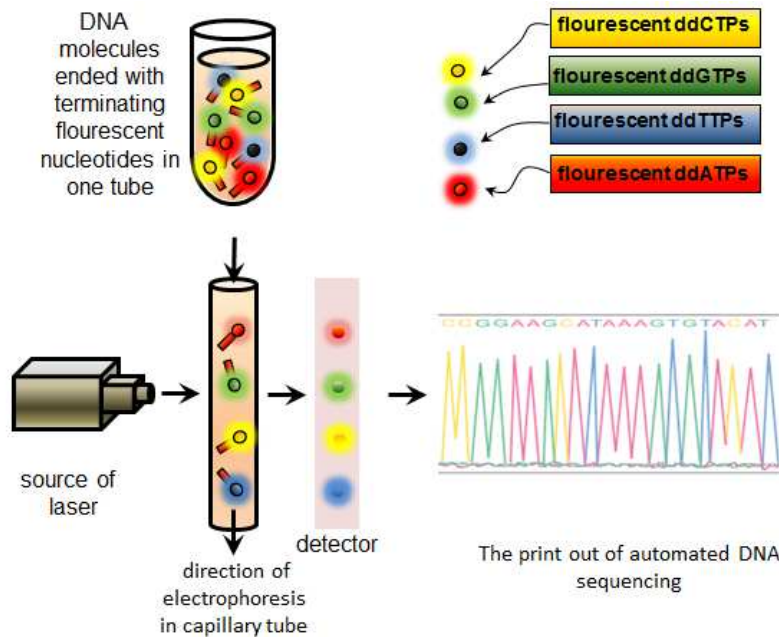
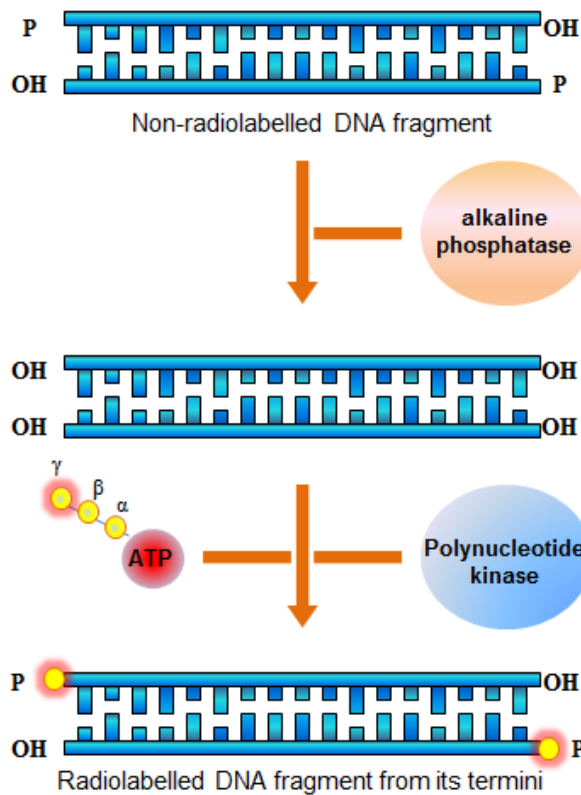


Figure (11.14). Reading the sequence generated by a chain termination for Sanger in a fluorescent dependent way. Each dideoxynucleotide is labeled with a different fluorochrome, so the chain-terminated polynucleotides are distinguished as they pass by the detector.

Labeling of Probes to Detect the Target Gene



Before we get involved in the detection of the gene we have to know how to label the specialized probes for identifying the target gene sequence. There are several ways to tag probes, such as end labeling, in which the enzyme polynucleotide kinase is used to transfer the terminal phosphate group of ATP onto 5'-hydroxyl termini of nucleic acid molecules. If the ATP donor is radioactively labelled, only the termini of each molecule become radioactive (figure 11.15).

Figure (11.15). End labelling DNA using polynucleotide kinase (PNK). (a) DNA is dephosphorylated using phosphatase to generate 5'-OH groups. (b) The terminal phosphate of [γ -³²P]ATP (solid circle) is then transferred to the 5' terminus by PNK. The reaction can also occur as an exchange reaction with 5'-phosphate termini.

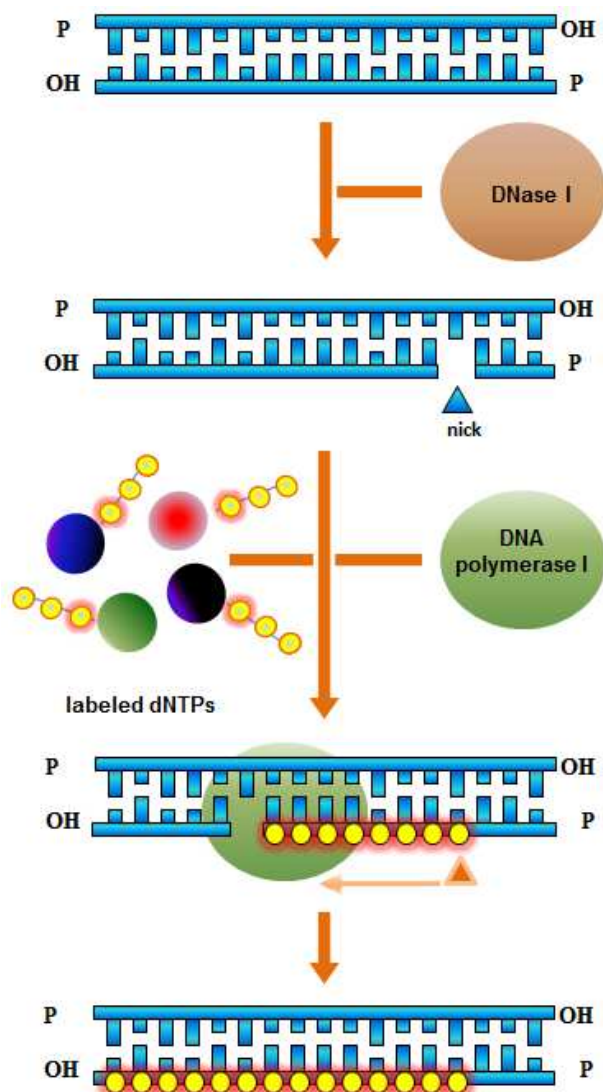
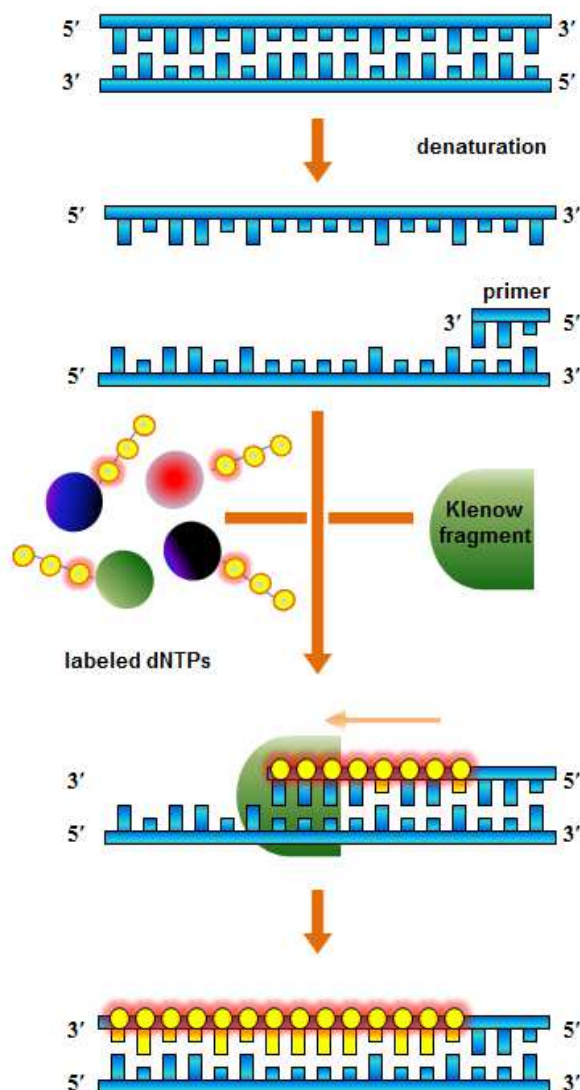


Figure (11.16). Labelling DNA by nick translation. At the beginning, a single-strand nick is introduced into the phosphodiester backbone of a DNA fragment using DNase I. Then, DNA polymerase I then synthesizes a copy of the template strand, degrading the non-template strand with its 5'→3' exonuclease activity. If [α - ^{32} P] dNTP is supplied this will be incorporated into the newly synthesized strand (solid circles).

There is other method of labelling known as labelling by primer extension, which is also called oligolabelling. It refers to a technique that uses random oligonucleotides (usually hexadeoxyribonucleotide molecules; sequences of six deoxynucleotides) to prime synthesis of a DNA strand in most labelling by DNA polymerase. The DNA to be labelled is denatured by heating, and the oligonucleotide primers annealed to the single-stranded DNAs. The Klenow fragment of DNA polymerase can then synthesize a copy of the template, primed from the 3'-hydroxyl group of the oligonucleotide. If a labelled dNTP is incorporated, a labelled DNA of is produced (figure 11.17). This method is also applied by PCR technique (see below).



However, whatever the technique used in labelling, the labelled DNA must be separated from the unlabelled DNA in the reaction mixture. It must be noted that in most labelling reactions, not all labelled dNTPs are incorporated into the target sequence, so, a certain method must be found to separate the labelled products from the unlabelled counterparts. A simple way to do so is summarized by the using of gel filtration in a suitable medium. This process can be done by Pasteur pipette, in which the labelled DNA is reached into the column first, then, it's followed by free nucleotides. The fractions are collected and the power of radiation is monitored.

Figure (11.17). Labelling DNA by primer extension (oligolabelling). First of all, DNA is denatured to give single-stranded molecules. Then, an oligonucleotide primer is then added to give a short double-stranded region with a free 3'-OH group. Then, the Klenow fragment of DNA polymerase I can then synthesize a copy of the template strand from the primer, incorporating [α - ^{32} P]dNTP (solid circles) to produce a labelled molecule.

How to Detect the Gene

After obtaining the probes by their chemical synthesis and their labelling, it's possible to be hybridized somehow in such away the target sequence can be identified. This fact has been employed by Edwin Southern at 1975 and used perfectly to identify the presence and location of the target sequence. Since the just the digestion of nucleic acids with restriction enzymes and their electrophoresis is not enough to identify a small sequence in a big complex of genetic material since the result is no more than a smear on the gel after it's staining with any DNA staining dye such as ethidium bromide. Therefore, Southern has invented a technique called after him, by which the sequence of interest can be detected. Southern blot



Edwin M. Southern

occupies a great position in genetic engineering experiments since this technique doesn't only give an answer to one question "is the gene present?", but it can give an answer to another question "where is the gene?"

The procedure of Southern blot will be dealt with some details due to its great importance in detection of the presence and insertion, and even the insertion of the gene. Moreover, many other detection methods depend on this method. The first step of Southern blot is to digest the DNA containing the target sequence with a restriction enzyme or many restriction enzymes. Then, these digested DNA fragments are subjected to electrophoresis on the agarose gel. This gel is placed on a filter paper that submerged on a reservoir containing the transfer buffer. The agarose gel is mounted on a filter-paper wick which dips into a reservoir containing transfer buffer. The hybridization membrane is sandwiched between the gel and a stack of paper towels, which serves to draw the transfer buffer through the gel by capillary action. The DNA molecules are carried out of the gel by the buffer flow and immobilized on the membrane (figure 11.18-a). Initially, the membrane material used was nitrocellulose. The main drawback with this membrane is its fragile nature. Supported nylon membranes have since been developed which have greater binding capacity for nucleic acids in addition to high tensile strength.

For efficient Southern blotting, gel pretreatment is important. Large DNA fragments (larger than 10 kb) require a longer transfer time than short fragments. To allow uniform transfer of a wide range of DNA fragment sizes, the electrophoresed DNA is exposed to a short depurination treatment for short period followed by alkali since the alkaline addition denatures the fragments prior to transfer, ensuring that they are in the single-stranded state and accessible for probing. Finally, the gel is equilibrated in neutralizing solution prior to blotting. An alternative method uses positively charged nylon membranes, which remove the need for extended gel pretreatment. With them the DNA is transferred in native (non-denatured) form and then alkali-denatured *in situ* on the membrane. After transfer, the nucleic acid needs to be fixed to the membrane and a number of methods are available. Oven baking at 80°C is the recommended method for nitrocellulose membranes and this can also be used with nylon membranes. Due to the flammable nature of nitrocellulose, it is important that it is baked in a vacuum oven. An alternative fixation method utilizes ultraviolet cross-linking. It is based on the formation of cross-links between a small fraction of the thymine residues in the DNA and positively charged amino groups on the surface of nylon membranes.

Following the fixation step, the membrane is placed in a solution of labelled (radioactive or non-radioactive) RNA, single-stranded DNA or oligodeoxynucleotide which is complementary in sequence to the blot transferred DNA band or bands to be detected. Conditions are chosen so that the labelled nucleic acid hybridizes with the DNA on the membrane. Since this labelled nucleic acid is used to detect and locate the complementary sequence, it is called the *probe*. After the hybridization reaction has been carried out, the membrane is washed to remove unbound radioactivity and regions of hybridization are detected autoradiographically by placing the membrane in contact with X-ray film (figure 11.18-b).

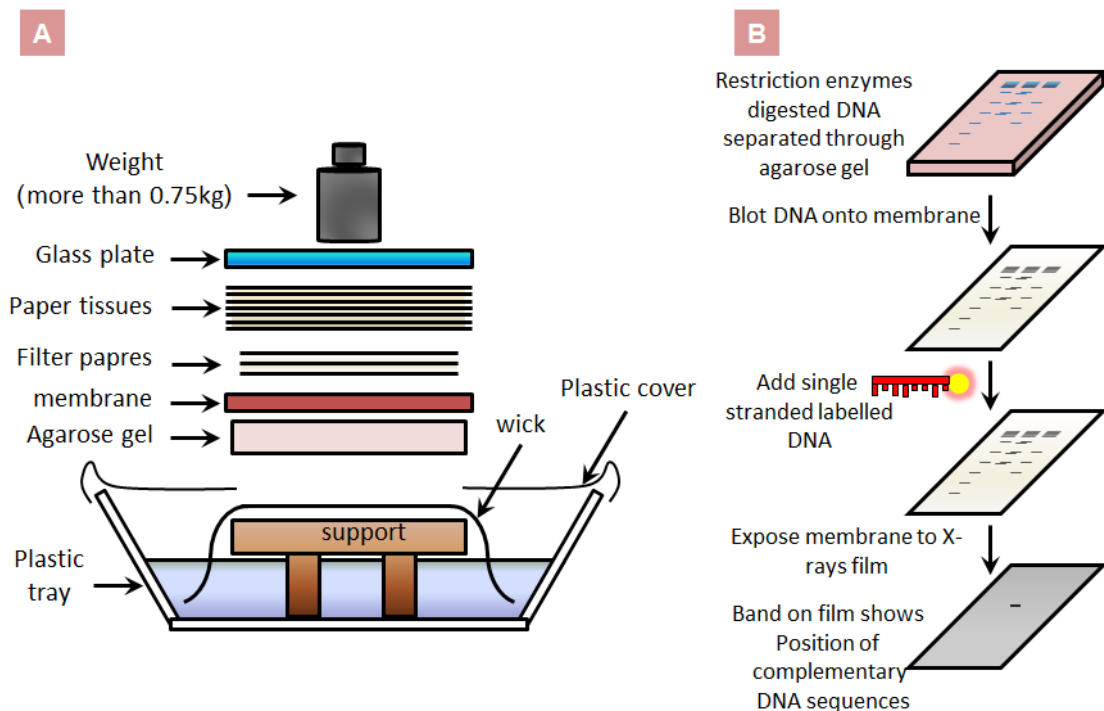
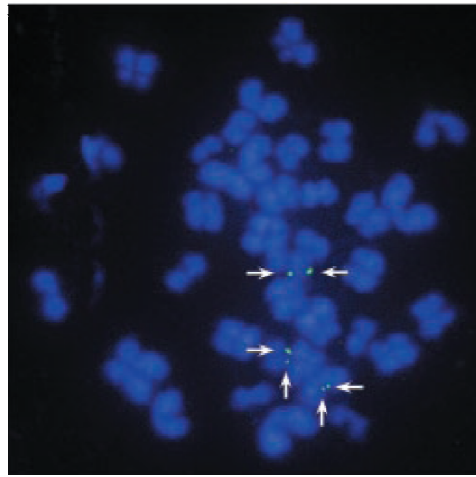


Figure (11.18). Southern blot. (a) Typical illustration for Southern blot, in which the capillary action of nucleic acids is shown. (b) Illustration shows how to find the target sequence in Southern blot, after separating DNA fragments on agarose gel, and their transfer into the membrane by the capillary action, and the conversion of DNA into a single stranded form, the probe binding with the membrane can be detected by exposing the membrane into a source of X-rays. The radioactively labelled probe forms a band on the film in a suitable position for the complementary sequence on the membrane.

The Southern blotting methodology can be extremely sensitive. It can be applied to restriction maps, the maps used in determination of the genes loci by the treatment with several restriction enzymes where the locus of any individual gene can be identified within hundred thousands of genes with different sequences. Southern blot has other applications such as forensic analysis. So, the researchers with this technique are able to use special probe to identify many details of the crime that let to know the criminal (see the appendix of this chapter). Add to that, the same principal steps can be used to transfer the RNA from the gel into similar membranes. This permits in the detection of specialized mRNA sequences of a particular length by the hybridization with specialized labelled probe. This is known as Northern blot. Moreover, there is another technique derived also from Southern blot, which is an immunological technique that doesn't rely on the presence of complementary sequences between nucleic acids. Instead, specialized probes for the target gene protein products known as Western blot (see below).



Despite the multiple advantages of Southern blot, this technique depends on the electrophoresis of the extracted, restriction enzyme digested DNA. Therefore, the locus of the gene on any chromosome cannot be identified. Thus, an alternative technique is used to do this task known as *in situ* hybridization (FISH). By this technique, it's possible to incubate radioactively (or fluorescence) labelled nucleic acid probes with tissue sections or even chromosomes. After removing the excess of probes, the possibility of the probe hybridization with its target sequence can be detected. The great efficiency this technique enjoys has been identified in the determination of the identity of cells that express on particular gene in certain tissue. This is not all the story, since this technique went far away in determining certain genes loci in chromosomes without the need to destroy the conformation of these chromosomes and to isolate their DNA, since to retain the natural conformation of the chromosome means to give the locus of the target gene in the chromosome (11.19).

Figure (11.19). Using of fluorescent in situ hybridization in the determination of the target genes loci by their hybridization with a fluorescence labelled probe specialized with the genes sequences (Lavitrano *et al.*, 2002).

Northern and Western blots are just extensions from Southern blot and for this reason they have acquired their jargon terms "Northern and Western".. These two blots never used in the direct detection of the gene as in the case of Southern; rather, they may be used in detecting the RNA copy of the gene (Northern blot) and the final product of the gene (Western blot). In case of Northern blot, the same essential steps of Southern blot are used, except the target of these steps is to detect RNA not DNA. Here, RNA is separated on agarose gel and transferred onto membrane in the same way of Southern blot that discussed above. The target RNA molecules are detected by the hybridizing with DNA probes that complementary to the target RNA sequences. As for Western blot, it involves in the detection of target protein using antibodies. Proteins are separated through a polyacrylamide gel containing the detergent SDS to keep them in an unfolded (denatured) state. The proteins are transferred from the gel onto a membrane in much the same way as described above for Southern blotting. Particular proteins are then detected using antibodies.

The specific interaction between the antibody and its antigen occurs on the membrane, and the position of the bound antibody is detected. Initially radio-labelled antibodies were used, but these have been largely superseded by antibody 'sandwiches'. The sandwiches work through the binding of one unlabelled antibody (the primary antibody) to the antigen on the membrane.

A second, labelled, antibody (the secondary antibody) is then used to detect the presence of the first antibody. This has several advantages; firstly, the multivalent nature of antibody binding means that a substantial increase in sensitivity is achieved, and secondly, a single secondary antibody can be used to detect a number of different primary antibodies (figure 11. 20).

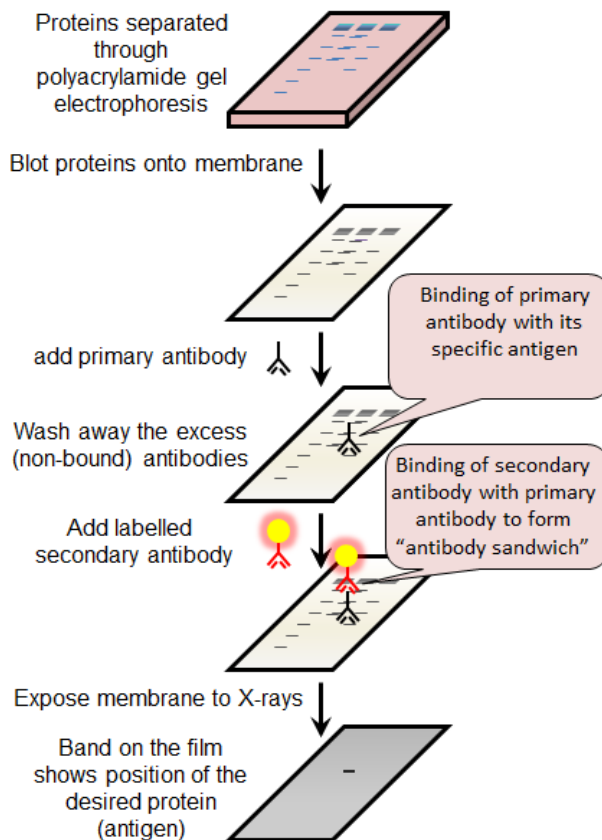


Figure (11.20). Western blotting using antibodies to detect specific proteins. Proteins are separated through a polyacrylamide gel, usually containing SDS to maintain them in a denatured state, prior to blotting onto a nitrocellulose membrane. Non-specific protein binding sites are blocked on the membrane – using solubilized milk powder – before the primary antibody is added. The primary antibody will specifically bind to the antigen to which it was raised. A labelled secondary antibody is then added to detect the location of the primary antibody. The antibody sandwich results in an amplification of the signal. The secondary antibody is often labelled with an enzyme whose activity, in the presence of appropriate substrates, results in either a colour change on the membrane or the emission of light that can be detected using X-ray film

How to Amplify the Gene

After identifying the gene sequence, we want to provide a large amount of it ready for further molecular genetics experiments. Thus, specialized primers can be designed for the gene from its flanking regions to amplify it into hundred thousands or millions of copies in a technique called polymerase chain reaction (PCR). In chapter 10, we discussed how bacteria can be used to amplify a piece of cloned DNA inside it, but this technique is tedious and time consuming compared with PCR technique. Now, we will discuss an *in vitro* amplification technique that can synthesize many copies of DNA. This technique, which costs hours instead of days, uses the same enzyme that *E. coli* use, which is DNA polymerase except of this reaction is taken place in a small test tube.

Polymerase Chain Reaction (PCR); it's a laboratory technique to "amplify" the DNA sequence in a specific form. The PCR is very efficient and sensitive. It makes millions or billions of copies off any specific sequence of DNA, even though this sequence is present within a complex mixture of other different sequences. Because of the power that PCR enjoys with, researchers employed it to amplify the target sequence even this sequence has very small ratio of DNA. For example, a single hair root contains, or a left small blood drop on the crime scene contains enormous amounts of DNA to perform PCR.

The technique of PCR made a revolution in the field of molecular biology. It enabled the researchers to perform experiments that are considered not-rational. Before mid of 1980s, when PCR is developed, the specialists in molecular biology must use monotonous and time consuming in order to detect, clone, and purify DNA sequences of interest. Kary Mullis was awarded Noble Prize at 1993 in chemistry due to his invention for PCR.



The PCR technique is based on the method by which cells replicate their own DNA. During DNA replication, the parental DNA strands are separated from each other. DNA polymerase polymerizes nucleotides to form two new daughter strands for each daughter strand. The parental strands act as templates to synthesize daughter strands. The new strands are assembled together where each nucleotide in the new two strands is determined by the complementary nucleotide in the template strand, A complements T and G complements C. Because of this specificity in base pairing, one of the newly synthesized two daughter strands has the same sequence of one of the parental strands; the other strand has the same sequence of the other parental strand. Thus, replication produces two identical copies for the original copy of DNA duplex (see chapter three).

The sequence that researchers desire to amplify in PCR is called the target sequence, which is exposed to about thirty cycles of replication in a small test tube. During each replication cycle, a number of target sequence molecules are paired since all products and templates for one cycle of replication serve as templates for the next cycle of the same replication. After a number of replication cycles that abbreviated (n), two copies of the target sequence are produced theoretically. After thirty cycles, PCR can produce more than ten billions copy of the target sequence. This is called polymerase chain reaction or PCR, since the DNA polymerase catalyses the chain reaction in replication.

Primer design; in order to replicate DNA, DNA polymerase doesn't only require the template, since it needs the presence of a primer as well. The primer is a sequence of a single stranded DNA, which anneals or binds with the template strand by a specific base pairing. Thus, it's possible for oligonucleotide synthesizer to produce primers for any desired sequence (see the direct gene synthesis).

The PCR primers are short sequences that usually about 15 - 20 mers. The sequences of these short primers are responsible about this high specificity in this technique. The researchers designed primers can anneal on both sequences that located on the either sites of the target sequence. They do it by making the primers complementary to the target sequence, and by making them long enough in order not to anneal to other site (figure 11.21).

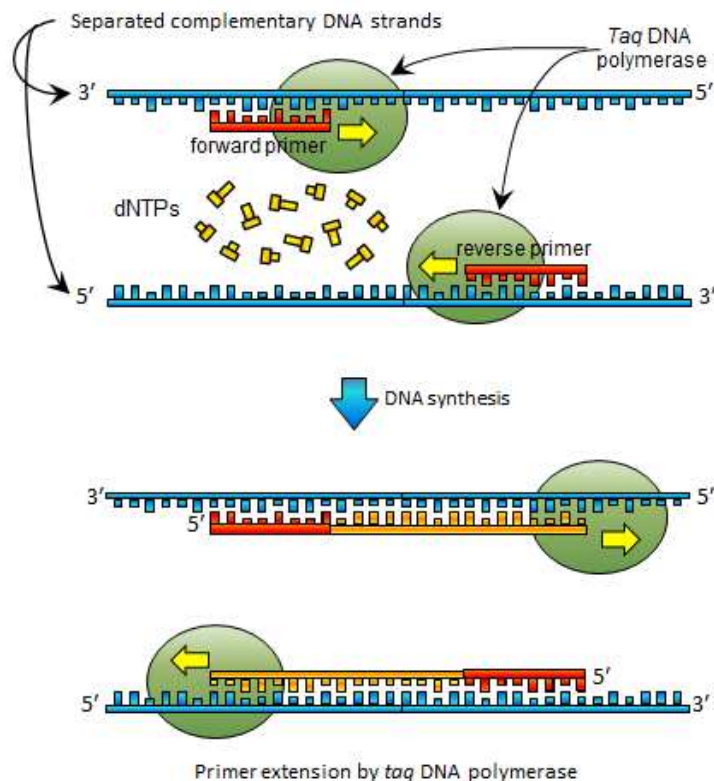


Figure (11.21). The polymerase chain reaction (PCR) that amplifies certain nucleotides *in vitro*. The DNA isolated from cells is heated to separate the complementary strands from each other. Then, these strands are bound with the presence of large quantities of short oligonucleotides that are chemically synthesized to anneal with the target sequence present in the heat separated complementary strands. These nucleotides act as specific primers to synthesize DNA *in vitro* in a way catalyzed by DNA polymerase, which synthesizes DNA between the two positions of forward and reverse primers.

As long the primer length is extended, as long the possibility of being complementary to the target sequence - rather than other sequences - is increased. Since any single position in the DNA sequence can be occupied either by A, G, C, or T, therefore, there is one out of four possibilities of this position will contain A for instance (this is for proximity since it's impossible for nucleotides to distribute evenly or randomly in the DNA). Therefore, the possibility of the presence of any particular sequence in the DNA, in which its length is a particular number of nucleotides (n) in any location of DNA sequence, is from 1 into 4 nucleotides as discussed in chapter ten. The possibility for the 20 mers length primer (the typical length for PCR primers) to exist randomly in any given region of DNA sequence is less than one to trillion (10^{-12}). The length of human genome is about 3×10^9 nucleotides. Consequently, any sequence with 20 nucleotides length is may be impossible to exist for more than once in the human genome. Researchers designed two of primers that can bind with the reverse strands of DNA on either sides of the target sequence. These primers are designed to "shoot" on the right direction, where the DNA fragment that located between them and not out of them is the copied fragment. The primer design to shoot on the right target simply requires building them in such away their 3' ends extend toward the target sequence, whereas their 5' ends extend outward of the target sequence. One of the ends is called 5' end, while the other end is called 3' end. In replication, nucleotides are usually added into the 3' end of the DNA growing strand. Thus, DNA synthesis proceeds from 5' to 3' direction. As its discussed in chapter three, the two complementary DNA strands are antiparallel, this means they walk in two reverse directions. Therefore, this fact has been exploited to generate PCR products with determined lengths (figure 11.21). The typical

PCR reaction consists of several components, mixed with each other in a solution of (25 – 100) μl . This solution must contain DNA template, primers, nucleotides to act like building blocks of the newly synthesized DNA, and DNA polymerase that catalyses the synthesis of DNA.

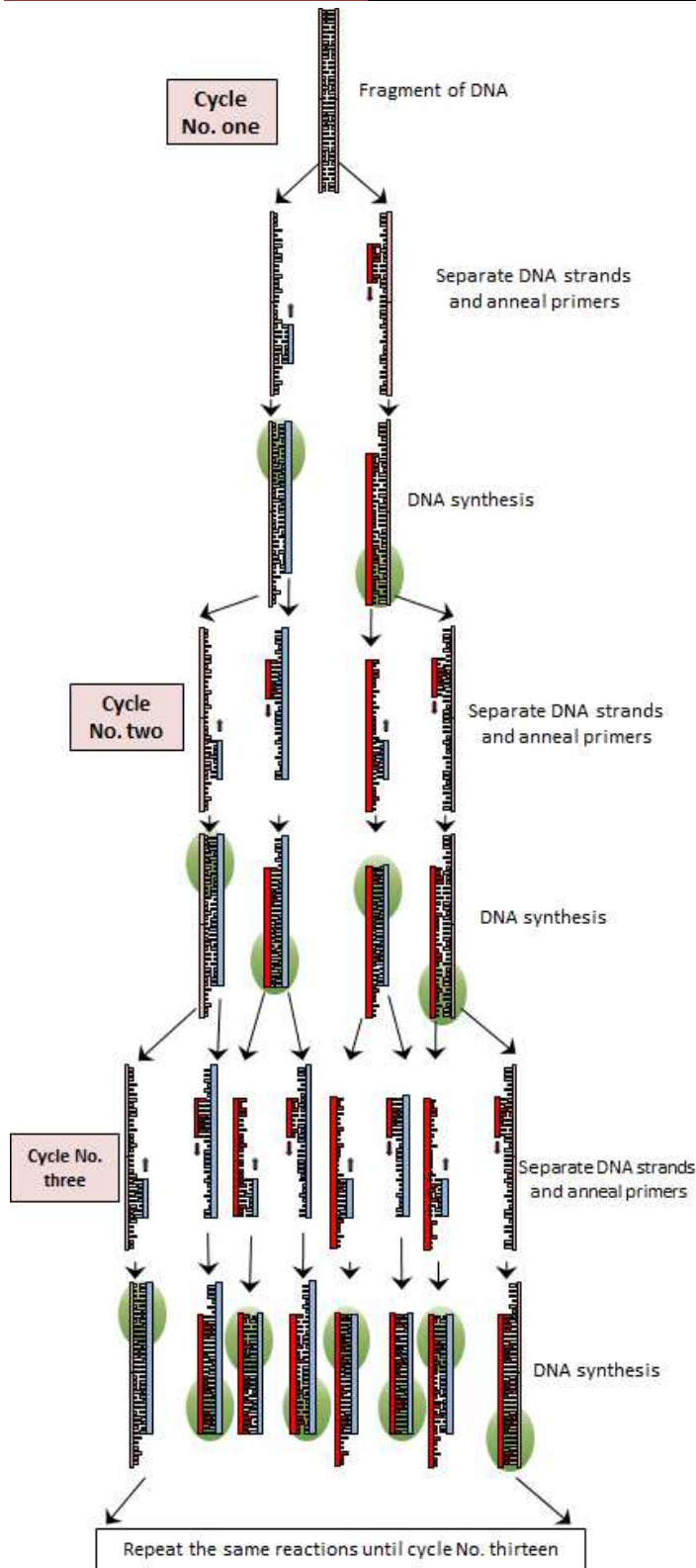
In addition to buffer and salts that usually contain the magnesium ion (Mg^{+2}) which is necessary for the optimum activity for DNA polymerase. It's possible for the template strand to be impure, such as the DNA sample taken from the crime scene. To perform PCR reaction, the solution containing tube is placed into a machine called DNA thermal cycler (figure 11.22). The thermocyclers are essentially programmable heating blocks. They usually contain a thick block of aluminum that composed of pores on which PCR tubes are placed. It's possible for this block to be cooled or heated rapidly into certain temperature degrees for certain periods under the control of computer. Each individual cycle of PCR reaction is regulated by changing the temperature of this aluminum block. Consequently, the temperature of the reaction mixture is changed.



The first step in PCR reaction is heating the mixture into high temperature degree, usually 94°C to 95°C , for about 5 min. the hydrogen bonds that hold the opposite strands with each other in the duplex are broken when exposed into these temperature degrees, and the DNA is separated into single strands. This process is called denaturation.

Figure (11.22). PCR thermocycler for gene amplification (manufactured by Eppendorf – Germany). This machine is lovely designed for the every researcher where it can be controlled by touching directly on its screen.

In second step, the PCR mixture is cooled into a lower temperature, which is between 50°C into 65°C . This allows primers to anneal to specific complementary sequences in the DNA template strand. The temperature degree is chosen precisely to be low enough to permit the primer to anneal, but not cooler than this degree. The lower primer annealing temperature degrees may allow to bind with regions in the template DNA strand that aren't complementary entirely to each others. This can leads to amplification of non-specific sequences. The optimum annealing temperature for many primers can be determined by the chemical formula that depend on the nucleotide composition of the primers, but it's a matter of try and error to find the best annealing temperature. The annealing step last from 15 to 30 seconds, which is an astonishingly short time considering that primer must "scan" DNA sequences looking for the template strand.



In third step (figure 11.23), the reaction is heated again, to reach into 72°C , which is the temperature at which DNA polymerase is more active, most of enzymes are destroyed at this temperature. The first days of PCR discovery, scientists used an *E. coli* derived DNA polymerase, which has higher activity at a temperature higher than the human body which is 37°C . But *E. coli* enzymes were destroyed at high temperature that required for denaturation and annealing reactions. Therefore, DNA polymerase must be added again in each reaction through each step of PCR steps. To solve this problem, scientists purified DNA polymerases from the microorganisms living in the hot springs, or in the deep thermal gaps in the sea.

Figure (11.23). PCR produces an amount of DNA that can pair in each cycle of DNA synthesis and contain types of DNA with unique sizes. Each cycle consists of three steps; denaturation, annealing, and polymerization. After several steps of the reaction, the reaction mixture is predominant with and within a few cycles the predominant DNA is identical to the sequence bracketed by and including the two primers in the original template. In this scheme, three cycles of reaction produce 16 DNA chains, 8 of which (yellow colour) are the same length as and correspond exactly to one or the other strand of the original bracketed sequence. But, after three more cycles, 240 of the 256 DNA chains.

The enzymes of these organisms are more active in the higher temperatures. The most used enzyme in PCR technique is called *Taq* DNA polymerase, which is originally purified from hot springs bacteria that are known as *Thermus aquaticus*. In 72°C, *Taq* DNA polymerase adds the nucleotides into the 3' ends of the annealed primers with a ratio about 2000 nucleotides per minute. Therefore, to amplify a sequence of 1000 nucleotides, the primer extension step must last 30 seconds at 72°C. At the end of this step, each template strand has new complementary strand. This finishes the first cycle of PCR cycles. This cycle can be repeated, at this point, by repeating the denaturation step again. In the next step, it's possible for the two DNA strands to act as templates, as it will happen for the newly synthesized strands. By this way, number of template strands, and thus they base pair again in each successful PCR step. At the end of replication, the reaction tube contains DNA fragments most of them are copies of the target sequence. The original DNA template strand still exist, but it's exist in a negligible amounts compared with the amounts of PCR products (figure 11.23).

The DNA amounts that amplified in the thermocycler are usually detected by gel electrophoresis. The so called enzyme linked immune adsorbent assay can be used with PCR primers that are tagged with a non-radioactive material such as biotin. Since PCR technique is a strong, easy, and non-cost effective, molecular geneticists developed means by which they can broaden the applications of PCR in different researches. Recently, many modifications are emerged in PCR technology, such as;

Hot start PCR; As soon as the PCR reagents have all been mixed together, it is possible for the DNA polymerase to start synthesis. This may happen while the reaction mixture is being heated for the first time, and is at a temperature low enough to allow nonspecific annealing of primer to template, generating a range of non-specific products. This problem would be prevented if DNA synthesis could not take place until the first cycle had reached its maximum temperature. This is the basis of hot-start PCR.

These conditions can be achieved by incorporating the polymerase or the magnesium salt (needed for the polymerase to function) into wax beads. These beads melt at the appropriate temperature, allowing the reaction to start. A different approach is to have the polymerase inactivated at the start by complexing with an antibody. The antibody denatures at high temperature, allowing the polymerase to function.

Reverse Transcription PCR (RT-PCR); researchers devised a technique not far away from the classical known PCR, but it's an extension to it or it's one of its derivatives, and it's called reverse transcription PCR or RT-PCR. Its initially includes the copying of RNA into DNA using reverse transcriptase, than, the classical PCR is used to amplify these complementary sequences that known as cDNA. Since the mRNA component in the cell or in the tissue is represented by the genes that are actively transcribed, so, this technique provides a superior way to analyze gene expression in the tissue of interest. By using this technique, it's often convenient to amplify RNA molecules, to estimate the abundance of a particular mRNA in a sample. This is usually done by having a round of reverse transcription, using a reverse transcriptase enzyme and a single primer, to make a single strand of cDNA prior to the PCR itself. The primer for reverse transcription could be oligo-dT for general cDNA synthesis from polyadenylated tail.

Quantitative Real Time PCR; despite the sensitivity of the classical PCR on the detection on the target sequence, it don't has the ability to gives "quantitative" or real results on the abundance of the target sequence. For example, when the researcher use this classical technique to detect certain gene, using amplification reactions in different PCR tubes, it's not necessary for the density of the resulting PCR product density to indicate the abundance of this gene because we don't have the suitable tool to force the primers to replicate the target gene from the first cycle (figure 11.24). When the intension is to know the amount of the target sequence number of copies a quantitative real time PCR is used.

There are many applications for PCR that have importance in the identification of the starting material. Theoretically, there is a quantitative relationship between the amount of the starting material (the target sequence) and the amount of the PCR product at any cycle of PCR. But, practically, PCR reactions give different amounts of the product, and this make the quantitative identification impossible. The first experiment paved the way of using the quantitative identification of PCR was published at 1993, where ethidium bromide is used for quantitative identification for PCR products during their accumulation.

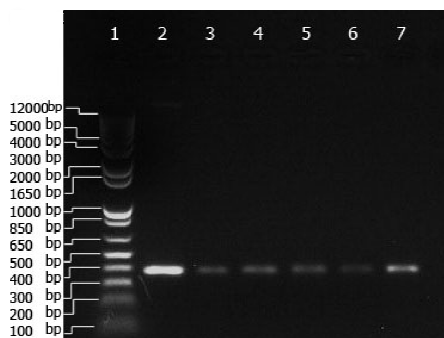
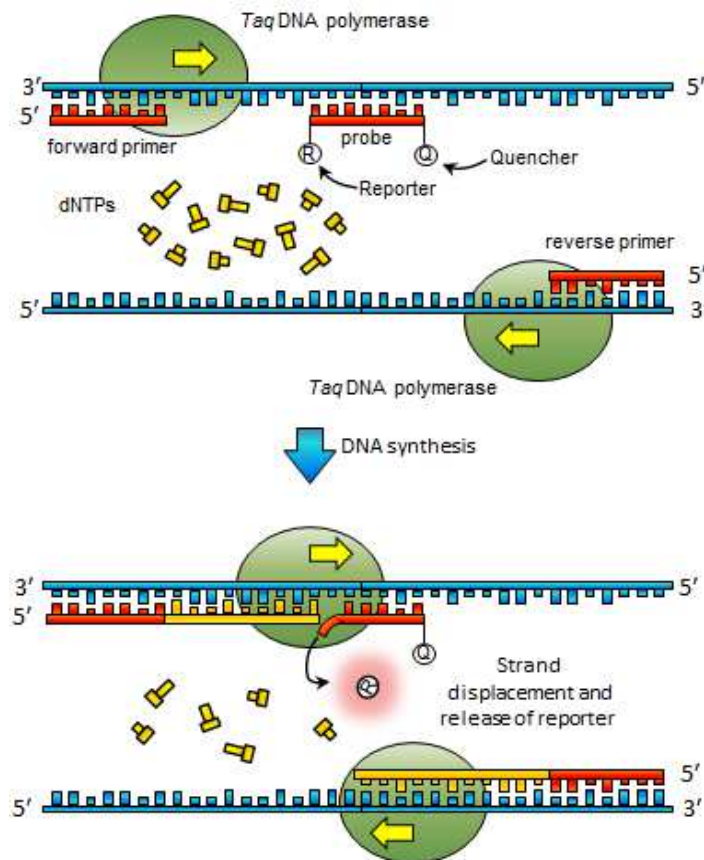


Figure (11.24); the inability of the classical PCR technique to give a real time quantitative description for the number of the target sequence copies. Although the target sequence is present according to the bands shown in lanes 3 to 7, their densities are different. These densities don't necessarily reflect the differences in the number of the copies of target sequence (Al-Shuhaib *et al.*, 2011).

Amplification produces increasing amounts of double stranded DNA, which binds ethidium bromide, resulting in an increase in fluorescence. By plotting the increase in fluorescence versus cycle number it is possible to analyze the PCR kinetics in real time. This is much more satisfactory than analyzing product accumulation after a fixed number of cycles. The principal drawback to the use of ethidium bromide is that both specific and non-specific products generate a signal. This can be overcome by the use of probe-based methods for assaying product accumulation. The probes used are oligonucleotides with a reporter fluorescent dye attached at the 5' end and a quencher dye at the 3' end. While the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter dye. If the target sequence is present, the probe anneals downstream from one of the primer sites. As the primer is extended, the probe is cleaved by the 5' nuclease activity of the *Taq* polymerase (figure 11.25).

This cleavage of the probe separates the reporter dye from the quencher dye, thereby increasing the reporter-dye signal. Cleavage removes the probe from the target strand, allowing primer extension to continue to the end of the template strand. Additional reporter-dye molecules are cleaved from their respective probes with each cycle, effecting an increase in fluorescence intensity proportional to the amount of amplicon produced. Instrumentation has been developed which combines thermal cycling with measurement of fluorescence, thereby enabling the progress of the PCR to be monitored in real time.



This revolutionizes the way one approaches PCR-based quantitation of DNA. Reactions are characterized by the point in time during cycling when amplification of a product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the target, the sooner a significant increase in fluorescence is noted. Quantitation of the amount of target in unknown samples is achieved by preparing a standard curve, using different starting copy numbers of the target sequence.

Figure (11.25). real time quantitative PCR reaction.

Certain method has been developed to allow monitoring of PCR reaction through the increasing of the fluorescence emanating from fluorescent probes. The instruments that have the ability to mix PCR reaction and detect on the fluorescence can be performed in capillary glass tubes. This permits the penetration of light to activate what is called the fluorophore by monitoring of the fluorescence emanation during performing PCR reaction. It's possible for the modern instruments to monitor several dyes in the same tube. As for the fluorescent probes that have increasing fluorescent ability after binding with DNA, they placed within the same PCR reaction mixture. By increasing the amount of the newly synthesized target DNA, the binding of probe with the target DNA is increased, and the fluorescence is increased in turn.

The simplest DNA binding probes have non-specific nature with certain sequence when they bind with DNA. For example, SYBR Green I, that has fluorescence at 520nm. These dye fluorescence once it binds with the double stranded DNA only. This dye tracks the total amount of the double stranded DNA but it doesn't has the ability to recognize between the different sequences. In order to make sure that the correct sequence is under amplification, a sequence specific fluorescent probe is used.

This idea can be exemplified through the probe that is known as TaqMan[®], which relies on fluorescence resonance energy transfer (FRET) of hybridization probes for quantitation. TaqMan probes are oligonucleotides that contain a fluorescent reporter dye, typically attached to the 5' base, and a quenching dye, typically attached to the 3' base. The

probe is designed to hybridize to an internal region of a PCR product. When irradiated, the excited reporter dye transfers energy to the nearby quenching dye molecule rather than fluorescing, resulting in a non-fluorescent substrate. During PCR, when the polymerase replicates a template on which a probe is bound, the 5' - 3' exonuclease activity of the polymerase cleaves the probe. This separates the fluorescent and quenching dyes and FRET no longer occurs. Fluorescence increases in each PCR cycle, proportional to the rate of probe cleavage, and is measured in a modified thermocycler. Real-time PCR is a powerful quantitative tool, but the cost of reagents and equipment is much higher than that of standard PCR reactions.

Chapter Eleven Appendix

How to Identify the Genetic Identity

Every year in court cases all over the world the ability to establish a person's identity is essential for a just decision. Genetics has come to the rescue of the courts and now the following new questions are routinely asked in the courts : (1) Is the drop of blood found at the crime scene from suspect on trial? Who is the child's father ? Until recently, there was no foolproof test. In a criminal case, if there was no identifiable fingerprint left behind at the crime scene, there was no case. Blood tests can determine who is not the parent not who is the parent for that child. A test has now been developed that provides hundred per cent positive identification. The test is called DNA fingerprinting. The test of DNA fingerprinting can show conclusively whether the genetic material in a drop of blood matches that of the suspect, or it can be used to solve paternity case. The technique of DNA finger- printing relies on developments from recombinant DNA technology and allows an examination of each individual's unique genetic blueprint-DNA. The technique was discovered in England by Alec Jeffreys.

It is based on the fact that the DNA of each individual is interrupted by a series of identical DNA sequences called repetitive DNA or tandem repeats. The pattern, length, and number of these repeats are unique for each individual. Jeffreys developed a series of DNA probes, which are short pieces of DNA that seek out any specific sequence they match, and base pair with that sequence. Such molecular probes are used to detect the unique repetitive DNA patterns characteristic of each individual. The procedure of DNA fingerprinting has the following steps: 1. DNA is purified from a small sample of blood, semen, or other DNA-bearing cells, and digested into smaller fragments with restriction endonucleases. 2. The fragments are separated by agarose gel electrophoresis. 3. The separated fragments are transferred to a nylon membrane by the technique of Southern blotting. 4. The DNA probes labelled with radioactive material are added to a solution containing the nylon membrane. 5. Wherever the probes fit a band containing repetitive DNA sequences, they attach. 6. The X-ray film is pressed against the nylon filter and exposed at bands carrying the radioactive probes attached to the fragments. The patterns of bands obtained on the film are 100 per cent unique for each person, except for identical twins who would have the same pattern.

The forensic application of the DNA fingerprinting technique involves a comparison between the DNA fingerprints obtained from cells at a crime scene with a DNA fingerprint from cells provided by the suspect. If the DNA pattern matches exactly, certain identification is made (figure 11.26). For paternity determination, DNA fingerprints of the mother, child and alleged father are compared. In this case, one-half of the bands in the child come from the mother and the other half from the father. All the paternal bands in child's DNA fingerprint must match with the alleged father for positive paternity identification



Alec Jeffreys

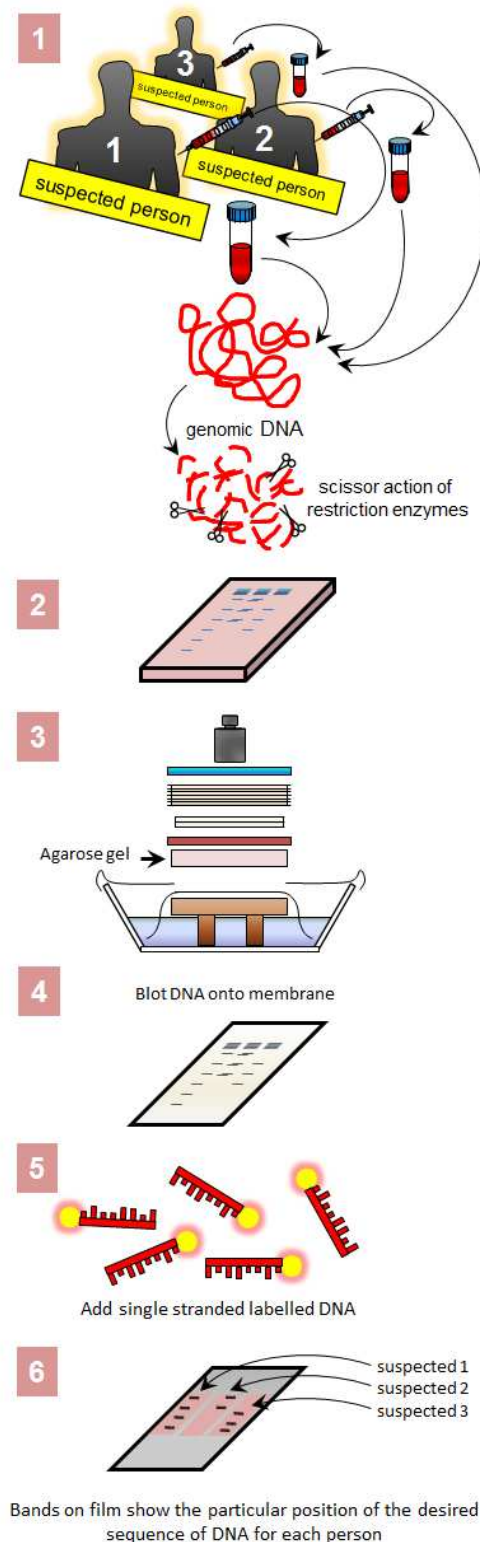


Figure (11.26). Classical steps of DNA fingerprinting protocol.

Chapter Eleven Revision Questions

Q1- Justify the following:

1. Despite the accuracy of density gradient centrifugation in purifying nucleic acids, its replaced other methods?
2. Despite the efficiency of obtaining the gene by cDNA library, this method losses the entire gene ID?
3. Its cannot construct a cDNA library for the gene that cannot express about itself at all?
4. The DNA synthesis machines synthesizes DNA in an reversal direction to the nature (5' to 3')?
5. Alkaline lysis method is usually used in the extraction of plasmid DNA?
6. Sanger and Coulson method in the identifying the DNA sequence is called chain termination method?
7. There is no need to separate DNA samples that intended to be sequenced in four lanes in the gel when flourescent markers are being used?
8. It's favored to use plastic membranes on nitrocellulose membranes in the hybridization experiments of nucleic acids as in Southern blot?
9. In Western blot, its favored to use the immunological ways over radiation ways in the labelling of antibodies?
10. In PCR reaction, when we want to amplify 1000 nucleotides, the primer extension step must be continued for 30 sec at 72°C?

Q2- Choose the correct answer:

1. The much more flourescent DNA in density gradient centrifugation is
a. linear form b. covalently closed circular form c. open circle form d. none
2. The way is the simplest way in the extraction and purification of DNA.
a. phenol based methods b. density gradient centrifugation c. spin column d. both a and b
3. When an experiment is performed to identify the purity of certain DNA sample, it was found that its purity is only 1.05, this means this samples is
a. pure b. contaminated with RNA c. contaminated with protein d. both b and c
4. The wavelength 260 nm is used in the identification of the concentration of in the sample.
a. proteins b. RNA and DNA c. DNA only d. RNA only
5. The technique is one of the most common technique in the identification of the size of DNA molecules.
a. HPLC b. FISH c. SDS-PAGE d. agarose gel electrophoresis
6. It's possible to isolate mRNA molecules in eukaryotes and differentiate them from other molecules in the tissues since they have
a. 5' cap b. 3' tail c. intervening sequences d. codons
7. The enzyme generates blunt ends from the sticky ends.
a. HindIII b. primase c. EcoRI d. SI nuclease

8. The main building blocks of *in vitro* DNA synthesis using DNA synthesis instruments are
a. nucleotides b. nucleosides monophosphates c. nucleoside diphosphates d. phosphoramidites
9. When a particular oligonucleotide is synthesized in DNA synthesis instrument, a TCA compound performs the step.
a. deprotection b. coupling c. capping d. stabilization
10. After synthesis of oligonucleotides in the specialized devices, the technician purifies them by
a. gel filtration b. agarose gel electrophoresis c. HPLC d. *in vitro* packaging
11. The labelling method that is known as can be performed in PCR.
a. 5' end labeling method b. 3' end labeling method c. nick translation method d. primer extension method
12. After completing the labelling, the labelled DNA is separated from the non-labelled DNA by
a. agarose gel electrophoresis b. gel filtration c. SDS-PAGE d. UV-visible spectrophotometer
13. After separating DNA samples in the sequencing gel, the technician read the bands G, C, A, T from top to bottom sequentially, this gives the sequence
a. 5'-GATC-3' b. 3'-GATC-5' c. 5'-CTAG-3' d. 3'-CTAG-5'
14. If the temperature at which the primer anneals with the DNA template is less than the optimum temperature, this may leads to generation of PCR products.
a. false positive results b. false negative results c. positive results d. negative results
15. It's impossible for technique to amplify introns entirely.
a. conventional PCR b. real time PCR c. jumping PCR d. RT-PCR

Q3- Define the following:

Genomic library, probe, nick translation, pyrosequencing, jumping PCR, quencher dye, DNA fingerprinting

Q4- Connect terms exist in right with its suitable counterparts in left

PCR	To amplify the gene
RT-PCR	To detect the location of the gene in the genome
FISH	To detect the location of the gene in the chromosome
Real time PCR	To detect the ability of the gene to give RNA copy
Southern blotting	To detect the ability of the gene to give protein
Western blotting	To detect to whom this gene belong

Further Readings

- Alberts B.**, Jonson A., Lewis J., Raff M., Roberts K., Walter P. Molecular biology of the cell. Fifth edition, Garland Science/USA. 2008.
- Applied biosystems** technical bulletin, 2012.
- Baker**, Catherine. Your genes, your choice: Exploring the Issues Raised by Genetic Research.
- Behleke M.A.** and Devor E.J. Chemical Synthesis of Oligonucleotides. Integrated DNA Technologies (IDT) 2005.
- Birnboim HC**, Doly J. (1982) Nucleic Acids Research 7, 1513-1523.
- Brown T. A.** Gene cloning and DNA analysis. Sixth edition, Wiley-Blackwell, 2010.
- Carey, L.**, and Mitnik, L. 2002. Trends in DNA forensic analysis. *Electrophoresis* 23, 1386-1397.
- Debenham, P. G.** 1992. Probing identity: The changing face of DNA fingerprinting. *Trends Biotechnol.* 10,96-102.
- Eppendorf** catalogue, 2012 – 2013.
- Fitzgerald-Hayes M.** and Reichsman F. DNA and Biotechnology. Third edition, Elsevier, 2010.
- Gasser, C. S.**, and R. T. Fraley. 1992. Transgenic crops. *Scientific American* 266(6):62–69.
- Higuchi R**, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (NY)*. 1993;11(9):1026-30.
- Howe**, Christopher. Gene cloning and manipulation. Second edition, Cambridge University Press, 2007.
- Jeffreys, A. J.**, Turner, M., and Debenham, P. 1991. The efficiency of multi-locus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *Am. J. Hum. Genet* 48, 824-840.
- Lavitrano M**, Bacci ML, Forni M, Lazzereschi D, Di Stefano C, Fioretti D, Giancotti P, Marfe' G, Pucci L, Renzi L, Wang H, Stoppacciaro A, Stassi G, Sargiacomo M, Sinibaldi P, Turchi V, Giovannoni R, Della Casa G, Seren E, Rossi G. 2002. Efficient production by sperm mediated gene transfer of hDAF transgenic pigs for xenotransplantation. *Proc Nat Acad Sci USA* 99:14230–14235.
- Lodge J.**, Lund P., and Minchin S. Gene cloning; principles and applications. Taylor and Francis group, 2007.
- Mullis, K. B.** 1990. The unusual origin of the polymerase chain reaction. *Scientific American* 262(4):56–65.
- Nowak, R.** 1994. Forensic DNA goes to court with O. J. *Science* 265:1352–1354.
- Nicholl D.** S.An Introduction to Genetic Engineering. Third edition, Cambridge University Press, 2008.
- Prescott L.M.** Microbiology. Fifth edition, The McGraw–Hill Companies 2002.
- Primrose S.B**, Twyman R.M., Old R.W. Principles of gene manipulation. Blackwell science. 2001.
- Reece R.J.** Analysis of genes and genomes. John Wiley & Sons, 2004.
- Tamarin** .Principles of genetics. Seventh edition. The McGraw–Hill. 2001.
- Turner P.**, McLennan A., Bates A., White M. Instant notes in molecular biology. Third edition, Taylor & Francis Group/UK. 2005.
- U.S. Congress**, Office of Technology Assessment, *Genetic Witness: Forensic Uses of DNA Tests, OTA-BA-438* (Washington, DC: U.S. Government Printing Office, July 1990).
- Venter, J. C**, Smith, H. O., and Hood, L. 1996.A new strategy for genome sequencing. *Nature* 381, 364-366.
- Wong D.W.S.** The ABCs of gene cloning. Second edition, Springer, 2006.

Glossary

A **site** of the ribosome is the site that an aminoacyl-tRNA enters to base pair with the codon.

Acceptor arm of tRNA is a short duplex that terminates in the CCA sequence to which an amino acid is linked.

An **activator** is a protein that stimulates the expression of a gene, typically by acting at a promoter to stimulate RNA polymerase. In eukaryotes, the sequence to which it binds in the promoter is called a response element.

Active site is the restricted part of an enzyme to which a substrate binds.

Adenylate cyclase is an enzyme that uses ATP as a substrate to generate cyclic AMP, in which 5' and 3' positions of the sugar ring are connected via a phosphate group.

An **allele** is one of several alternative forms of a gene occupying a given locus on a chromosome.

Allosteric regulation describes the ability of a protein to change its conformation (and therefore activity) at one site as the result of binding a small molecule to a second site located elsewhere on the protein.

Alternative splicing describes the production of different RNA products from a single product by changes in the usage of splicing junctions.

The **Alu family** is a set of dispersed, related sequences, each ~300 bp long, in the human genome. The individual members have *Alu* cleavage sites at each end (hence the name).

Amanitin (more fully α -amanitin) is a bicyclic octapeptide derived from the poisonous mushroom *Amanita phalloides*; it inhibits transcription by certain eukaryotic RNA polymerases, especially RNA polymerase II.

The **amber** codon is the triplet UAG, one of the three termination codons that end protein synthesis.

Aminoacyl-tRNA is a tRNA linked to an amino acid. The COOH group of the amino acid is linked to the 3'- or 2'-OH group of the terminal base of the tRNA.

Amplification refers to the production of additional copies of a chromosomal sequence, found as intrachromosomal or extrachromosomal DNA.

Annealing of DNA describes the renaturation of a duplex structure from single strands that were obtained by denaturing duplex DNA.

Antibody is a protein (immunoglobulin) produced by B lymphocyte cells that recognizes a particular 'foreign antigen', and thus triggers the immune response.

Anticodon is a trinucleotide sequence in tRNA which is complementary to the codon in mRNA and enables the tRNA to place the appropriate amino acid in response to the codon.

Anticodon arm of tRNA is a stem loop structure that exposes the anticodon triplet at one end.

Antigen is any foreign substance whose entry into an organism provokes an immune response by stimulating the synthesis of an antibody (an immunoglobulin protein that can bind to the antigen).

Antiparallel strands of the double helix are organized in opposite orientation, so that the 5' end of one strand is aligned with the 3' end of the other strand.

Antisense strand (template strand) of DNA is complementary to the sense strand, and is the one that acts as the template for synthesis of mRNA.

att sites are the loci on a phage and the bacterial chromosome at which recombination integrates the phage into, or excises it from, the bacterial chromosome.

Attenuation describes the regulation of bacterial operons by controlling termination of transcription at a site located before the first structural gene.

Attenuator is a terminator sequence at which attenuation occurs.

B-form DNA is a right-handed double helix with 10 base pairs per complete turn (360°) of the helix. This is the form found under physiological conditions whose structure was proposed by Crick and Watson.

Bacterial artificial chromosome (BAC) is a synthetic DNA molecule that contains the sequences needed for replication and segregation in bacteria. This is used in genomic cloning to amplify sequences typically 100-200 kb long. They are usually derived from the naturally-occurring F factor episome.

Base pairing describes the specific (complementary) interactions of adenine with thymine or of guanine with cytosine in a DNA double helix (thymine is replaced by uracil in double helical RNA).

Bidirectional replication describes a system in which an origin generates two replication forks that proceed away from the origin in opposite directions.

Branch migration describes the ability of a DNA strand partially paired with its complement in a duplex to extend its pairing by displacing the resident strand with which it is homologous.

C-value paradox describes the lack of relationship between the DNA content (C-value) of an organism and its coding potential. A **CAAT box** is part of a conserved sequence located upstream of the start points of eukaryotic transcription units; it is recognized by a large group of transcription factors.

Cap is the structure at the 5' end of eukaryotic mRNA, introduced after transcription by linking the terminal phosphate of 5' GTP to the terminal base of the mRNA. The added G (and sometimes some other bases) are methylated, giving a structure of the

form 7MeG5'ppp5'Np . . .

Carcinogen is a chemical that increases the frequency with which cells are converted to a cancerous condition.

cDNA is a single-stranded DNA complementary to an RNA, synthesized from it by reverse transcription *in vitro*.

Cell cycle is the set of stages through which a cell progresses from one division to the next.

Central dogma describes the basic nature of genetic information: sequences of nucleic acid can be perpetuated and interconverted by replication, transcription, and reverse transcription, but translation from nucleic acid to protein is unidirectional, because nucleic acid sequences cannot be retrieved from protein sequences.

Centriole is a small hollow cylinder consisting of microtubules. It occurs in the centrosome (a type of microtubule organizing center) and is thought to play a role in organizing the microtubules.

Centromere is a constricted region of a chromosome that includes the site of attachment (the kinetochore) to the mitotic or meiotic spindle.

Centrosomes are the regions from which microtubules are organized at the poles of a mitotic cell. In animal cells, each centrosome contains a pair of centrioles surrounded by a dense amorphous region to which the microtubules attach.

Chaperones are a class of proteins which bind to incompletely folded or assembled proteins in order to assist their folding or prevent them from aggregating.

Chloroplast DNA (ctDNA) is an independent genome (usually circular) found in a plant chloroplast.

Chromatids are the copies of a chromosome produced by replication. The name is usually used to describe the copies in the period before they separate at the subsequent cell division.

Chromatin describes the state of nuclear DNA and its associated proteins during the interphase (between mitoses) of the eukaryotic cell cycle.

Chromosome is a discrete unit of the genome carrying many genes. Each chromosome consists of a very long molecule of duplex DNA and an approximately equal mass of proteins. It is *cis* configuration describes two sites on the same molecule of DNA.

The **cisternae** of the Golgi apparatus are the successive stacks, each bounded by a membrane, that make up individual compartments.

Cistron is the genetic unit defined by the *cis/trans* test; equivalent to gene.

Cloverleaf describes the structure of tRNA drawn in two dimensions, forming four distinct arm-loops.

Coactivators are factors required for transcription that do not bind DNA but are required for (DNA-binding) activators to interact with the basal transcription factors.

Coding strand (sense strand) of DNA has the same sequence as the mRNA and is related to the protein sequence that it represents by the genetic code.

Codon is a triplet of nucleotides that represents an amino acid or a termination signal.

Cofactor is a small inorganic component (often a metal ion) that is required for the proper structure or function of an enzyme.

Cognate tRNAs (is an accepting tRNAs) are those recognized by a particular aminoacyl-tRNA synthetase. They all are charged with the same amino acid.

Cointegrate structure is produced by fusion of two replicons, one originally possessing a transposon, the other lacking it; the cointegrate has copies of the transposon present at both junctions of the replicons, oriented as direct repeats.

Conjugation is a process in which two cells come in contact and exchange genetic material. In bacteria, DNA is transferred from a donor to a recipient cell. In protozoa, DNA passes from each cell to the other.

Consensus sequence is an idealized sequence in which each position represents the base most often found when many actual sequences are compared.

Conservative transposition refers to the movement of large elements, originally classified as transposons, but now considered to be episomes. The mechanism of movement resembles that of phage excision and integration.

Constitutive heterochromatin describes the inert state of permanently nonexpressed sequences, usually satellite DNA.

Copy number is the number of copies of a plasmid that is maintained in a bacterium (relative to the number of copies of the origin of the bacterial chromosome).

Core histone is one of the four types (H2A, H2B, H3, and H4) found in the core particle derived from the nucleosome (this excludes histone H1).

CpG island is a stretch of 1-2 kb in a mammalian genome that is rich in unmethylated CpG doublets.

CRP activator (CAP activator) is a positive regulator protein activated by cyclic AMP. It is needed for RNA polymerase to initiate transcription of certain (catabolite-sensitive) operons of *E. coli*.

D arm of tRNA has a high content of the base dihydrouridine.

D loop is a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region. The same term is used also to describe the displacement of a region of one strand of duplex DNA by a complementary single-stranded invader.

Daughter strand or duplex of DNA refers to the newly synthesized DNA. The two cells that result from a cell division are referred to as

De novo methylase adds a methyl group to an unmethylated target sequence on DNA.

Deletions are generated by removal of a sequence of DNA, the regions on either side being joined together.

Demethylase is a casual name for an enzyme that removes a methyl group, typically from DNA, RNA, or protein.

Denaturation of protein describes its conversion from the physiological conformation to some other (inactive) conformation.

Density gradient is used to separate macromolecules on the basis of differences in their density. It is prepared from a heavy soluble compound such as CsCl.

Deoxyribonuclease (DNAase) is an enzyme that specifically digests DNA. It may cut only one strand or may cut both strands.

Deoxyribonucleic acid (DNA) is a nucleic acid molecule consisting of long chains of polymerized (deoxyribo)nucleotides. In double-stranded DNA the two strands are held together by hydrogen bonds between complementary nucleotide base pairs.

DNA fingerprinting analyzes the differences between individuals of the fragments generated by using restriction enzymes to cleave regions that contain short repeated sequences. Because these are unique to every individual, the presence of a particular

subset in any two individuals can be used to define their common inheritance (e.g. a parent-child relationship).

DNA ligase makes a bond between an adjacent 3'-OH and 5'-phosphate end where there is a nick in one strand of duplex DNA.

DNA polymerase is an enzyme that synthesizes a daughter strand(s) of DNA (under direction from a DNA template). Any particular enzyme may be involved in repair or replication (or both).

DNA replicase is a DNA-synthesizing enzyme required specifically for replication.

DNA topoisomerase is an enzyme that changes the number of times the two strands in a closed DNA molecule cross each other. It does this by cutting the DNA, passing DNA through the break, and resealing the DNA.

DNAases are enzymes that attack bonds in DNA.

Domain of a chromosome may refer *either* to a discrete structural entity defined as a region within which supercoiling is independent of other domains; *or* to an extensive region including an expressed gene that has heightened sensitivity to degradation by the enzyme DNAase 1.

Domain of a protein is a discrete continuous part of the amino acid sequence that can be equated with a particular function.

Double-strand break (DSB) occurs when both strands of a DNA duplex are cleaved at the same site. Genetic recombination is initiated by double-strand breaks. The cell also has repair systems that act on double-strand breaks created at other times.

Downstream identifies sequences proceeding farther in the direction of expression; for example, the coding region is downstream of the initiation codon.

Effector is the target protein for the activated G protein.

Elongation is the stage in a macromolecular synthesis reaction (replication, transcription, or translation) when the nucleotide or polypeptide chain is being extended by the addition of individual subunits.

Elongation factors (EF in prokaryotes, eEF in eukaryotes) are proteins that associate with ribosomes cyclically, during addition of each amino acid to the polypeptide chain.

End labeling describes the addition of a radioactively labeled group to one end (5' or 3') of a DNA strand.

Endonucleases cleave bonds within a nucleic acid chain; they may be specific for RNA or for single-stranded or double stranded DNA.

Endoplasmic reticulum is an organelle involved in the synthesis of lipids, membrane proteins, and secretory proteins. It consists of a highly convoluted sheet of membranes, extending from the outer layer of the nuclear envelope into the cytoplasm.

Enhancer is a ds-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter.

Episome is a plasmid able to integrate into bacterial DNA.

Error-prone synthesis occurs when DNA incorporates non-complementary bases into the daughter strand.

Euchromatin comprises all of the genome in the interphase nucleus except for the heterochromatin. The euchromatin is less tightly coiled than heterochromatin, and contains the active or potentially active genes.

Excision of phage or episome or other sequence describes its release from the host chromosome as an autonomous DNA molecule.

Excision repair describes a type of repair system in which one strand of DNA is directly excised and then replaced by resynthesis using the complementary strand as template.

Exon is any segment of an interrupted gene that is represented in the mature RNA product.

Exonucleases cleave nucleotides one at a time from the end of a polynucleotide chain; they may be specific for either the 5' or 3' end of DNA or RNA.

Extra arm of tRNA lies between the TyC and anticodon arms. It is the most variable in length in tRNA, from 3-21 bases. TRNAs are called class 1 if they lack it, and class 2 if they have it.

Extrachromosomal genome in a bacterium is a self-replicating set of genes that is not part of the bacterial chromosome. In many cases, the genes are necessary for bacterial growth under certain environmental conditions.

F plasmid is an episome that can be free or integrated in *E. coli*, and which in either form can sponsor conjugation.

Facultative heterochromatin describes the inert state of sequences that also exist in active copies—for example, one mammalian X chromosome in females.

Feedback inhibition describes the ability of a small molecule product of a metabolic pathway to inhibit the activity of an enzyme that catalyzes an earlier step in the pathway.

Forward mutations inactivate a wild-type gene.

Frameshift mutations arise by deletions or insertions that are not a multiple of 3 bp; they change the frame in which triplets are translated into protein.

GO is a noncycling state in which a cell has ceased to divide.

G1 is the period of the eukaryotic cell cycle between the last mitosis and the start of DNA replication.

G2 phase is the period of the cell cycle separating the replication of a cell's chromosomes (S phase) from the following mitosis (M phase).

Gene (cistron) is the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

Gene cluster is a group of adjacent genes that are identical or related.

Genetic code is the correspondence between triplets in DNA (or RNA) and amino acids in protein.

Genome is the complete set of sequences in the genetic material of an organism. It includes the sequence of each chromosome plus any DNA in organelles.

Glycolipid has a head consisting of an oligosaccharide, linked to a fatty acid tail.

Golgi apparatus consists of individual stacks of membranes near the endoplasmic reticulum; involved in glycosylating proteins and sorting them for transport to different cellular locations. an increase in temperature (and other abuses to the cell). They occur in all organisms. They usually include chaperones that act on denatured proteins.

Helicase is an enzyme that uses energy provided by ATP hydrolysis to separate the strands of a nucleic acid duplex.

Helix-turn-helix motif describes an arrangement of two a helices that form a site that binds to DNA, one fitting into the major groove of DNA and other lying across it.

A **hemi-methylated** site is a palindromic sequence that is methylated on only one strand of DNA.

Hemimethylated DNA is methylated on one strand of a target sequence that has a cytosine on each strand.

Heterochromatin describes regions of the genome that are highly condensed, are not transcribed, and are late-replicating. Heterochromatin is divided into two types, which are called constitutive and facultative.

Heterogeneous nuclear RNA (hnRNA) comprises transcripts of nuclear genes made by RNA polymerase II; it has a wide size distribution and low stability.

Hfr cell is a bacterium that has an integrated F plasmid within its chromosome. Hfr stands for high frequency recombination, referring to the fact that chromosomal genes are transferred from an Hfr cell to an F' cell much more frequently than from an F+ cell.

Histones are conserved DNA-binding proteins that form the basic subunit of chromatin in eukaryotes. Histones H2A, H2B, H3, and H4 form an octameric core around which DNA coils to form a nucleosome.

Histone HI is external to the nucleosome.

Histone acetyltransferase (HAT) enzymes modify histones by addition of acetyl groups; some transcriptional coactivators have HAT activity.

hnRNP is the ribonucleoprotein form of hnRNA (heterogeneous nuclear RNA), in which the hnRNA is complexed with proteins.

Holliday structure is an intermediate structure in homologous recombination, where the two duplexes of DNA are connected by the genetic material exchanged between two of the four strands, one from each duplex. A joint molecule is said to be resolved when nicks in the structure restore two separate DNA duplexes.

Holoenzyme (complete enzyme) is the complex of all subunits including the permanent subunits and their temporary counterparts.

Homologous recombination (generalized recombination) involves a reciprocal exchange of sequences of DNA, e.g. between two chromosomes that carry the same genetic loci.

Hotspot describes a site in the genome at which the frequency of mutation (or recombination) is very much increased.

Housekeeping genes (constitutive genes) are those (theoretically) expressed in all cells because they provide basic functions needed for sustenance of all cell types.

Hybridization describes the pairing of complementary RNA and DNA strands to give an RNA-DNA hybrid.

In situ hybridization (cytological hybridization) is performed by denaturing the DNA of cells squashed on a microscope slide so that reaction is possible with an added single-stranded RNA or DNA; the added preparation is radioactively labeled and its hybridization is followed by autoradiography.

Induced mutations result from the action of a mutagen. The mutagen may act directly on the bases in DNA or it may act indirectly to trigger a pathway that leads to a change in DNA sequence.

Inducer is a small molecule that triggers gene transcription by binding to a regulator protein.

Initiation describes the stages of transcription up to synthesis of the first bond in RNA. This includes binding of RNA polymerase to the promoter and melting a short region of DNA into single

Initiation codon is a special codon (usually AUG) used to start synthesis of a protein.

Initiation complex in bacterial protein synthesis contains a small ribosome subunit, initiation factors, and initiator aminoacyl-tRNA bound to mRNA at an AUG initiation codon.

Initiation factors (IF in prokaryotes, eIF in eukaryotes) are proteins that associate with the small subunit of the ribosome specifically at the stage of initiation of protein synthesis.

Insertion is identified by the presence of an additional stretch of base pairs in DNA.

Insertion sequence (IS) is a small bacterial transposon that carries only the genes needed for its own transposition.

Integrase is an enzyme that is responsible for a site-specific recombination that inserts one molecule of DNA into another.

Integration of viral or another DNA sequence describes its insertion into a host genome as a region covalently linked on either side to the host sequences.

Interphase is the period between mitotic cell divisions; divided into G1, S, and G2.

Intron (intervening sequence) is a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it.

kilobase (kb) is a measure of length and may be used to refer to DNA (1000 base pairs) or to RNA (1000 bases).

Kinetochores are the structural features of the chromosome to which microtubules of the mitotic spindle attach. Its location determines the centromeric region.

Lagging strand of DNA must grow overall in the 3'-5' direction and is synthesized discontinuously in the form of short fragments (5'-3') that are later connected covalently.

Large subunit of the ribosome (50S in bacteria, 60S in eukaryotes) has the peptidyl transferase active site that synthesizes the peptide bond.

Lariat is an intermediate in RNA splicing in which a circular structure with a tail is created by a 5'-2' bond.

Leader of a protein is a short N-terminal sequence responsible for initiating passage into or through a membrane.

Leader (5' UTR) of an mRNA is the non-translated sequence at the 5' end that precedes the initiation codon.

Leading strand of DNA is synthesized continuously in the 5'-3' direction.

Library is a set of cloned fragments together representing the entire genome (genomic library) or all the expressed genes (cDNA library).

Linker DNA is all DNA contained on a nucleosome in excess of the 146 bp core DNA.

Linking number is the number of times the two strands of a closed DNA duplex cross over each other.

Locus is the position on a chromosome at which the gene for a particular trait resides; a locus may be occupied by any one of the alleles for the gene.

Long terminal repeat (LTR) is the sequence that is repeated at each end of the integrated retroviral genome.

Loop is a single-stranded region at the end of a hairpin in RNA (or single-stranded DNA); it corresponds to the sequence between inverted repeats in duplex DNA.

Luxury genes are those coding for specialized functions synthesized (usually) in large amounts in particular cell types.

Lysis describes the death of bacteria at the end of a phage infective cycle when they burst open to release the progeny of an infecting phage (because phage enzymes disrupt the bacterium's cytoplasmic membrane or cell wall). The same term also applies to eukaryotic cells; for example, when infected cells are attacked by the immune system.

Lysogeny describes the ability of a phage to survive in a bacterium as a stable prophage component of the bacterial genome.

Lytic infection of a bacterium by a phage ends in the destruction of the bacterium with release of progeny phage.

Maintenance methylase adds a methyl group to a target site that is already hemimethylated.

Megabase (Mb) is 1 million base pairs of DNA.

Messenger RNA (mRNA) is the intermediate that represents one strand of a gene coding for protein. Its coding region is related to the protein sequence by the triplet genetic code.

Metastasis describes the ability of tumor cells to leave their site of origin and migrate to other locations in the body, where a new colony is established.

Methotrexate is a drug that inhibits the enzyme DHFR (dihydrofolate reductase).

Methyltransferase (methylase) is an enzyme that adds a methyl group to a substrate, which can be a small molecule, a protein, or a nucleic acid.

Micrococcal nuclease is an endonuclease that cleaves DNA; in chromatin, DNA is cleaved preferentially between nucleosomes.

Microsatellite DNAs consist of repetitions of extremely short (typically <10 bp) units.

Minisatellite DNAs consist of ~10 copies of a short repeating sequence, the length of the repeating unit is measured in 10s of base pairs. The number of repeats varies between individual genomes.

Minor groove of DNA is 12Å across.

Minus strand DNA is the single-stranded DNA sequence that is complementary to the viral RNA genome of a plus strand virus.

Mismatch describes a site in DNA where the pair of bases does not conform to the usual G-C or A-T pairs. It may be caused by incorporation of the wrong base during replication or by mutation of a base.

Mismatch repair is a DNA repair mechanism that fixes bases that do not pair properly. The mechanism preferentially corrects the sequence of the daughter strand, by distinguishing the daughter strand and parental strand on the basis of their states of methylation.

Missense mutations change a single codon and so may cause the replacement of one amino acid by another in a protein sequence.

Mitochondrial DNA (mtDNA) is an independent DNA genome, usually circular, that is located in the mitochondrion.

Mitosis (M phase) is the process by which the cell nucleus divides, resulting in daughter cells that contain the same amount of genetic material as the parent cell.

Monocistronic mRNA codes for one protein.

Monolayer describes the growth of eukaryotic cells in culture as a layer only one cell deep.

Mutagens increase the rate of mutation by inducing changes in DNA sequence, directly or indirectly.

N-formyl-methionyl-tRNA is the aminoacyl-Trna that initiates bacterial protein synthesis. The amino group of the methionine is formylated.

Nascent RNA is a ribonucleotide chain that is still being synthesized, so that its 3' end is paired with DNA where RNA polymerase is elongating.

Nick translation describes the ability of *E. coli* DNA polymerase I to use a nick as a starting point from which one strand of a duplex DNA can be degraded and replaced by resynthesis of new material; is used to introduce radioactively labeled nucleotides into DNA *in vitro*.

Non-homologous end-joining (NHEJ) ligates blunt ends. It is common to many repair pathways and to certain recombination pathways (such as immunoglobulin recombination).

Nonsense mutation is any change in DNA that causes a (termination) codon to replace a codon representing an amino acid.

Nuclear envelope is a layer of two concentric membranes (inner and outer nuclear membranes) that surrounds the nucleus and its underlying intermediate filament lattice, the nuclear lamina. The nuclear envelope is penetrated by nuclear pores. The

outer membrane is continuous with the membrane of the rough endoplasmic reticulum.

Nucleic acids are molecules that encode genetic information. They consist of a series of nitrogenous bases connected to ribose molecules that are linked by phosphodiester bonds. DNA is deoxyribonucleic acid, and RNA is ribonucleic acid.

Nucleoid is the region in a prokaryotic cell that contains the genome. The DNA is bound to proteins and is not enclosed by a membrane.

Nucleosome is the basic structural subunit of chromatin, consisting of ~200 bp of DNA and an octamer of histone proteins.

Ochre codon is the triplet UAA, one of the three termination

Okazaki fragments are the short stretches of 1000-2000 bases produced during discontinuous replication; they are later joined into a covalently intact strand.

Oncogenes are genes whose products have the ability to transform eukaryotic cells so that they grow in a manner analogous to tumor cells. Oncogenes carried by retroviruses have names of the form *v-onc*.

Opal codon is the triplet UGA, one of the three termination codons that end protein synthesis. It has evolved to code for an amino acid in a small number of organisms or organelles.

Open complex describes the stage of initiation of transcription when RNA polymerase causes the two strands of DNA to separate to form the "transcription bubble".

Open reading frame (ORF) is a sequence of DNA consisting of triplets that can be translated into amino acids starting with an initiation codon and ending with a termination codon.

Operator is the site on DNA at which a repressor protein binds to prevent transcription from initiating at the adjacent promoter.

Operon is a unit of bacterial gene expression and regulation, including structural genes and control elements in DNA recognized by regulator gene product(s).

Opine is a derivative of arginine that is synthesized by plant cells infected with crown gall disease.

P site of the ribosome is the site that is occupied by peptidyl tRNA, the tRNA carrying the nascent polypeptide chain, still paired with the codon to which it bound in the A site.

Palindrome is a DNA sequence that reads the same on each strand of DNA when the strand is read in the 5' to 3' direction. It consists of adjacent inverted repeats.

Parental strand or duplex of DNA refers to the DNA that will be replicated.

Passive transport describes movement of molecules along their electrochemical gradient; no energy is required.

Peptidyl transferase is the activity of the ribosomal 50S subunit that synthesizes a peptide bond when an amino acid is added to a growing polypeptide chain. The actual catalytic activity is a property of the rRNA.

Phage T4 is a virus that infects E. coli causing lysis of the bacterium.

Phospholipid is a lipid that has a positively charged head that is linked by a phosphate group to the fatty acid tails.

Photoreactivation is a type of direct repair that involves a light dependent enzyme.

Pilin is the subunit that is polymerized into the pilus in bacteria.

Pilus (pili) is a surface appendage on a bacterium that allows the bacterium to attach to other bacterial cells. It appears like a short, thin, flexible rod. During conjugation, pili are used to transfer DNA from one bacterium to another.

Plaque is an area of clearing in a bacterial lawn. It is created by a single phage particle that has undergone multiple rounds of lytic growth.

Plasmid is a circular, extra-chromosomal DNA. It is autonomous and can replicate itself.

Plus strand DNA is the strand of the duplex sequence representing a retrovirus that has the same sequence as that of the RNA.

Point mutation is a change in the sequence of DNA involving a single base pair.

Polarity refers to the effect of a mutation in one gene in influencing the expression (at transcription or translation) of subsequent genes in the same transcription unit.

Poly(A) is a stretch of 200 bases of adenylic acid that is added to the 3' end of mRNA following its synthesis.

Poly(A) polymerase is the enzyme that adds the stretch of polyadenylic acid to the 3' of eukaryotic mRNA. It does not use a template.

Polycistronic mRNA includes coding regions representing more than one gene.

Polyribosome (polysome) is an mRNA that is simultaneously being translated by several ribosomes.

Post-translational translocation (posttranslational) is the movement of a protein across a membrane after the synthesis of the protein is completed and it has been released from the ribosome.

Pre-mRNA is used to describe the nuclear transcript that is processed by modification and splicing to give an mRNA.

Premature termination describes the termination of protein or of RNA synthesis before the chain has been completed. In protein synthesis it can be caused by mutations that create termination codons within the coding region. In RNA synthesis it is caused by various events that act on RNA polymerase. A protein to be imported into an organelle or secreted from bacteria is called a "**preprotein**" until its signal sequence has been removed.

Primary transcript is the original unmodified RNA product corresponding to a transcription unit.

Primase is a type of RNA polymerase that synthesizes short segments of RNA that will be used as primers for DNA replication.

Primer is a short sequence (often of RNA) that is paired with one strand of DNA and provides a free 3'-OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

Primosome describes the complex of proteins involved in the priming action that initiates replication on 4>X-type origins. It is also involved in restarting stalled replication forks.

Prion is a proteinaceous infectious agent, which behaves as an inheritable trait, although it contains no nucleic acid. Examples are PrP^{Sc}, the agent of scrapie in sheep and bovine spongiform encephalopathy.

Processing of RNA describes changes that occur after its transcription, including modification of the 5' and 3' ends, internal methylation, splicing, or cleavage.

Processivity describes the ability of an enzyme to perform multiple catalytic cycles with a single template instead of dissociating after each cycle.

Promoter is a region of DNA where RNA polymerase binds to initiate transcription.

Proofreading refers to any mechanism for correcting errors in protein or nucleic acid synthesis that involves scrutiny of individual units *after* they have been added to the chain.

Prophage is a phage genome covalently integrated as a linear part of the bacterial chromosome.

Provirus is a duplex sequence of DNA integrated into a eukaryotic genome that represents the sequence of the RNA genome of a retrovirus.

PrP is the protein that is the active component of the prion that causes scrapie and related diseases. The form involved in the disease is called PrP^{Sc}.

Pseudogenes are inactive but stable components of the genome derived by mutation of an ancestral active gene. Usually they are inactive because of mutations that block transcription or translation or both.

Puromycin is an antibiotic that terminates protein synthesis by mimicking a tRNA and becoming linked to the nascent protein chain.

Pyrimidine dimer is formed when ultraviolet irradiation generates a covalent link directly between two adjacent pyrimidine bases in DNA. It blocks DNA replication.

Reading frame is one of the three possible ways of reading a nucleotide sequence. Each reading frame divides the sequence into a series of successive triplets. There are three possible reading frames in any sequence, depending on the starting point. If the

first frame starts at position 1, the second frame starts at position 2, and the third frame starts at position 3.

Receptor is a transmembrane protein, located in the plasma membrane, which binds a ligand in a domain on the extracellular side, and as a result has a change in activity of the cytoplasmic domain.

Recessive allele is obscured in the phenotype of a heterozygote by the dominant allele, often due to inactivity or absence of the product of the recessive allele.

Reciprocal translocation exchanges part of one chromosome with part of another chromosome.

Recombinant progeny have a different genotype from that of either parent.

Recombination-repair is a mode of filling a gap in one strand of duplex DNA by retrieving a homologous single strand from another duplex.

Regulator gene codes for a product (typically protein) that controls the expression of other genes (usually at the level of transcription).

Release factor (RF) is required to terminate protein synthesis to cause release of the completed polypeptide chain and the ribosome from mRNA.

Renaturation describes the reassociation of denatured complementary single strands of a DNA double helix.

Repair of damaged DNA can take place by repair synthesis, when a strand that has been damaged is excised and replaced by the synthesis of a new stretch. It can also take place by recombination reactions, when the duplex region containing the damaged is replaced by an undamaged region from another copy of the genome.

Repetitive DNA behaves in a reassociation reaction as though many (related or identical) sequences are present in a component, allowing any pair of complementary sequences to reassociate.

Replication of duplex DNA takes place by synthesis of two new strands that are complementary to the parental strands. The parental duplex is replaced by two identical daughter duplexes, each of which has one parental strand and one newly synthesized strand. It is called semiconservative because the conserved units are the single strands of the parental duplex.

Replication eye is a region in which DNA has been replicated within a longer, unreplicated region.

Replication fork (growing point) is the point at which strands of parental duplex DNA are separated so that replication can proceed. A complex of proteins including DNA polymerase is found at the fork.

Replicative transposition describes the movement of a transposon by a mechanism in which first it is replicated, and then one copy is transferred to a new site.

Replicon is a unit of the genome in which DNA is replicated. Each replicon contains an origin for initiation of replication.

Replisome is the multiprotein structure that assembles at the bacterial replicating fork to undertake synthesis of DNA. It contains DNA polymerase and other enzymes.

Repression describes the ability of bacteria to prevent synthesis of certain enzymes when their products are present; more generally, refers to inhibition of transcription (or translation) by binding of repressor protein to a specific site on DNA (or mRNA).

Repressor is a protein that inhibits expression of a gene. It may act to prevent transcription by binding to an operator site in DNA, or to prevent translation by binding to RNA.

Resolution occurs by a homologous recombination reaction between the two copies of the transposon in a cointegrate. The reaction generates the donor and target replicons, each with a copy of the transposon.

Resolvase is the enzyme activity involved in site-specific recombination between two transposons present as direct repeats in a cointegrate structure.

Restriction endonucleases recognize specific short sequences of DNA and cleave the duplex (sometimes at target site, sometimes elsewhere, depending on type).

Restriction fragment length polymorphism (RFLP) refers to inherited differences in sites for restriction enzymes (for example, caused by base changes in the target site) that result in differences in the lengths of the fragments produced by cleavage with the relevant restriction enzyme. RFLPs are used for genetic mapping to link the genome directly to a conventional genetic marker.

Retroposon (retrotransposon) is a transposon that mobilizes via an RNA form; the DNA element is transcribed into RNA, and then reverse-transcribed into DNA, which is inserted at a new site in the genome. The difference from retroviruses is that the retroposon does not have an infective (viral) form.

Retrovirus is an RNA virus with the ability to convert its sequence into DNA by reverse transcription.

Reverse transcriptase is an enzyme that uses a template of single-stranded RNA to generate a double-stranded DNA copy.

Reverse transcription is synthesis of DNA on a template of RNA; accomplished by reverse transcriptase enzyme. Revertants are derived by reversion of a mutant cell or organism.

Rho factor is a protein involved in assisting *E. coli* RNA polymerase to terminate transcription at certain terminators (called rho-dependent terminators).

Rho-dependent terminators are sequences that terminate transcription by bacterial RNA polymerase in the presence of the rho factor.

Ribonucleases (RNAases) are enzymes that cleave RNA. They may be specific for single-stranded or for double-stranded RNA, and may be either endonucleases or exonucleases.

Ribonucleoprotein is a complex of RNA with proteins.

Ribosomal DNA (rDNA) is usually a tandemly repeated series of genes coding for a precursor to the two large rRNAs.

Ribosomal RNA (rRNA) is a major component of the ribosome. Each of the two subunits of the ribosome has a major rRNA as well as many proteins.

Ribosome is a large assembly of RNA and proteins that synthesizes proteins under direction from an mRNA template. Bacterial ribosomes sediment at 70S, eukaryotic ribosomes at 80S. A ribosome can be dissociated into two subunits.

Ribosome-binding site is a sequence on bacterial mRNA that includes an initiation codon that is bound by a 30S subunit in the initiation phase of protein synthesis.

Ribozyme is an RNA that has catalytic activity.

A helix is said to be **right-handed** if the turns runs clockwise along the helical axis.

RNA polymerases are enzymes that synthesize RNA using a DNA template (formally described as DNA-dependent RNA polymerases).

RNA splicing is the process of excising the sequences in RNA that correspond to introns, so that the sequences corresponding to exons are connected into a continuous mRNA.

Rolling circle is a mode of replication in which a replication fork proceeds around a circular template for an indefinite number of revolutions; the DNA strand newly synthesized in each revolution displaces the strand synthesized in the previous revolution, giving a tail containing a linear series of sequences complementary to the circular template strand.

Rough endoplasmic reticulum (rough ER) refers to the region of the endoplasmic reticulum to which ribosomes are bound. It is the site of synthesis of membrane proteins and secretory proteins.

S phase is the restricted part of the eukaryotic cell cycle during which synthesis of DNA occurs.

Satellite DNA (simple-sequence DNA) consists of many tandem repeats (identical or related) of a short basic repeating unit.

Scrapie is an infective agent made of protein.

Selfish DNA describes sequences that do not contribute to the genotype of the organism but have self-perpetuation within the genome as their sole function.

Semiconservative replication is accomplished by separation of the strands of a parental duplex, each then acting as a template for synthesis of a complementary strand.

Semidiscontinuous replication is a mode in which one new strand is synthesized continuously while the other is synthesized discontinuously.

Shine-Dalgarno sequence is part or all of the polypurine sequence TATAATG centered about 10 bp before the AUG initiation codon on bacterial mRNA. It is complementary to the sequence at the 3' end of 16S rRNA.

Shotgun cloning analyzes an entire genome in the form of randomly generated fragments.

Sigma factor is the subunit of bacterial RNA polymerase needed for initiation; it is the major influence on selection of promoters.

Silencer is a short sequence of DNA that can inactivate expression of a gene in its vicinity.

Silencing describes the repression of gene expression in a localized region, usually as the result of a structural change in chromatin.

Silent mutations do not change the product of a gene.

Single-strand binding protein (SSB) attaches to single stranded DNA, thereby preventing the DNA from forming a duplex.

Site-specific recombination (specialized recombination) occurs between two specific (not necessarily homologous) sequences, as in phage integration/excision or resolution of cointegrate structures during transposition.

Small nuclear RNA (snRNA) is one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions.

Small subunit of the ribosome (30S in bacteria, 40S in eukaryotes) binds the mRNA.

Smooth ER consists of a regions of endoplasmic reticulum devoid of ribosomes.

snRNPs (snurps) are small nuclear ribonucleoproteins (snRNAs associated with proteins).

Somatic mutation is a mutation occurring in a somatic cell, and therefore affecting only its daughter cells; it is not inherited by descendants of the organism. Somatic mutations are generally spontaneous, but a case of directed mutation occurs in the

immune system where more diversity is generated in rearranged immunoglobulin genes by somatic mutation.

SOS response in *E. coli* describes the coordinate induction of many enzymes, including repair activities, in response to irradiation or other damage to DNA; results from activation of protease activity by RecA to cleave LexA repressor.

Spacer is a sequence in a gene cluster that separates the repeated copies of the transcription unit.

Spindle guides the movement of chromosomes during cell division. The structure is made up of microtubules.

Spliceosome is a complex formed by the snRNPs that are required for splicing together with additional protein factors.

Spontaneous mutations occur in the absence of any added reagent to increase the mutation rate, as the result of errors in replication (or other events involved in the reproduction of DNA) or by environmental damage.

Startpoint (start site) refers to the position on DNA corresponding to the first base incorporated into RNA.

Stem is the base-paired segment of a hairpin structure in RNA.

Stop codon (termination codon) is one of three triplets (UAG, UAA, UGA) that causes protein synthesis to terminate. They are also known historically as *nonsense codons*. The UAA codon is called ochre, and the UAA codon as called amber, after the names of the nonsense mutations by which they were originally identified.

Strand displacement is a mode of replication of some viruses in which a new DNA strand grows by displacing the previous (homologous) strand of the duplex.

Structural gene codes for any RNA or protein product other than a regulator.

Supercoiling describes the coiling of a closed duplex DNA in space so that it crosses over its own axis.

Suppression describes the occurrence of changes that eliminate the effects of a mutation without reversing the original change in DNA.

Suppressor is a second mutation that compensates the effects of a primary mutation, so that the combination of the two mutations restores wild phenotype.

T-DNA is the segment of the Ti plasmid of *Agrobacterium tumefaciens* that is transferred to the plant cell nucleus during infection. It carries genes that transform the plant cell.

TATA box is a conserved A-T-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcription unit; may be involved in positioning the enzyme for correct initiation.

Telomerase is the ribonucleoprotein enzyme that creates repeating units of one strand at the telomere, by adding individual bases to the DNA 3' end, as directed by an RNA sequence in the RNA component of the enzyme.

Telomere is the natural end of a chromosome; the DNA sequence consists of a simple repeating unit with a protruding single-stranded end that may fold into a hairpin.

Termination is a separate reaction that ends a macromolecular synthesis reaction (replication, transcription, or translation), by stopping the addition of subunits, and (typically) causing dissolution of the synthetic apparatus.

Terminator is a sequence of DNA that causes RNA polymerase to terminate transcription.

Ti plasmid is an episome of the bacterium *Agrobacterium tumefaciens* that carries the genes responsible for the induction of crown gall disease in infected plants.

Transcript is the RNA product produced by copying one strand of DNA. It may require processing to generate a mature RNA.

Transcription describes synthesis of RNA on a DNA template.

Transcription factor is required for RNA polymerase to initiate transcription at specific promoter(s), but is not itself part of the enzyme.

Transfection of eukaryotic cells is the acquisition of new genetic markers by incorporation of added DNA.

Transfer RNA (tRNA) is the intermediate in protein synthesis that interprets the genetic code. Each tRNA can be linked to an amino acid. The tRNA has an anticodon sequence that complementary to a triplet codon representing the amino acid.

Transformation of bacteria describes the acquisition of new genetic markers by incorporation of added DNA.

Transformation (oncogenesis) of eukaryotic cells refers to their conversion to a state of unrestrained growth in culture.

Transformed cells are cultured cells that have acquired many of the properties of cancer cells.

Transforming principle is DNA that is taken up by a bacterium and whose expression then changes the properties of the recipient cell.

Transgene is a gene that is introduced into a cell or animal from an external source.

Transgenic animals are created by introducing new DNA sequences into the germ line via addition to the egg.

Transition is a mutation in which one pyrimidine is substituted by the other or in which one purine is substituted for the other.

Translation is synthesis of protein on the mRNA template.

Translocation describes a rearrangement in which part of a chromosome is detached by breakage and then becomes attached to some other chromosome.

Transposase is the enzyme activity involved in insertion of transposon at a new site.

Transposition refers to the movement of a transposon to a new site in the genome.

Transposon (transposable element) is a DNA sequence able to insert itself at a new location in the genome, without having any sequence relationship with the target locus.

Transversion is a mutation in which a purine is replaced by a pyrimidine or vice versa.

Tumor suppressor is identified by a loss-of-function mutation that contributes to cancer formation. They usually function to prevent cell division or to cause death of abnormal cells. The two most important are p53 and RB.

Tumor virus has the ability to transform an animal cell into a cancerous state.

Ty stands for transposon yeast, the first transposable element to be identified in yeast.

Type I topoisomerase is an enzyme that changes the topology of DNA by nicking and resealing one strand of DNA.

Type II topoisomerase is an enzyme that changes the topology of DNA by nicking and resealing both strands of DNA.

Ubiquitin has a highly conserved sequence of 76 amino acids. It is linked via its COOH group to the ε NH₂ group of a lysine residue in a target protein.

Unidirectional replication refers to the movement of a single replication fork from a given origin.

Upstream identifies sequences proceeding in the opposite direction from expression; for example, the bacterial promoter is upstream of the transcription unit, the initiation codon is upstream of the coding region.

Virion is the physical virus particle (irrespective of its ability to infect cells and reproduce).

Viroid is a small infectious nucleic acid that does not have a protein coat.

Virulent phage mutants are unable to establish lysogeny.

VNTR (variable number tandem repeat) regions describe very short repeated sequences, including microsatellites and minisatellites.

Wobble hypothesis accounts for the ability of a tRNA to recognize more than one codon by unusual (non-G-C, non-A-T) pairing with the third base of a codon.

Yeast artificial chromosome (YAC) is a synthetic DNA molecule that contains an origin for replication, a centromere to support segregation, and telomeres to seal the ends. It is used as a means to propagate whatever genes it carries in yeast cells.