Chapter 1. Introduction

1. Molecular Biology

The term molecular biology was first used in 1945 by William Astbury who was referring to the study of the chemical and physical structure of biological macromolecules. By that time, biochemists had discovered many fundamental intracellular chemical reactions. The importance of specific reactions and of protein structure in defining the numerous properties of cells was also appreciated. However, the development of molecular biology had to await the understanding that the most advantageous approaches would be made by studying "simple" systems such as bacteria and bacteriophages which yield information about the basic biological processes more readily than animal cells. In fact, the faith in the basic uniformity of life processes was an important factor in rapid growth of molecular biology. That is, it was believed that fundamental biological principles that govern the activity of simple organisms, such as bacteria and viruses, must apply to more complex cells; only the details should vary. This faith has been amply justified by experimental results.

The roots of molecular biology were established in 1953 when an Englishman, Francis Crick and a young American, James Watson working at Medical Research Council Unit, Cavendish Laboratory, Cambridge, proposed a double helical model for the structure of DNA (deoxyribonucleic acid) molecule which was well known as the chemical bearer of genetic informations of certain microorganisms (bacteria, bacteriophages, etc.) due to pioneer discoveries made by Grifith (1928), Avery, Macleod and McCarthy (1944) and Hershey and Chase (1952). This discovery was followed by a thorough search of occurrence of DNA as the genetic material in other microorganisms, plants and animals and also by investigations of the molecular and atomic nature of different reactions of living cells. From all these studies has emerged the realization that the basic chemical organization and the metabolic processes of all living things are remarkably similar despite their morphological diversity and that the physical and chemical principles governing living systems are similar to those governing non-living systems.

The present understanding of molecular biology is that in most organisms the phenotype or the body structure and function ultimately depend for their determination on the structural and functional (i.e., enzymatic) proteins or polypeptides. The synthesis of polypeptides is specified, directed and regulated by self-duplicating genes which are borne within molecules of DNA which is the universally accepted chemical bearer of genetic informations of most living organisms except certain viruses in which this function is carried by RNA, another nucleic acid. The genetic informations for polypeptide synthesis are initially dictated by the disposition of nitrogen bases in DNA molecule and are copied down by the process of transcription. During transcription stage copies (that is, transcripts) of an individual gene or genes are synthesized. These copies are molecules of RNA that include such familiar classes as ribosomal RNA, messenger RNA and transfer RNA. The biochemical interplay of these RNA copies which leads to the synthesis of a polypeptide chain, is called translation, meaning, literally, that the genetic message encoded in a messenger RNA molecule is translated into the linear sequence of amino acids in a polypeptide. The polypeptide in its turn determines the phenotype of the organism.

1. HISTORICAL BACKGROUND

The molecular biology is a very young biological discipline and has a very short history. Certain notable accomplishments of molecular biologists can be summarized as follows :

1928 F. Grifth discovered the phenomenon of transformation in bacteria.

1934 M. Schlesinger demonstrated that the bacteriophages are composed of DNA and protein.

1941 G.W. Beadle and E.L.Tatum published their classical study on the biochemical genetics of Neurospora.

1944 O.T. Avery, C.M. MacLeod and M. McCarthy recognized the DNA nature of transforming principle of pneumococcus bacteria. The fact suggested that it is DNA and not protein which is the hereditary chemical.

1948 A. Boivin, R.Vendrely and C. Vendrely showed that in the different cells of an organism the quantity of DNA for each haploid set of chromosome is constant.

1950 E. Chargaff demonstrated that in DNA the numbers of adenine and thymine groups are always equal and so are the numbers of guanine and cytosine groups.

1952 A.D. Hershey and M. Chase demonstrated that only the DNA of T2 bacteriophage enters the host, the bacterium Escherichia coli, whereas the protein remains behind.

1953 J. D. Watson and F.H.C. Crick proposed a model for DNA comprising of two helically intertwined chains tied together by hydrogen bonds between the purines and pyrimidines.

1956 A. Gierer and G. Schramm demonstrated that RNA is the genetic material of tobacco mosaic virus (TMV).

1957 H. Fraenkel-Contrat and B.Singer separated RNA from the protein of TMV viruses, produced hybrid RNA viruses and confirmed the view that RNA is the genetic material of some viruses. Mathew Meselson and Franklin W. Stahl performed a density-gradient experiment (using heavy isotope of nitrogen, 15N) in bacteria to confirm the Watson and Crick's semiconservative theory of DNA replication.

1958 G. Beadle and E. Tatum received Nobel Prize for their contribution in biochemical genetics of fungus. J. Laderberg got Nobel Prize for the discovery of bacterial recombination.

1959 R.L. Sinsheimer isolated singlestranded DNA from a small virus φ-X- 174 which attacks Escherichia coli. S. Ochoa ; A. Kornberg received Nobel Prize for artificial synthesis of nucleic acids.

1961 M.W. Nirenberg and J.H. Matthaei cracked the messenger RNA code. F.H.C.Crick and his colleagues showed that the genetic language is made up of three-letter words (i.e., triplet codons). F.Jacob and J. Monod put forward the operon concept.

1962 J. Watson and F. Crick ; M. Wilkens got Nobel Prize for the discovery of molecular nature of DNA.

1963 J.P. Waller reported that nearly one-half of all proteins in E.coli cells have the amino acid methionine in the N-terminal position.

1964 K.A. Marcker and F. Sanger discovered a peculiar aminoacyltRNA in E.coli, called N-formyl- methionyl - tRNA and suggested that this molecule may play a role in the special mechanism of chain elongation. R.W. Holley and his colleagues gave detailed structure of alanyl tRNA (tRNA ala) from yeast. Holley died in 1993.

1965 F.H.C. Crick proposed the wobble hypothesis for anticodons of tRNA and explained how several codons meant for same amino acid are recognized by same tRNA. H.

Wallace and M.L. Birnstiel isolated ribosomal RNA genes in Xenopus. F. Jacob., A. Lwoff, and J. Monod received Nobel Prize for the discovery of protein synthesis mechanism in virus.

1968 R.W. Holley ; H.G. Khorana and M.W. Nirenberg got Nobel Prize for deciphering the genetic code.

1969 A.D. Hershey, M. Delbruck and S.E. Luria shared Nobel Prize in medicine for their contribution to replication and recombination in viruses (bacteriophages). Britten and Davidson proposed the gene-battery model for regulation of protein synthesis in eukaryotes.

1970 Howard Temin and David Baltimore demonstrated the synthesis of DNA on RNA template tumour viruses. Both were awarded Nobel Prize in 1975 for the discovery of an enzyme called RNA directed DNA polymerase (or reverse transcriptase) which is present in the core of virus particle (rous sarcoma virus). Biotechnology emerged as a new discipline due to marriage of biological science with technology (see Dubey, 1995). Knippers ; Kornberg and Gefter ; Moses and Richardson isolated DNA-polymerase-II enzyme.

1972 Mertz and Davis in 1972 demonstrated that cohesive termini of cleaved DNA molecule could be covalently sealed with *E.coli* DNA ligase and were able to produce recombinant DNA molecules. Cohen *et al*., for the first time reported the cloning of DNA by using plasmid as vector. R. Porter; G.M. Edelman received Nobel Prize (physiology and medicine) for the discovery of chemical structure of antibodies. C.B. Anfinsen; S. Moore and W.H. Stein got Nobel Prize (chemistry) for the discovery of chemical structure and activity of the enzyme ribonuclease.

1973 S.H. Kim suggested three dimensional structure, i.e., L-shaped model, of tRNA.

1975 E.M. Southern developed a method, called Southern blotting technique for analysing the related genes in a DNA restriction fragment. D. Pribnow discovered Pribnow box or minus ten sequence in E. coli genome.

1977 P.A. Sharp and R.J. Roberts discovered split genes of adenovirus. D.S. Hogness, I.B. David and N. Davidson studied split genes for 28 S rRNA in Drosophila. P. Chambon, P. Leder and R.A. Flavell studied split genes of B'globin, ovalbumin and tRNA. Itakura et al., first of all produced human insulin (humulin) by means of recombinant technology.

1978-79 W. Gilbert first of all used the terms exon and intron (for split genes).

1978 Hinnen et al., first of all described the transformation of yeast (Saccharomyces cervisae) by the help of plasmid of E.coli.

1979 Khorana reported completion of the total synthesis of a biologically functional gene. Alwine et al., developed northern blotting technique in which mRNA bands are blot transferred from the gel onto chemically reactive paper. Towbin et al., developed the western blotting technique to find out the newly encoded protein by a transformed cell.

1980 Fredrick Sanger got the second Nobel Prize for discovering complete sequence of 5400 nucleotides of single stranded DNA of $\varphi \times 174$ bacteriophage.

1982 A. Klug was awarded Nobel Prize in chemistry for providing three-dimensional structure of tRNAs. Rubin and Spradling for the first time introduced Drosophila gene of xanthine dehydrogenase into a P-element (= parental element) which then was microinjected into embryo deficient for this gene. R.D. Palmiter and R.L. Brinter produced transgenic mice by genetic engineering.

1983 Marilyn Kozak proposed the scanning hypothesis for initiation of translation by eukaryotic ribosomes.

1984 Robert Tijan identified a DNA-binding protein called SP1 which is involved in eukaryotic gene regulation.

1985 Karry Mullis discovered polymerase chain reaction (PCR) which is widely exploited in gene cloning for genetic engineering. He made use of a thermostable enzyme (acts best on 720 C temperature), called Taq DNA polymerase, isolated from Thermus aquaticus.

1984,86 Alec Jeffreys discovered the technique of DNA fingerprinting.

1987 S. Tonegawa was awarded Nobel Prize for discovering the mode of rearrangements of DNA sequences of mammalian immunoglobulin genes to produce a large variety of antibodies. Stanford and coworkers developed the particle bombardment gun which shooted foreign DNA into plant cells or tissues at a very high speed.

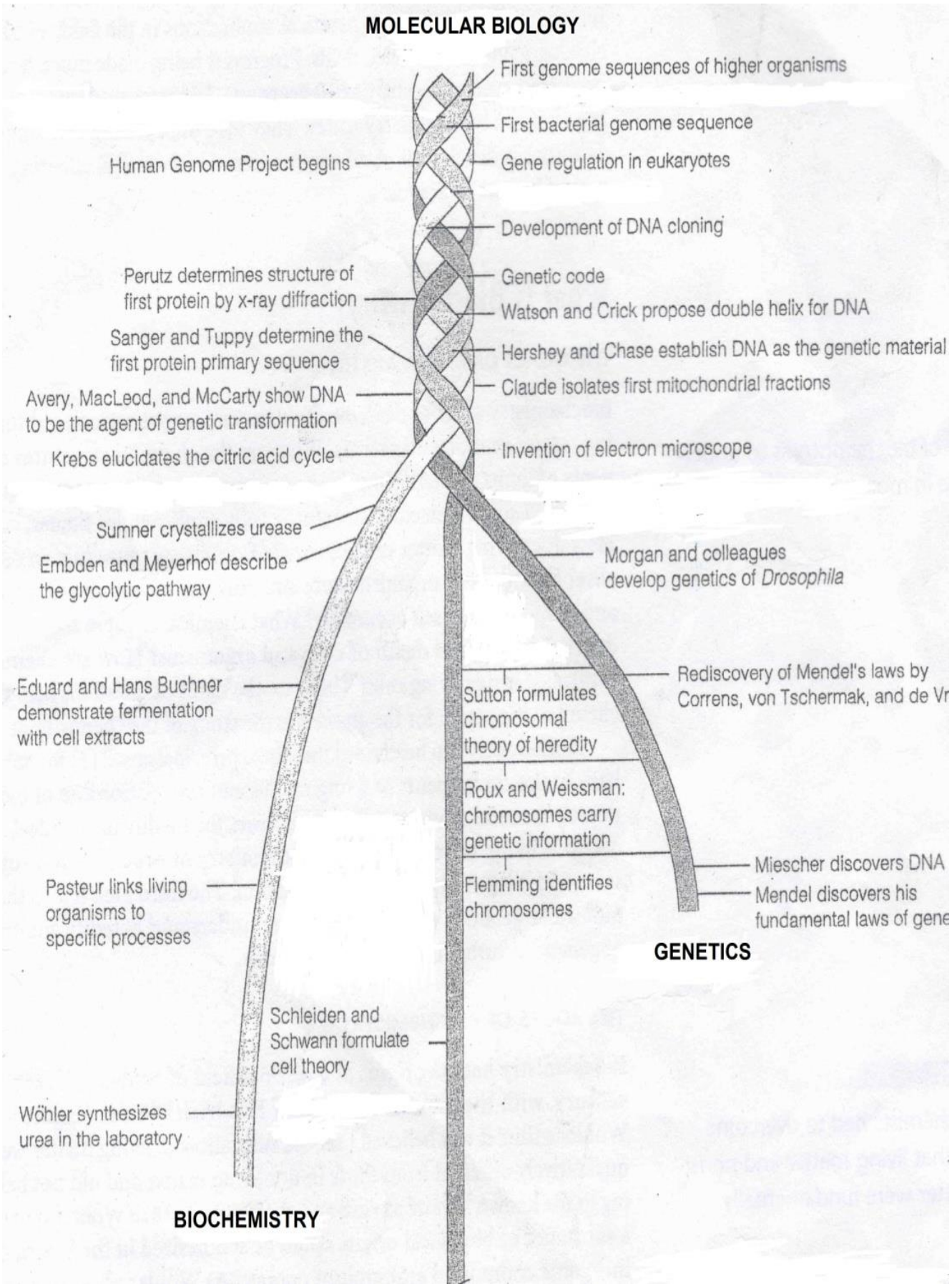
1988 J.W. Black, G.B. Elion and G.H. Hitchings were awarded Nobel Prize for formulating drugs such as 6- mercaptopurine and thioguanine, which lead to inhibition of DNA synthesis and of cell division. This prove effective in cancer chemotherapy. They also designed drugs for treating gout, malaria and viral infections such as herpes.

1989 T. Cech and S. Altman awarded Nobel Prize for showing enzymatic role of some RNA molecules, called ribozymes.

1991 Dr. Lalji Singh at CCMB, Hyderabad has developed a new technique of DNA fingerprinting by using BKM-DNA probe (BKM = banded krait minor satellite). He discovered this probe while he was working on sex determination in snake, the banded krait (Bungarus fasciatus) for his Ph.D. work.

1992 Edwin G. Krebs and Edmond H. Fisher were awarded Nobel Prize for the pioneering work on "reversible protein phosphorylation as a biological regulator mechanism." Phosphorylation of proteins is shown to affect transcription, translation, cell division and many other cellular processes. Prof. Asis Dutta of JNU, New Delhi, was selected for the Birla Award for Science and Technology for cloning and characterization of two novel genes–gene for oxalate decarboxylase from Lathyrus sativus (in 1991) and gene for a seed specific nutritionally balanced protein from Amaranthus (in 1992).

1993 M.J. Chamberlain proposed the inchworm model for elongation of transcript of DNA template. This year's Nobel Prize in chemistry was shared by Kary Mullis (for the discovery of PCR) with Michael Smith (for site directed mutagenesis).


MOLECULAR BIOLOGY
First genome sequences of higher organisms
First bacterial genome sequence
Human Genome Project begins
Gene regulation in eukaryotes
Development of DNA cloning
Perutz determines structure of first protein by x-ray diffraction
Genetic code
Watson and Crick propose double helix for DNA
Sanger and Tuppy determine the first protein primary sequence
Hershey and Chase establish DNA as the genetic material
Avery, MacLeod, and McCarty show DNA to be the agent of genetic transformation
Claude isolates first mitochondrial fractions
Krebs elucidates the citric acid cycle
Invention of electron microscope
Sumner crystallizes urease
Embden and Meyerhof describe the glycolytic pathway
Morgan and colleagues develop genetics of Drosophila
Eduard and Hans Buchner demonstrate fermentation with cell extracts
Rediscovery of Mendel's laws by Correns, von Tschermak, and de Vr
Sutton formulates chromosomal theory of heredity
Roux and Weissman: chromosomes carry genetic information
Miescher discovers DNA
Pasteur links living organisms to specific processes
Flemming identifies chromosomes
Mendel discovers his fundamental laws of gene
GENETICS
Schleiden and Schwann formulate cell theory
Wöhler synthesizes urea in the laboratory
BIOCHEMISTRY

2. Achievements of Molecular Biology

Asifa Akhtar. There have been a number of important concepts that have emerged. One that particularly jumps to mind is the importance of epigenetics in gene regulation. The field of epigenetics has flourished over the past 10 years. It is clear that chromatin provides an ideal platform for various posttranslational modifications on DNA and histones, which act as a signalling platform for various cellular processes. I also think that the discovery that a combination of four transcription factors can induce a pluripotent state was phenomenal and has stimulated a lot of research in the stem cell field[1]. Last, but not least, the involvement of non-coding RNAs in various cellular and nuclear processes is totally fascinating. The mechanisms by which long non-coding RNAs regulate gene expression await exciting discoveries in the coming years.

Elaine Fuchs. For the stem cell field, there is no question that the findings of Shinya Yamanaka and his co-worker Kazutoshi Takahashi were paradigm-shifting. Their work reported the creation of induced pluripotent stem (iPS) cells from mouse skin fibroblasts when cultured in embryonic stem cell (ESC) conditions[1]. It was remarkable that transient overexpression of a mere four transcription factors, OCT4, SOX2, MYC and Krüppel-like factor 4 (KLF4) — all naturally expressed by ESCs — could achieve this dramatic dedifferentiation of fibroblasts. This finding has allowed researchers to derive patient-tailored iPS cells to study the biology of a host of different human diseases — a first step, but a major one, for the future development of new drugs and treatments in medicine.

Tim Mitchison. Reaction–diffusion gradients specifying positional information inside cells. Gradients of signalling molecules were long known in developmental biology and paracrine physiology. But gradients inside cells being used as a spatial organizing system is a new concept. Bicoid, a classic developmental morphogen, diffuses inside a syncytium, but this is a special case. Gradients of RAN•GTP from mitotic chromatin and of Aurora B activity from chromatin in M phase and midzones in cytokinesis are classic cellular signals that we now know organize space inside cells using a reaction–diffusion mechanism. I attribute this concept to Eric Karsenti, who mooted the idea in the mid 1980s for signals diffusing away from DNA in eggs. However, it wasn't proven until the development of fluorescence resonance energy transfer (FRET)-based activity biosensors in the past decade[2,3]. In general, fluorescence sensors of biochemical activity are a very important development.

Reuben J. Shaw. One area close to our own work is the unexpected re-emergence of metabolism and its relationship to growth control and cancer. Advances in autophagy continue to amaze me in terms of how little basic information we actually have on how a cell works. Autophagy regulators are highly conserved proteins in a central cell biological process that is deregulated in common human diseases, yet much of the biochemical framework for this process

has been decoded only recently. Other newly decoded central regulators and processes, ranging from cilia to sirtuins, microRNAs (miRNAs) and pathways such as those involving Hippo and mammalian target of rapamycin (mTOR), underlying so much biology, have changed half of what we know. These are very exciting times.

Daniel St Johnston. Several surprising concepts have emerged during the past decade: first, the amazing extent to which basic cell biological processes have been conserved during the evolution of eukaryotes; second, how much gene regulation is post-transcriptional, particularly through small non-coding RNAs; third, how basic cellular processes, such as endocytic trafficking, microtubule dynamics or mitochondrial behaviour are modulated during the course of normal development; and last, the wide range of cell biological and developmental events that are regulated in response to cellular stresses, such as DNA damage or nutrient deprivation, and how these are used as signals during normal development.

The most important technical advances have been high-throughput sequencing, which has provided the complete sequence of many genomes, and the use of RNA interference (RNAi) to knock down gene function.

Andreas Strasser. One important concept to emerge is the ability to reprogramme differentiated cells, such as fibroblasts or hepatocytes, to assume a pluripotent stem cell fate. Another key finding has been the discovery that signal transducers undergo complex processes of modification by different forms of ubiquitin linkages and that these regulate cellular responses to extracellular signals, such as ligands of the tumour necrosis factor (TNF) family. In addition, an important result has been the discovery that caspase 8 regulates both apoptosis and another cell death process, termed necroptosis. It will now be important to determine the roles of necroptosis in cell death processes that are thought to shape embryonic development but are not affected by mutations that block apoptosis. Moreover, the mechanism by which caspase 8 prevents receptor-interacting protein 1 (RIP1)- and RIP3-mediated necroptosis is now an area of immense interest.

Susan Taylor. Genomic science has transformed the way we think about biology and provides us with a new paradigm for asking biological questions and for thinking about evolution. Sequencing technology has advanced at an extraordinary pace, as has computing, so that sequencing whole genomes is becoming rapid and inexpensive. This has changed the face of biology. The human microbiome and our dramatic co-evolution with microbes is one of the most surprising discoveries to emerge from genome science. In parallel, and also of comparable magnitude, is the recognition of the role that small RNA molecules have and their enormous importance in regulating biology.

Claire E. Walczak. The past 10 years have been remarkable in terms of our understanding of genome organization, chromatin structure and gene expression, which provide the foundation for specifying individual cell function. This information has also provided the basis for many

genome-wide studies looking at a multitude of biological processes and disease states. Such studies have provided fundamental new insights into epigenetics, have elucidated a molecular understanding of altered gene regulation in disease, and have enabled fundamental new discoveries, such as RNAi and the existence of miRNAs in the genome.

Marino Zerial. In the past decade, we have progressively shifted our view and approaches drastically towards a genomic perspective. Today, our research of biological processes is no longer focused on single genes or proteins but tends to widen to the complexes, pathways or even systems level. Owing to this change in dimensionality, we have 'changed gear' and routinely benefit from comparing species, interrogating genomes and manipulating cells and organisms. This was unthinkable in the 1990s. For example, consider how RNAi has changed our approach to exploring the function of genes. In general, the genomic revolution has disclosed a horizon of interesting problems, such as the role of both coding and non-coding RNAs, to name one.

 *There has been increasing collaboration between different research communities both within, and outside, cell biology. Where do you think the most interesting interfaces in molecular cell biology reside, and what do you envisage the most fruitful collaborations will be in the future?*

A.A. Indeed, in this post-genomic era, the way we do research has changed dramatically. On average, papers have a more interdisciplinary and collaborative flavour, especially in combining biochemical and genomic analyses. In the future, I can foresee even more fruitful collaborations between cell and molecular biologists and bioinformaticians or even physicists. In fact, I think the next generation of scientists are already on the way who will perform both wet and dry laboratory research equally well.

E.F. I find the interface between human genetics, cell biology and the pharmaceutical and biotechnology industries to be the most exciting. The ability to rapidly sequence many human cancer samples has led to the identification of frequent mutations in particular types of cancer. The advances in small-molecule screening and design have led to fruitful collaborations between basic and pharmaceutical chemists, as they begin to design drugs that target only the mutant form of the protein and not its wild-type counterpart. An example is the recent development of inhibitors against the Val600Glu mutation in BRAF, a frequent mutation in melanomas[4,5]. While tumour resistance still makes eradication problematic, application of this approach to tumour resistance mutations should lead to drugs that can overcome the tumour cell resistance. With a bit of vision towards the future, we can begin to imagine drug cocktails that will make many more cancers treatable.

T.M. The development of new microscopy technology is currently extremely exciting, and has been for two decades or more. This includes new instrumentation, which typically involves

physicists, and new probes, which often involves chemists. Super-resolution is one exciting direction (photo-activated localization microscopy (PALM), stochastic optical reconstruction microscopy (STORM) and stimulated emission depletion (STED))[6–8]. Activity biosensors is another[2,3]. Intravital imaging in mice and humans is yet another.

The success of mathematical modelling has been more mixed. I am optimistic that it will help us truly understand collective protein behaviour in the future but, so far, I think the impact has been modest. One big problem is groups saying, "Our model works, therefore we have solved the problem". This is rarely true, and questionable assumptions are often hidden. But we do know that human intuition alone cannot explain collective protein behaviour in complex systems, and there seems to be no alternative to formal modelling. But, we do have to get it better integrated and be more critical.

Finally, DNA sequencing is getting cheaper by the day, and this will have a huge impact. If you can apply this resource to your question, you will make rapid progress. It will also greatly enable work on non-traditional organisms, which opens up all of biology for molecular cell biology approaches.

R.J.S. There have been incredible breakthroughs by technology-driven laboratories with expertise in physics, microscopy and mass spectrometry, which have revolutionized the ability to detect and quantify protein abundance, modifications and interactions, as well as the concentrations of intracellular metabolites that would have been unimaginable 10 years ago. This has also been coupled with advances in RNAi technology, DNA sequencing, ChIP–seq (chromatin immunoprecipitation followed by sequencing) and other techniques that bring high-throughput approaches to cell biology laboratories. Collaborations among adept practitioners using these disparate technologies to decode tissue-specific regulators and rate-limiting pathway components will uncover much fundamental cell biology as well as new targets for many different disease states.

D. St J. Cell biological research is becoming increasingly quantitative, and this has stimulated exciting collaborations with mathematicians and physicists in diverse areas, for example to measure forces in biological systems, to model morphogenetic events and to automate image analysis. Input from the physical sciences has also played an important role in the development of super-resolution microscopy, which has the potential to revolutionize cell biology if and when it can be improved to image faster and deeper inside cells.

A.S. Interactions between bioinformaticians and cellular and molecular biologists have been highly productive, for example by allowing one to make sense out of large data sets, such as gene expression profiles. Interactions between structural biologists, medicinal chemists and cell biologists have allowed us to define complex interactions of proteins in cell signalling, such as the functions of the B cell lymphoma 2 (BCL-2) family members in apoptosis. Importantly, such

interdisciplinary interactions have facilitated the development of small molecules to manipulate these processes in a therapeutic setting, such as the treatment of certain cancers.

S.T. Building bridges between computational science and experimental biology is one of our great opportunities and also one of our greatest challenges. There are enormous opportunities for bridging biology with theory and computer science, as well as with translational medicine. Advances in computing have allowed us to gather enormous amounts of data; however, the relevance of the data is compromised without a mechanistic understanding of biological systems. It is absolutely essential to bring these communities together so that we find ways to truly speak a common language. Only in this way can we achieve a comprehensive view of biological systems.

C.E.W. Even only 10 years ago, cell biology was largely a 'fuzzy' science, in which we looked at pretty pictures and described what we saw. Quantitative approaches, aided by advances in imaging, have transformed the field to a more quantitative science. The development of mathematical models for biological processes has grown increasingly more complex, offering new insight into protein function. In the future, we need to increase collaboration with chemists and physicists utilizing nanotechnology to visualize and perturb proteins on smaller and smaller scales. Computer scientists are essential to help us organize, process and analyse the large datasets being generated by high-throughput methods.

M.Z. Biological research today makes use of more quantitative approaches than before. For example, imaging and image analysis can be very quantitative, sensitive and precise, and thus are tremendously powerful for exploring biological mechanisms in time and space. This makes the collaboration with theoreticians particularly productive. Biologists need to work with mathematicians, physicists and engineers interested in biological problems because they can help us to understand mechanisms in a more precise and predictive fashion. Previously, our problem was the ability to identify some components that could give us clues into molecular mechanisms. Now that we can get to such components relatively easily (for example, with 'omics'), our problem becomes how to understand the ways in which the structure and functional properties of biological systems emerge from the interplay of individual components. For this, we need the support of theory.

A.A. The dynamics and quantitative nature of how various pathways and macromolecular complexes function remain poorly understood. We are also beginning to appreciate that spatial and temporal control contribute important regulatory steps in gene regulation. The same molecule in different cellular compartments may have very different regulatory functions, which could be missed during biochemical analyses. If we can gear our research to go from qualitative to quantitative biology and understand the real dynamics of our favourite molecules *in vivo*, we will make a great leap in our understanding of various cellular pathways.

E.F. The most pressing questions in my field are in many ways no different than they were 20–30 years ago, but the answers are closer at hand. How do stem cells build tissues during normal homeostasis and wound repair, and how does this go awry in human diseases, including cancers? And how can we exploit this information to understand the bases of these different diseases and develop new and improved therapies for the treatment of these disorders? With the recombinant DNA technology revolution of the early 1980s and the human genome revolution at the turn of the century, the interface between basic science and medicine is closing at a pace we never imagined possible as students. The tools and technologies available to address fundamental biological questions are advancing at a ferocious rate. The challenge ahead will be to ask the right questions and creatively develop strategies that exploit these tools to bridge this gap and revolutionize medicine.

R.J.S. A big challenge going forward comes out of this explosion of data from different systems: bridging the omics studies (RNAi screens, ChIP–seq, phosphoproteomes and mass spectrometry interactomes) to define what the key rate-limiting proteins in any biological process are. The world still needs careful mechanistic dissection of individual proteins and functions, which sometimes gets lost amidst the push for larger and larger datasets. Taking the findings in cellular systems and then bridging that to the physiology and pathology of diseases in the intact higher organism also remains a key challenge.

D. St J. Most recent cell biology has focused on a relatively small number of cell types (most often, unpolarized, transformed tissue culture cells) and has largely overlooked the astonishing array of different cell types with specialized functions that occur *in vivo*. I think that one of the key challenges for the future is to develop better ways of performing*in vivo* cell biology to examine cellular behaviours in the context of organs and tissues. The ability to induce iPS cells to form organs in culture will be an enormous help for this type of work.

A.S. One challenge is elucidating the precise definition of how cellular differentiation and functional activation are controlled; that is, how the many transcriptional regulators, modifications to the genome (for example, through methylation) and posttranscriptional regulatory processes (for example, through the impact of miRNAs) interact to regulate stepwise changes towards a differentiated state. Another is defining the mechanisms that regulate non-apoptotic, but still genetically programmed, cell death pathways and the definition of their role in normal physiology (for example, during embryonic development and tissue homeostasis in adulthood).

S.T. The biggest challenge for biology is always to ask the right question, and this is even more important now as technologies advance so rapidly. In our frenzy to collect more and more data, we need to learn how to ask the right questions and how to extract useful information from that data. In parallel with systems biology, we must have a mechanistic understanding of biology. Without understanding the underlying biochemical principles, the data mean little. Just as we

need classical physiology to understand how molecules work in whole animals, we need biochemistry to have a true mechanistic understanding of biological events.

C.E.W. While the genomic revolution has provided us with a wealth of potentially important molecules, the large-scale functional genomics screens only scratch the surface of understanding the mechanisms by which these proteins act. The challenge is to develop creative approaches to answer the most fundamental biological questions. For example, although proteomic approaches have identified all of the components of the mitotic spindle and genome-wide screens have identified an array of molecules that affect the mitotic spindle, we still do not understand the fundamental mechanism by which each chromosome moves to the spindle equator and then is partitioned to the daughter cells.

M.Z. Cell biology must move to tissues and organisms. An outstanding problem is bridging between scales. Understanding how cellular components form complexes, how these assemble into organelles and how organelles form cells, which build organs and organisms, poses enormous technical and conceptual challenges. The integration of biological processes is one of the most difficult problems we face. Solving these problems requires trespassing across the traditional borders between fields and developing new experimental and analytical methods. At present, we can explain only small parts of biological mechanisms: we see a few pieces of a puzzle, but for the whole picture we must draw in complexity. There are no current solutions at the modelling or computational level. This problem requires the development of new theories.

What would you consider to be the main bottlenecks for the productivity of your research and what advice would you give to young researchers facing these challenges?

A.A. Despite technical advances, such as high-throughput sequencing strategies, which have been tremendous in providing us with a global view relatively rapidly, the real bottleneck remains the in-depth analysis of data from these strategies and how to make biological sense out of such information. I also think the challenge remains to understand how various biological pathways work mechanistically and, even more importantly, how various pathways are interconnected. These are challenges for all generations of scientists. For young laboratories, one of the major challenges is to hire the right group of people and to focus on a particular biological question. My advice would be to use a multidisciplinary approach to address the questions of interest, as this provides one with the possibility to look at the question from different angles and may reveal unexpected and exciting findings.

E.F. In the United States, the main bottleneck is the precarious funding climate we face and the diminished emphasis our country places on higher education. Our country has spent decades investing in biomedical research and we are now poised to capitalize on this foundation and make major breakthroughs in the coming decades. It is paramount that we work harder to educate policy makers and the public about the time it takes to translate scientific discoveries

into cures. I am optimistic that we can do so, and I would encourage young researchers to be optimistic as well, to pursue their passion for science but also to become involved in efforts to communicate with policy makers and the public who hold the strings to our future.

T.M. Funding is a major bottleneck everywhere, and it particularly affects young scientists. One huge challenge is proving our worth to society, which we must do if we expect to be funded by taxpayers' money. The huge progress in molecular cell biology in the past two decades has not translated into a whole lot of improvement in the human condition — for example, new ways of preventing and treating disease. I believe basic science is having, and will have, that impact, but over long time periods and, often, in unpredictable ways. Solving this problem requires that some bright young people (but not all — basic progress is as important as ever) take the road of: first, translating basic research into useful applications, which in my mind includes synthetic biology as well as more conventional ideas like drugs and stem cell-based organ replacements; and second, educating and energizing the public, which includes innovating at all stages of conventional education as well as public outreach, political action and internet-based science activity.

My advice to young scientists is to avoid the road well travelled. There is no surefire route to success in sciences. But if you take the common road of doing what your advisor and her colleagues already do, you are guaranteed to end up in a crowded field in which it is hard to compete for resources and gain independent recognition. You need to take risks — in approach, in system and in questions.

R.J.S. Ironically, in the face of the kinds of technologies that have been developed for cost-effective RNAi screens and full proteomic mass spectrometry analyses of all cellular proteins and metabolites, one unmet need that is still rate-limiting in nearly every area of cell biology research is having tools and reagents that can selectively visualize pools of a given protein with specific post-translational modifications (for example, acetylation, sumoylation and phosphorylation) in intact cells and organisms. My advice to young scientists would be to explore areas at those interfaces between fields and always be incorporating new techniques and ways of thinking about a biological problem.

D. St J. Apart from the usual ones of too much bureaucracy and too little funding, one of the major difficulties we face in the laboratory is how to examine the functions of proteins that have multiple roles in the same cell lineage. Although there are a few specific tricks that allow one to knock out the function of a protein at a specific place or time, most of these actually knock out the gene and not the protein, leaving the problem of perdurance. It would be a huge help to have a standard way of engineering conditional mutant forms of proteins that are either light or heat sensitive.

My advice to young scientists is to not assume that everything that has been published is correct and that it is better to tackle interesting questions that are hard than minor questions that are easy.

A.S. First, I believe that it is important to work on a problem that you are passionate about (that is, for which you really, really want to be the first to know the answers). Second, it is a great joy to work in an environment that is highly collaborative and collegial. Nobody can cover all areas of expertise that are necessary to tackle 'big issues'. Therefore, having ready access, as a cellular or molecular biologist, to structural biologists and bioinformaticians who are eager to collaborate is a great boost for your ability to answer important questions. Finally, when choosing your postdoctoral position, it is in my opinion a good idea to join a group that desperately needs your expertise and has projects and/or techniques on offer that you want to learn. This will create a mutually beneficial or 'symbiotic' relationship, whereas joining a research programme that already has all the expertise that you can offer is like 'shipping coals to Newcastle'.

S.T. Formulate your questions carefully and then delve deeply into your system. Do not be afraid of collaborations and of learning new technologies and new systems. Do not fear reaching out. A major bottleneck for all biological research in the United States now is funding of RO1 investigator-initiated grants. The US National Institutes of Health have nurtured an explosion of biological discoveries over the past few decades, and these discoveries are having an enormous, and often unanticipated, impact on our understanding of disease. We have, unquestionably, been the dominant player in the international community. Other countries, however, are now outpacing us in their funding and we will have trouble maintaining our eminence and dominance.

C.E.W. Carrying out studies that will have a sustained impact requires an ever-increasing amount of multidisciplinary resources, including people, expertise, equipment and funding. Our educational system is not keeping pace with the rapid development of new technologies and still maintains fairly traditional disciplines. This makes it challenging to recruit young scientists who are willing to break new ground to make the exciting new discoveries. My advice to young researchers is to take every opportunity available to learn and discover, and never forget to take time to just think and reflect about what is interesting and cool.

M.Z. My strong advice to young researchers is to think in a truly multidisciplinary way and to go to institutes which can support this approach. Addressing a problem from different sides is essential. Good funding is not everything: I would also encourage young scientists to choose institutes where they can get good mentoring and be stimulated and challenged by faculty members who demonstrate true interest in their work. Importantly, treasure the value of central facilities, available to everyone and capable of supporting research beyond what a single laboratory can do. This kind of support raises the level of ambition and productivity of a young starting group more than any seemingly rich 'start-up package'.

Chapter 2. DNA Structure

4. Nucleic Acids

Nucleic acids are biopolymers, or large biomolecules, essential for all known forms of life. Nucleic acids, which include DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), are made from monomers known as nucleotides. Each nucleotide has three components: a 5-carbon sugar, a phosphate group, and a nitrogenous base. If the sugar is deoxyribose, the polymer is DNA. If the sugar is ribose, the polymer is RNA. When all three components are combined, they form a nucleotide. Nucleotides are also known as phosphate nucleotides.

Nucleic acids are among the most important biological macromolecules (others being amino acids/proteins, sugars/carbohydrates, and lipids/fats). They are found in abundance in all living things, where they function in encoding, transmitting and expressing genetic information in other words, information is conveyed through the nucleic acid sequence, or the order of nucleotides within a DNA or RNA molecule. Strings of nucleotides strung together in a specific sequence are the mechanism for storing and transmitting hereditary, or genetic information via protein synthesis.

Nucleic acids were discovered by Friedrich Miescher in 1869. Experimental studies of nucleic acids constitute a major part of modern biological and medical research, and form a foundation for genome and forensic science, as well as the biotechnology and pharmaceutical industries.

Deoxyribonucleic acid

Deoxyribonucleic acid (DNA) is a nucleic acid containing the genetic instructions used in the development and functioning of all known living organisms. The DNA segments carrying this genetic information are called genes. Likewise, other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information. Along with RNA and proteins, DNA is one of the three major macromolecules that are essential for all known forms of life. DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are, therefore, anti-parallel. Attached to each sugar is one of four types of molecules called nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription. Within cells DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as

mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

Ribonucleic acid

Ribonucleic acid (RNA) functions in converting genetic information from genes into the amino acid sequences of proteins. The three universal types of RNA include transfer RNA (tRNA), messenger RNA (mRNA), and ribosomal RNA (rRNA).Messenger RNA acts to carry genetic sequence information between DNA and ribosomes, directing protein synthesis.Ribosomal RNA is a major component of the ribosome, and catalyzes peptide bond formation. Transfer RNA serves as the carrier molecule for amino acids to be used in protein synthesis, and is responsible for decoding the mRNA. In addition, many other classes of RNA are now known.

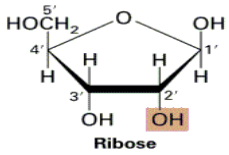
Artificial nucleic acid analogs

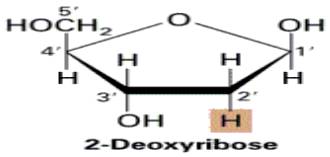
Artificial nucleic acid analogues have been designed and synthesized by chemists, and include peptide nucleic acid,morpholino- and locked nucleic acid, as well as glycol nucleic acid and threose nucleic acid. Each of these is distinguished from naturally occurring DNA or RNA by changes to the backbone of the molecule.

5. Chemical composition of DNA

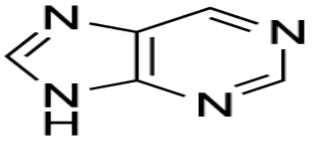
The chemical DNA was first discovered in 1869, but its role in genetic inheritance was not demonstrated until 1943. In 1953James Watson and Francis Crick determined that the structure of DNA is a double-helix polymer, a spiral consisting of two DNA strands wound around each other. Each strand is composed of a long chain of monomer nucleotides. Thenucleotide of DNA consists of a deoxyribose sugar moleculeto which is attached a phosphate group and one of four nitrogenous bases: two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). The nucleotides are joined together by covalent bonds between the phosphate of one nucleotide and the sugar of the next, forming a phosphate-sugar backbone from which the nitrogenous bases protrude. One strand is held to another by hydrogen bonds between the bases; the sequencing of this bonding is specific—i.e., adenine bonds only with thymine, and cytosineonly with guanine.

Deoxyribose (a pentose sugar derivative)

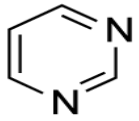




Nitrogenous Bases

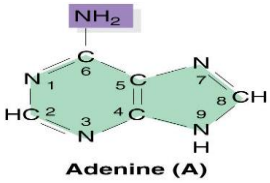


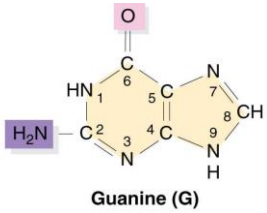
purine



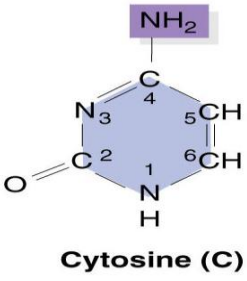
pyrimidine

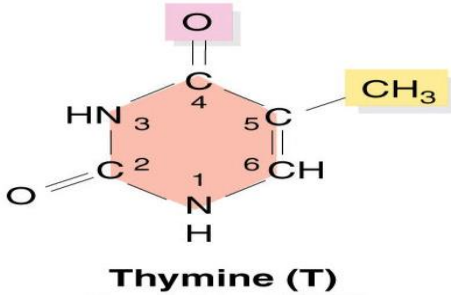
Purines



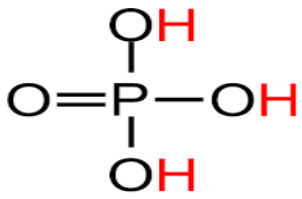


Pyrimidines





Phosphoric acid



H_3PO_4

6. Nucleoside & Nucleotide

A nucleoside consists of a nitrogenous base covalently attached to a sugar (ribose or deoxyribose) but without the phosphate group. A nucleotide consists of a nitrogenous base, a sugar (ribose or deoxyribose) and one to three phosphate groups.

Nucleoside=Sugar+Base

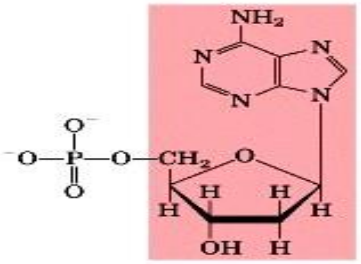
Nucleotide = Sugar + Base + Phosphate

Comparison chart

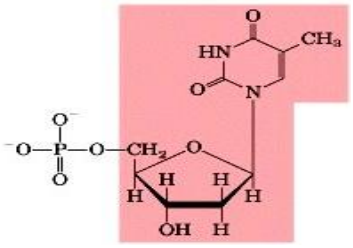
	Nucleoside	Nucleotide
Relevance in medicine	Several nucleoside analogues are used as antiviral or anticancer agents.	Malfunctioning nucleotides are one of the main causes of all cancers known of today.
Examples	Examples of nucleosides include cytidine, uridine, adenosine, guanosine, thymidine and inosine.	Nucleotides follow the same names as nucleosides, but with the indication of phosphate groups. For example, 5'-uridine monophosphate.
Chemical Composition	Sugar + Base. A nucleoside consists of a nitrogenous base covalently attached to a sugar (ribose or deoxyribose) but without the phosphate group. When phosphate group of nucleotide is removed by hydrolysis, the structure remaining is nucleoside.	Sugar + Base + Phosphate. A nucleotide consists of a nitrogenous base, a sugar (ribose or deoxyribose) and one to three phosphate groups.

7. Types of Deoxyribonucleotides

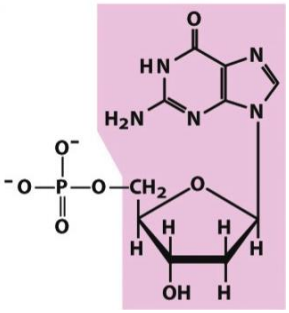
There are four types of Deoxy-ribonucleotides with respect to Nitrogen bases



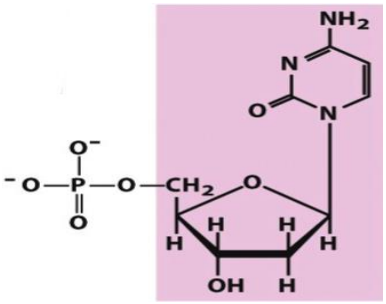
Deoxyadenylate



Deoxythymidylate



Deoxyadenylate



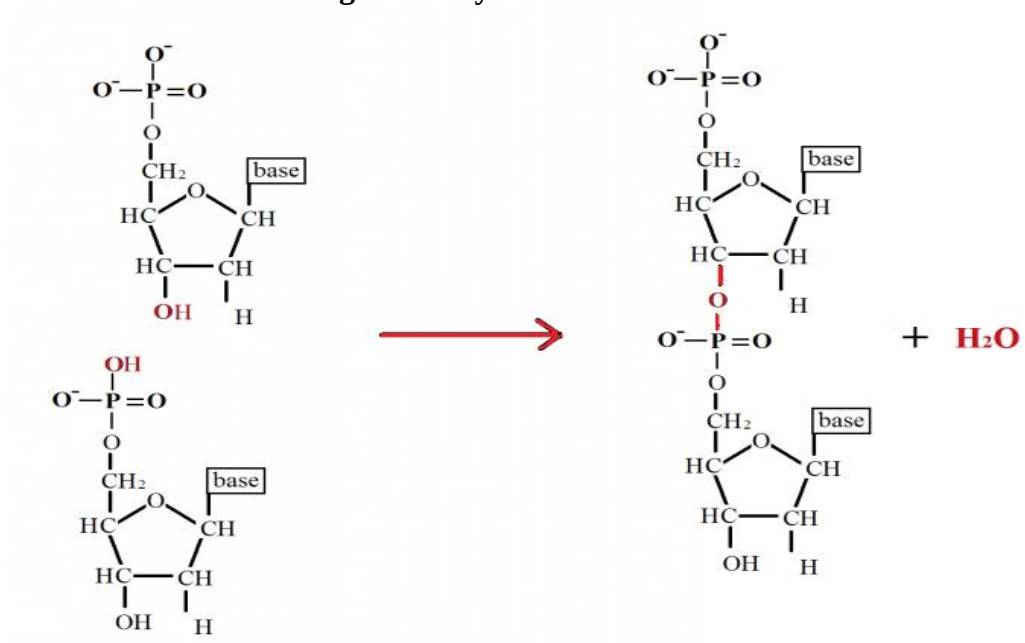
Deoxythymidylate

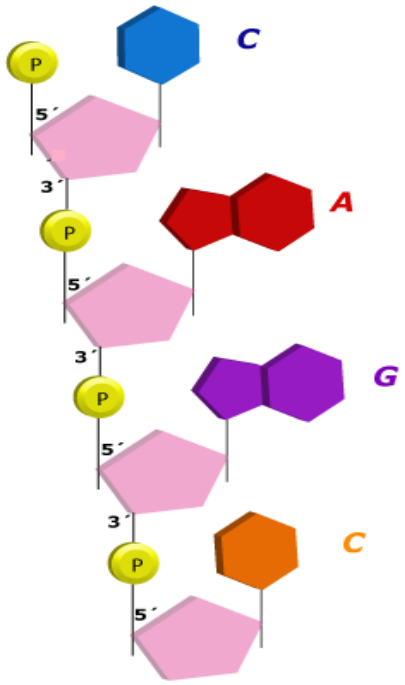
8. How do Deoxyribonucleotides Join?

A deoxyribonucleotide is the monomer, or single unit, of DNA, or deoxyribonucleic acid. Each deoxyribonucleotide comprises three parts: a nitrogenous base, a deoxyribose sugar, and one phosphate group. The nitrogenous base is always bonded to the 1' carbon of the deoxyribose, which is distinguished from ribose by the presence of a proton on the 2' carbon rather than an OH group. The phosphate groups bind to the 5' carbon of the sugar.

When deoxyribonucleotides polymerize to form DNA, the phosphate group from one nucleotide will bond to the 3' carbon on another nucleotide, forming a phosphodiester bond via dehydration synthesis. New nucleotides are always added to the 3' carbon of the last nucleotide, so synthesis always proceeds from 5' to 3'.

Joining of Deoxyribonucleotides





9. Structure of DNA

Miescher ’s DNA structure

Although few people realize it, 1869 was a landmark year in genetic research, because it was the year in which Swiss physiological chemist Friedrich Miescher first identified what he called "nuclein" inside the nuclei of human white blood cells. (The term "nuclein" was later changed to "nucleic acid" and eventually to "deoxyribonucleic acid," or "DNA.") Miescher's plan was to isolate and characterize not the nuclein (which nobody at that time realized existed) but instead the protein components of leukocytes (white blood cells). Miescher thus made arrangements for a local surgical clinic to send him used, pus-coated patient bandages; once he received the bandages, he planned to wash them, filter out the leukocytes, and extract and identify the various proteins within the white blood cells. But when he came across a substance from the cell nuclei that had chemical properties unlike any protein, including a much higher phosphorous content and resistance to proteolysis (protein digestion), Miescher realized that he had discovered a new substance (Dahm, 2008). Sensing the importance of his findings, Miescher wrote, "It seems probable to me that a whole family of such slightly varying phosphorous-containing substances will appear, as a group of nucleins, equivalent to proteins" (Wolf, 2003).

More than 50 years passed before the significance of Miescher's discovery of nucleic acids was widely appreciated by the scientific community. For instance, in a 1971 essay on the history of nucleic acid research, Erwin Chargaff noted that in a 1961 historical account of nineteenth-century science, Charles Darwin was mentioned 31 times, Thomas Huxley 14 times, but Miescher not even once. This omission is all the more remarkable given that, as Chargaff also noted, Miescher's discovery of nucleic acids was unique among the discoveries of the four major cellular components (i.e., proteins, lipids, polysaccharides, and nucleic acids) in that it could be "dated precisely... [to] one man, one place, one date."

Levene Investigates the Structure of DNA

Meanwhile, even as Miescher's name fell into obscurity by the twentieth century, other scientists continued to investigate the chemical nature of the molecule formerly known as nuclein. One of these other scientists was Russian biochemist Phoebus Levene. A physician turned chemist, Levene was a prolific researcher, publishing more than 700 papers on the chemistry of biological molecules over the course of his career. Levene is credited with many firsts. For instance, he was the first to discover the order of the three major components of a single nucleotide (phosphate-sugar-base); the first to discover the carbohydrate component of RNA (ribose); the first to discover the carbohydrate component of DNA (deoxyribose); and the first to correctly identify the way RNA and DNA molecules are put together.

During the early years of Levene's career, neither Levene nor any other scientist of the time knew how the individual nucleotide components of DNA were arranged in space; discovery of the sugar-phosphate backbone of the DNA molecule was still years away. The large number of molecular groups made available for binding by each nucleotide component meant that there were numerous alternate ways that the components could combine. Several scientists put forth suggestions for how this might occur, but it was Levene's "polynucleotide" model that proved to

be the correct one. Based upon years of work using hydrolysis to break down and analyze yeast nucleic acids, Levene proposed that nucleic acids were composed of a series of nucleotides, and that each nucleotide was in turn composed of just one of four nitrogen-containing bases, a sugar molecule, and a phosphate group. Levene made his initial proposal in 1919, discrediting other suggestions that had been put forth about the structure of nucleic acids. In Levene's own words, "New facts and new evidence may cause its alteration, but there is no doubt as to the polynucleotide structure of the yeast nucleic acid" (1919).

Indeed, many new facts and much new evidence soon emerged and caused alterations to Levene's proposal. One key discovery during this period involved the way in which nucleotides are ordered. Levene proposed what he called a tetranucleotide structure, in which the nucleotides were always linked in the same order (i.e., G-C-T-A-G-C-T-A and so on). However, scientists eventually realized that Levene's proposed tetranucleotide structure was overly simplistic and that the order of nucleotides along a stretch of DNA (or RNA) is, in fact, highly variable. Despite this realization, Levene's proposed polynucleotide structure was accurate in many regards. For example, we now know that DNA is in fact composed of a series of nucleotides and that each nucleotide has three components: a phosphate group; either a ribose (in the case of RNA) or a deoxyribose (in the case of DNA) sugar; and a single nitrogen-containing base. We also know that there are two basic categories of nitrogenous bases: the purines (adenine [A] and guanine [G]), each with two fused rings, and the pyrimidines (cytosine [C], thymine [T], and uracil [U]), each with a single ring. Furthermore, it is now widely accepted that RNA contains only A, G, C, and U (no T), whereas DNA contains only A, G, C, and T (no U) (Figure 1).

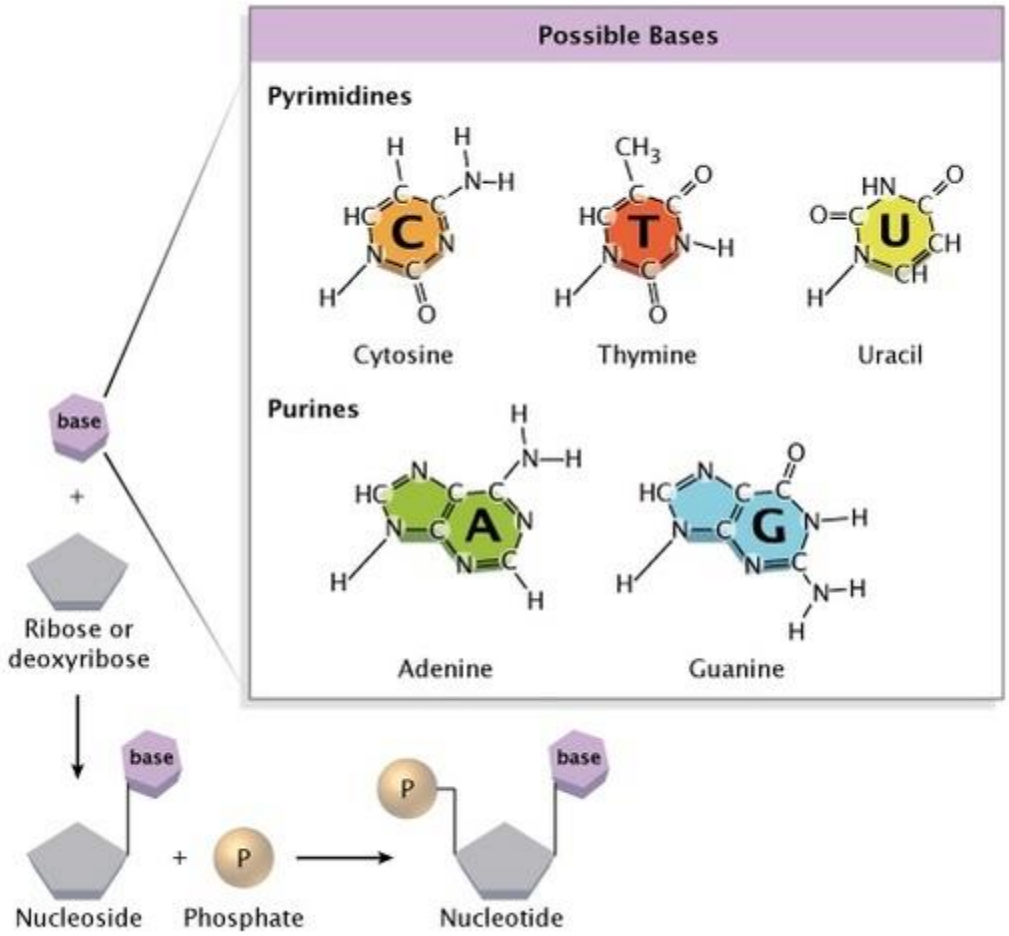


Figure : The chemical structure of a nucleotide.

Chargaff Rules for DNA Structure

Erwin Chargaff was one of a handful of scientists who expanded on Levene's work by uncovering additional details of the structure of DNA, thus further paving the way for Watson and Crick. Chargaff, an Austrian biochemist, had read the famous 1944 paper by Oswald Avery and his colleagues at Rockefeller University, which demonstrated that hereditary units, or genes, are composed of DNA. This paper had a profound impact on Chargaff, inspiring him to launch a research program that revolved around the chemistry of nucleic acids. Of Avery's work, Chargaff (1971) wrote the following:

"This discovery, almost abruptly, appeared to foreshadow a chemistry of heredity and, moreover, made probable the nucleic acid character of the gene... *Avery gave us the first text of a new language, or rather he showed us where to look for it. I resolved to search for this text."*

As his first step in this search, Chargaff set out to see whether there were any differences in DNA among different species. After developing a new paper chromatography method for separating and identifying small amounts of organic material, Chargaff reached two major conclusions (Chargaff, 1950). First, he noted that the nucleotide composition of DNA varies among species. In other words, the same nucleotides do not repeat in the same order, as proposed by Levene. Second, Chargaff concluded that almost all DNA--no matter what organism or tissue

type it comes from--maintains certain properties, even as its composition varies. In particular, the amount of adenine (A) is usually similar to the amount of thymine (T), and the amount of guanine (G) usually approximates the amount of cytosine (C). In other words, the total amount of purines (A + G) and the total amount of pyrimidines (C + T) are usually nearly equal. (This second major conclusion is now known as "Chargaff's rule.") Chargaff's research was vital to the later work of Watson and Crick, but Chargaff himself could not imagine the explanation of these relationships--specifically, that A bound to T and C bound to G within the molecular structure of DNA (Figure 2).

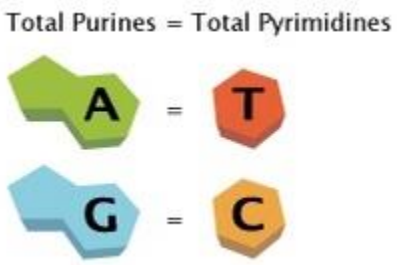


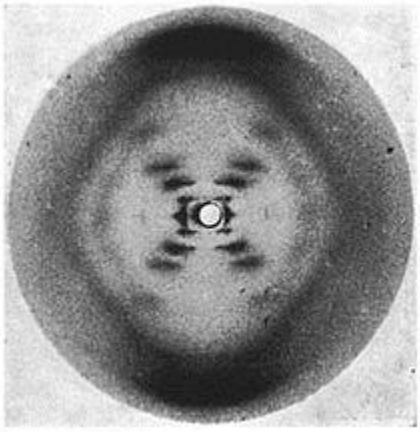
Figure : What is Chargaff's rule?

10. Wilkins and Franklin work on DNA

Maurice Wilkins and Rosalind Franklin, together with Ray Gosling, Alec Stokes and Herbert Wilson and other colleagues at the Randall Institute at King's, made crucial contributions to the discovery of DNA's structure in 1953.

Wilkins began using optical spectroscopy to study DNA in the late 1940s. In 1950 he and Gosling obtained the first clearly crystalline X-ray diffraction patterns from DNA fibres. Alec Stokes suggested that the patterns indicated that DNA was helical in structure.

The discovery of the structure of DNA in 1953 revealed the physical and chemical basis of how characteristics are passed down through the generations and how they are expressed in individual organisms.



X-ray diffraction pattern of DNA

11. Watson and Crick Model of DNA Structure

Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA. Watson and Crick's discovery was also made possible by recent advances in model building, or the assembly of possible three-dimensional structures based upon known molecular distances and bond angles, a technique advanced by American biochemist Linus Pauling. In fact, Watson and Crick were worried that they would be "scooped" by Pauling, who proposed a different model for the three-dimensional structure of DNA just months before they did. In the end, however, Pauling's prediction was incorrect.

Using cardboard cutouts representing the individual chemical components of the four bases and other nucleotide subunits, Watson and Crick shifted molecules around on their desktops, as though putting together a puzzle. They were misled for a while by an erroneous understanding of how the different elements in thymine and guanine (specifically, the carbon, nitrogen, hydrogen, and oxygen rings) were configured. Only upon the suggestion of American scientist Jerry Donohue did Watson decide to make new cardboard cutouts of the two bases, to see if perhaps a different atomic configurationwould make a difference. It did. Not only did the complementary bases now fit together perfectly (i.e., A with T and C with G), with each pair held together by hydrogen bonds, but the structure also reflected Chargaff's rule (Figure 3).

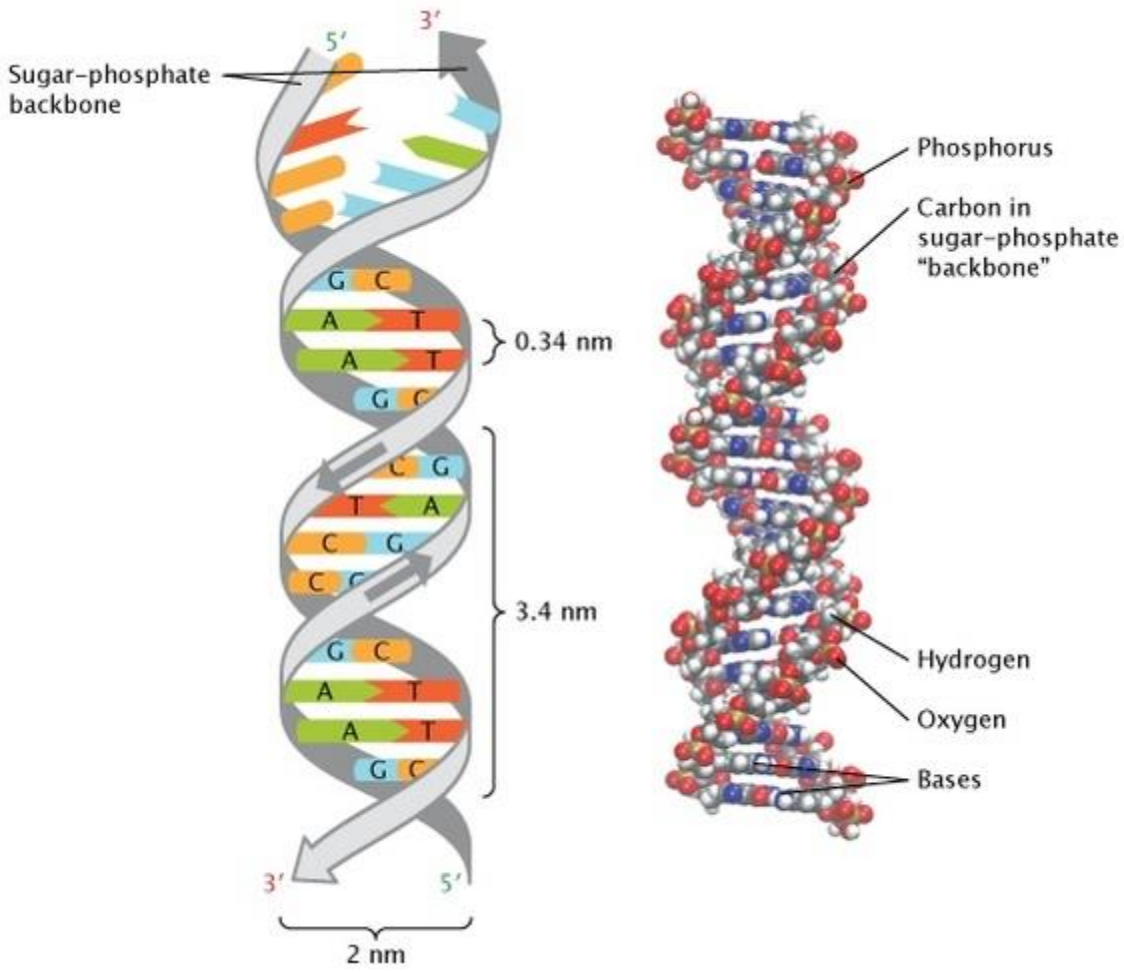


Figure : The double-helical structure of DNA.

Chapter 3. RNA Structure

12. Chemical Composition of RNA

Usually ribonucleic acid (RNA) is single stranded and made up of long, unbranched polynucleotide chain. The polynucleotide chain is formed by joining of ribonucleotides, with the help of 3' – 5' phosphodiester bonds in the same fashion as in case of DNA. But RNA is more stable than DNA because of intermolecular pairing.

Ribonucleotides = Pentose sugar (ribose) + N-base + phosphate group
Nitrogen bases are of two types

Purines

Adenine (A)
Guanine (G)

Pyrimidines

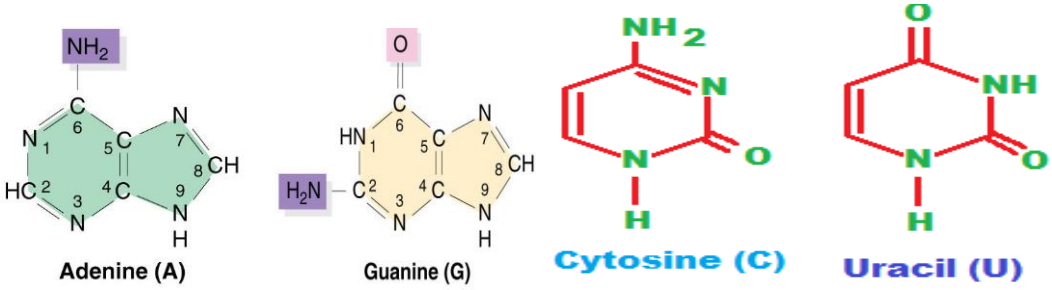
Cytosine (C)
Uracil (U)

Many RNAs possess number of minor bases in addition to above four bases, so there are unusual nucleotides like pseudouridine, inosine, dihydroxyuridine etc.

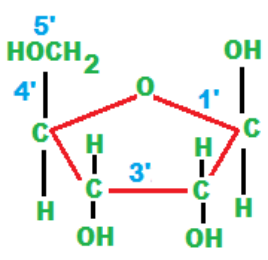
Following table summarizes N – bases, ribonucleosides and ribonucleotides.

N – base	Ribonucleoside	Ribonucleotide	Abbreviation
Adenine	Adenosine	Adenosine monophosphate (Adenylic acid)	AMP
Guanine	Guanosine	Guanosine monophosphate (Guanylic acid)	GMP
Cytosine	Cytidine	Cytidine monophosphate (Cytidylic acid)	CMP
Uracil	Uridine	Uridine monophosphate (Uridylic acid)	UMP

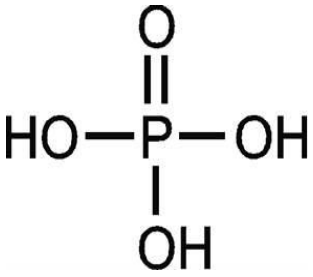
Only difference between DNA and RNA chemical composition is RNA has ribose sugar, has hydroxyl group (- OH) at the 2' position and N-base thymine is replaced by uracil.



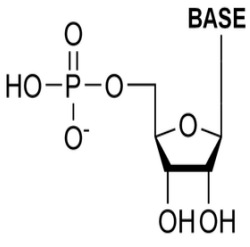
Nitrogenous Bases



Ribose (a pentose sugar)



Phosphoric acid

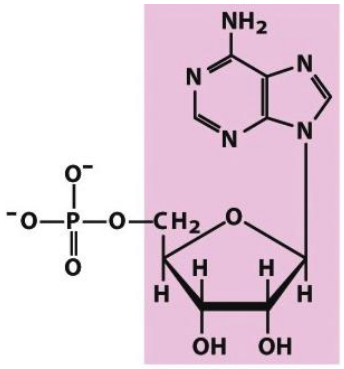


A Ribonucleotide

13. Types of Ribonucleotides

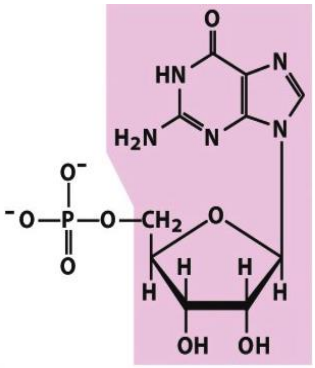
There are mainly four types of ribonucleotides depending upon the types of nitrogenous bases present in RNA.

Nucleotides	Symbols	Nucleoside
Adenylate (adenosine 5'-monophosphate)	A, AMP	Adenosine
Guanylate (guanosine 5'-monophosphate)	G, GMP	Guanosine
Uridylate (uridine 5'-monophosphate)	U, UMP	Uridine
Cytidylate (cytidine 5'-monophosphate)	C, CMP	Cytidine



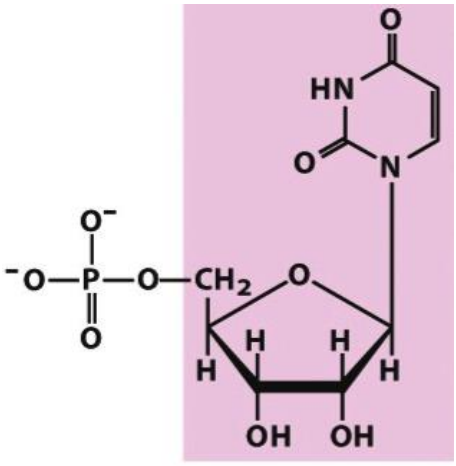
Adenylate (adenosine 5′-monophosphate)

Adenosine



Guanylate (guanosine 5′-monophosphate)

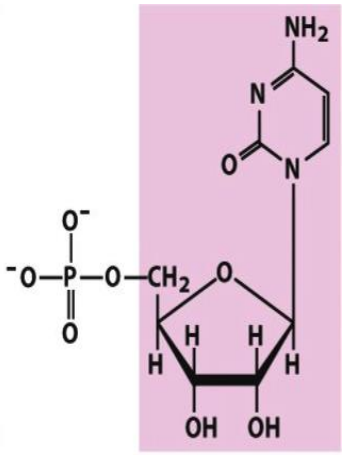
Guanosine



Uridylate (uridine 5′-monophosphate)

U, UMP

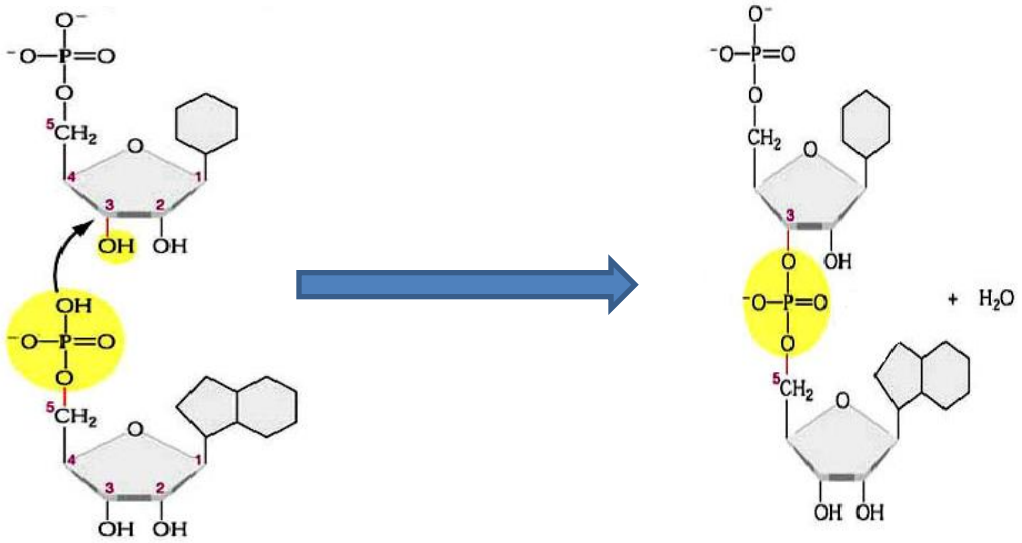
Uridine



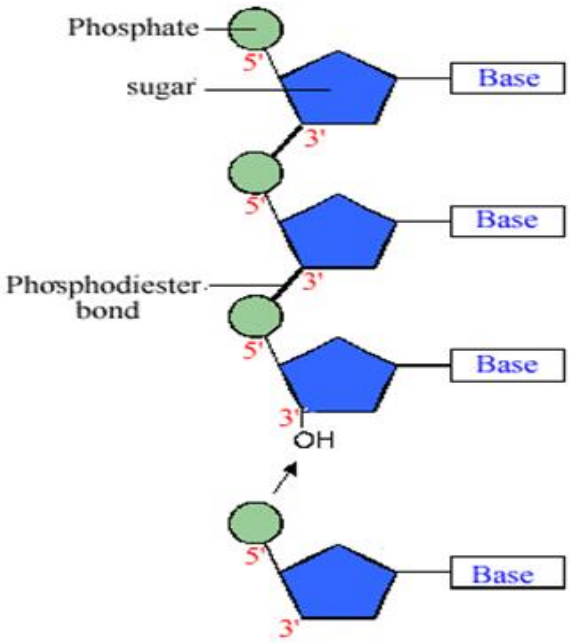
Cytidylate (cytidine 5′-monophosphate)

C, CMP

Cytidine



Joining of Ribonucleotides



A Poly-Ribonucleotide

14. Types of RNAs

There are actually several types of ribonucleic acids or RNAs, but mainly three types of Ribonucleic acids (RNAs) present in the cells of living organisms.

1. Messenger RNA (mRNA)
2. Transfer RNA (tRNA)
3. Ribosomal RNA (rRNA)

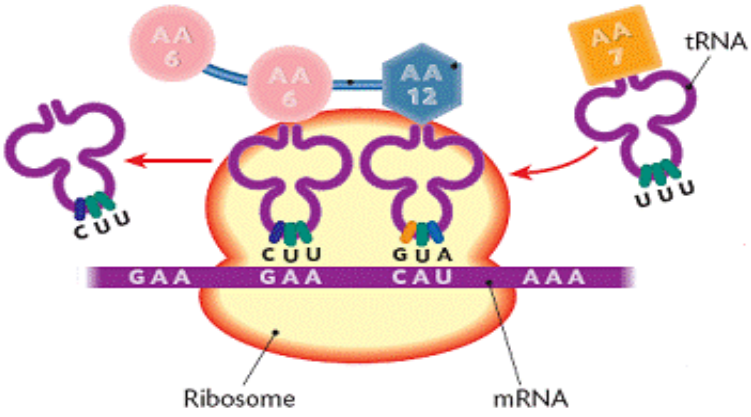
Messenger RNA (mRNA)

- It is the type of RNA that carries genetic information from DNA to the protein biosynthetic machinery of the ribosome.
- It provides the templates that specify amino acid sequences in polypeptide chains.
- The process of forming mRNA on a DNA template is known as transcription.
- It may be monocistronic or polycistronic.
- The length of mRNA molecules is variable and it depends on the length of gene.

Transfer RNA (tRNA)

- Transfer RNAs serve as adapter molecules in the process of protein synthesis.
- They are covalently linked to an amino acid at one end.

- They pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence.

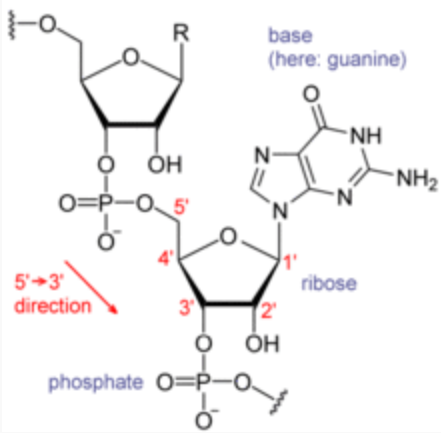


Ribosomal RNA (rRNA)

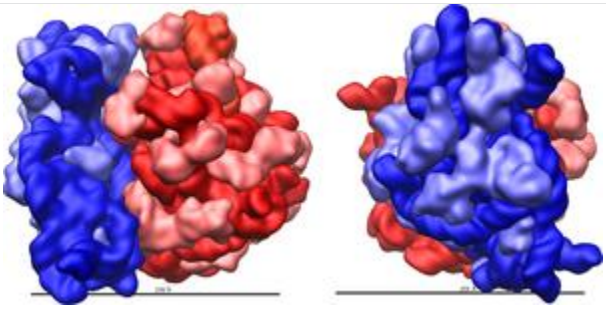
- Ribosomal RNAs are components of ribosomes.
- rRNA is a predominant material in the ribosomes constituting about 60% of its weight.
- It has a number of functions to perform in the ribosomes.

15. Structures of RNAs

Each nucleotide in RNA contains a ribose sugar, with carbons numbered 1' through 5'. A base is attached to the 1' position, in general, adenine (A), cytosine(C), guanine (G), or uracil (U). Adenine and guanine are purines, cytosine and uracil are pyrimidines. A phosphate group is attached to the 3' position of one ribose and the 5' position of the next. The phosphate groups do not have a negative charge each at physiological pH, making RNA a charged molecule (polyanion). The bases form hydrogen bonds between cytosine and guanine, between adenine and uracil and between guanine and uracil.[8] However, other interactions are possible, such as a group of adenine bases binding to each other in a bulge,[9] or the GNRA tetraloop that has a guanine–adenine base-pair.



Ribosomal RNA

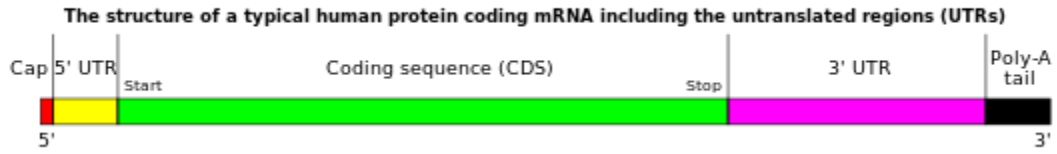


Ribosomal RNA (rRNA) is the RNA component of a ribosome

Ribosomes are non membranous organelles that participate in the translation of mRNA into a protein product. The ribosome structure is composed of 2 subunits. A small and a large subunit each of which primarily consists of rRNA of various size and a small quantity of proteins. rRNA constitutes about the 80% of the whole RNA present in an eukaryotic cell. The large subunit consists of rRNA of 5S, 5.8S and 28S sizes whereas the small subunit consists of rRNA of 18S size. (where S is the unit for rRNA size). These rRNAs are synthesized by transcription of the rRNA genes. However, the rRNA genes encode for all rRNAs except from the 5S rRNA, which is synthesized by the tRNA genes along with all the nuclear tRNAs. RNA polymerase type I is responsible for the transcription of rRNA genes by binding on the core element−CE, which overlaps the Transcription Start Site−TSS, along with the Transcription Factors inducing the so called Transcription Initiation Complex designated as TIC. The rate of the transcription is controlled by an Upstream Control Sequence−UCS located 100 base pairs upstream to the TSS. The transcription process begins and the genes are transcribed into pre−rRNA in the following order as situated on the gene: -18S - 5.8S - 28S-. The transcription comes to an end when the Transcription Complex reaches an area rich in Adenines found at about 600 base pairs downstream of the gene, indicating its end. The pre−rRNA formed includes all rRNAs on a

single strand so that cleavage has to be performed so that different size rRNAs are separated. This task is untertaken by RNases that cleave the rRNA giving rise to the differential size rRNAs.

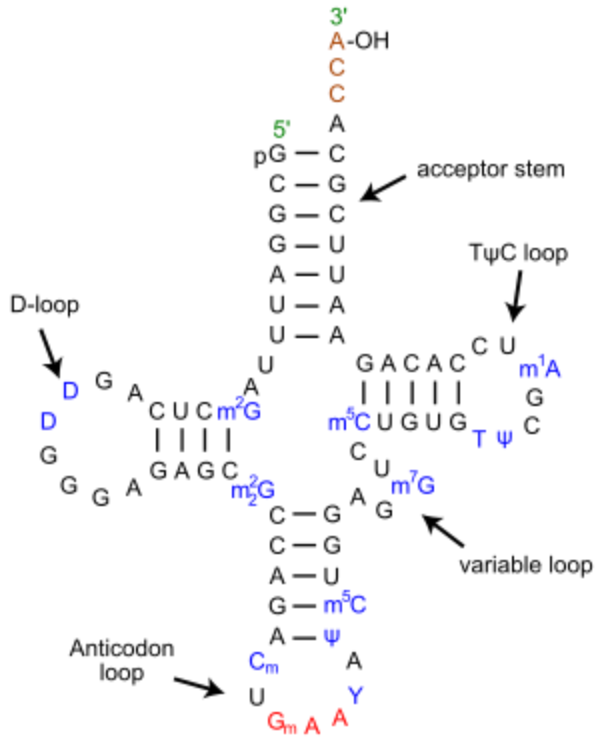
Messenger RNA



The structure of a mature eukaryotic mRNA. A fully processed mRNA includes a 5' cap, 5' UTR,coding region, 3' UTR, and poly(A) tail.

mRNA genes are the genes that encode only for proteins but this encoding has an RNA intermediate. The DNA is firstly transcribed into mRNA and subsequently translated into a protein product. So the mRNA genes are the genes that encode for mRNA in order to synthesize proteins. mRNA constitutes only the 5% of the total RNA. RNA polymerase II is the enzyme responsible for the transcription of the corresponding genes into mRNA. The polymerase binds on the TATA box which acts more or less as a promoter, located about 25 base pairs upstream the Transcription Start Site−TSS, along with the transcription factors giving rise to the Transcription Initiation Complex−TIC. In order for this complex to be functional a proper sequence of events of binding the TF and the polymerase on the promoter must occur as: TFII-D, TFII-A, TFII-B, RNA pol II, TFII-F, TFII-H TFII-E TFII-J. As soon as the TIC is formed then transcription begins giving rise to pre−mRNA which include both exons and introns. Transcription ends without recognition of an adenine rich area but rather by automatic disassembling of the Transcription Complex. The pre−mRNA is then submitted to processing that involves splicing - removal of introns and merging of the adjacent exons- and capping - addition of 7−methylguanosine on the 5' end of the mRNA so that mRNA cannot be cleaved by exonucleases. It also serves as a recognition site of the mRNA prior to translation for the small ribosomal subunit.

Transfer RNA



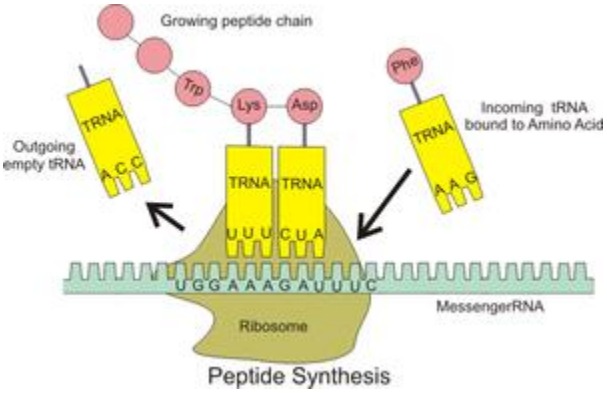
Secondary *cloverleaf structure* of $tRNA^{Phe}$ from yeast.

Transfer RNA is encoded by genes that also encode for the 5S size rRNA. RNA polymerase III is responsible for the transcription of these genes by binding on the promoter, situated about 100 base pairs downstream the Transcription Start Site -TSS, along with the Transcription Factors giving rise to the Transcription Initiation Complex. As soon as this complex is formed transcription process can begin and when the Transcription Complex faces an Adenine rich region transcription comes to an end as this area is an indication for the gene end. tRNA constitutes 15% of the total RNA and is directly involved in the translation of the mRNA. More specificaly tRNA binds onto a specific amino acid and brings it along the translation site so that it is bound on the newly synthesized peptide.

- tRNA binds to its specific amino acid recognized by its side R chain in presence of the aminoacyl tRNA synthetase enzyme. The synthetase binds the 5'-CCA-OH-3' acceptor arm with the —COOH group of the amino acid.
- When the small ribosomal subunit faces an AUG codon on the mRNA it indicates the commencing of the peptide formation. As soon as the AUG codon is recognized then the first tRNA binds on the small ribosomal subunit and on the mRNA through its anticodon arm, giving rise to the Translation Initiation Complex designated as $tRNA_i^{met}$. Eventually the large ribosomal

subunit binds on the complex indicating the initiation of the translation process. Translation always begins with the methionine amino acid on the newly synthesized peptide.

- After translocation of the Translation complex the $tRNA_i^{met}$ enter the Peptidyl site of the complex leaving the Aminoacyl site vacant for the next tRNA to enter, bringing together the two adjacent amino acids so that a peptide bond can be formed in presence of the peptidyl transferase enzyme. As soon as the peptide bond is formed, the tRNA is released from its amino acid in presence of the tRNA deacylase.
-



The interaction of tRNA and mRNA in protein synthesis

Chapter 4. DNA a genetic material

16. Nature of Genetic Material

After establishment of the fact that genes are the physical units located on the chromosomes. A major problem for the biologists was to find out the molecules responsible for carrying the hereditary information.

Characteristics of Genetic Material
Genetic material must contain complex information.
Genetic material must replicate faithfully.
Genetic material must encode phenotype.

Three sets of experiments provided a pivotal evidence that DNA rather than protein, is the hereditary material.

Griffith's Experiments (1928)
Avery's Experiments (1944)
Hershey-Chase experiments (1952)

17. Griffith's Experiments (1928)

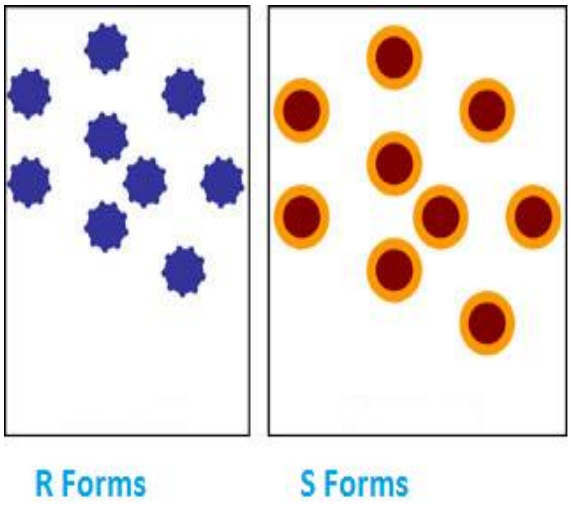
Griffith's experiment, reported in 1928 byFrederick Griffith,[1] was the first experiment suggesting that bacteria are capable of transferring genetic information through a process known as transformation. Griffith's findings were followed by research in the late 1930s and early 40s that isolated DNA as the material that communicated this genetic information.

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine. Griffith used two strains of pneumococcus (*Streptococcus pneumoniae*) bacteria which infect mice – a type III-S (smooth) which was virulent, and a type II-R (rough) strain which was nonvirulent. The III-S strain covered itself with apolysaccharide capsule that protected it from the host's immune system, resulting in the death of the host, while the II-R strain did not have that protective capsule and was defeated by the host's immune system. A German bacteriologist, Fred Neufeld, had discovered the three pneumococcal types (Types I, II, and III) and discovered the Quellung reaction to identify them in vitro. Until Griffith's experiment, bacteriologists believed that the types were fixed and unchangeable, from one generation to another.

In this experiment, bacteria from the III-S strain were killed by heat, and their remains were added to II-R strain bacteria. While neither alone harmed the mice, the combination was able to kill its host. Griffith was also able to isolate both live II-R and live III-S strains of pneumococcus

from the blood of these dead mice. Griffith concluded that the type II-R had been "transformed" into the lethal III-S strain by a "transforming principle" that was somehow part of the dead III-S strain bacteria.

Today, we know that the "transforming principle" Griffith observed was the DNA of the III-s strain bacteria. While the bacteria had been killed, the DNA had survived the heating process and was taken up by the II-R strain bacteria. The III-S strain DNA contains the genes that form the protective polysaccharide capsule. Equipped with this gene, the former II-R strain bacteria were now protected from the host's immune system and could kill the host. The exact nature of the transforming principle (DNA) was verified in the experiments done by Avery, McLeod and McCarty and by Hershey and Chase.



18. Griffith's experimental work

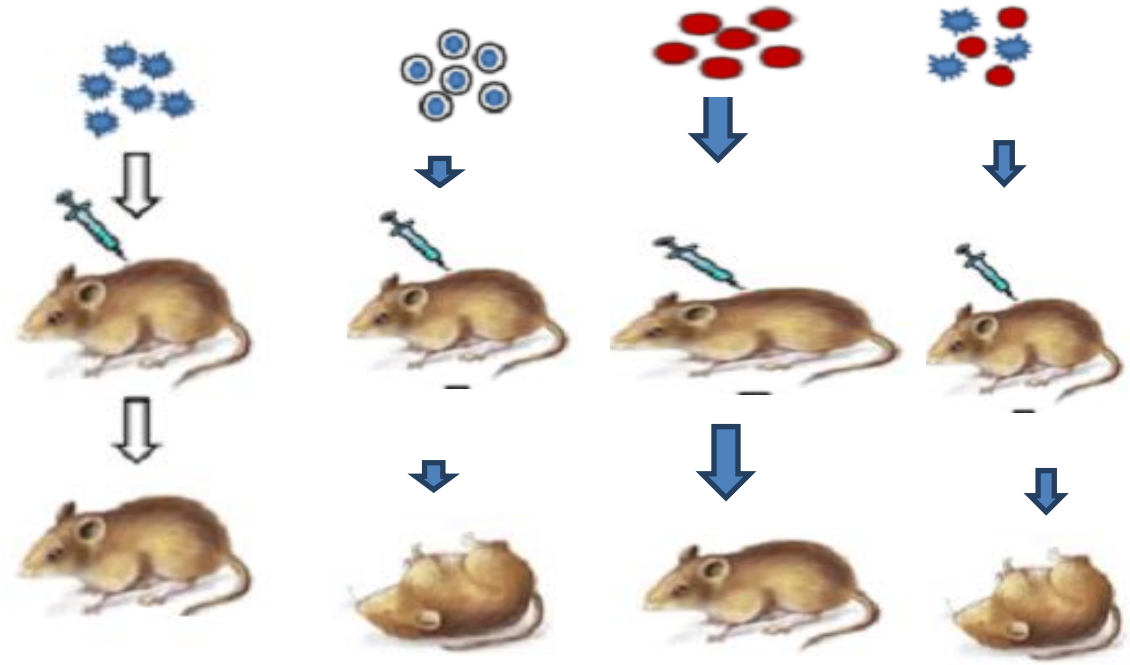


Fig. Griffith's experiment discovering the "transforming principle"

Possible Interpretations

It could have been the case that S- Type bacteria were not completely killed and a few live bacteria remained in the culture.

A second interpretation was that the live R-Type bacteria had mutated to the virulent S form.

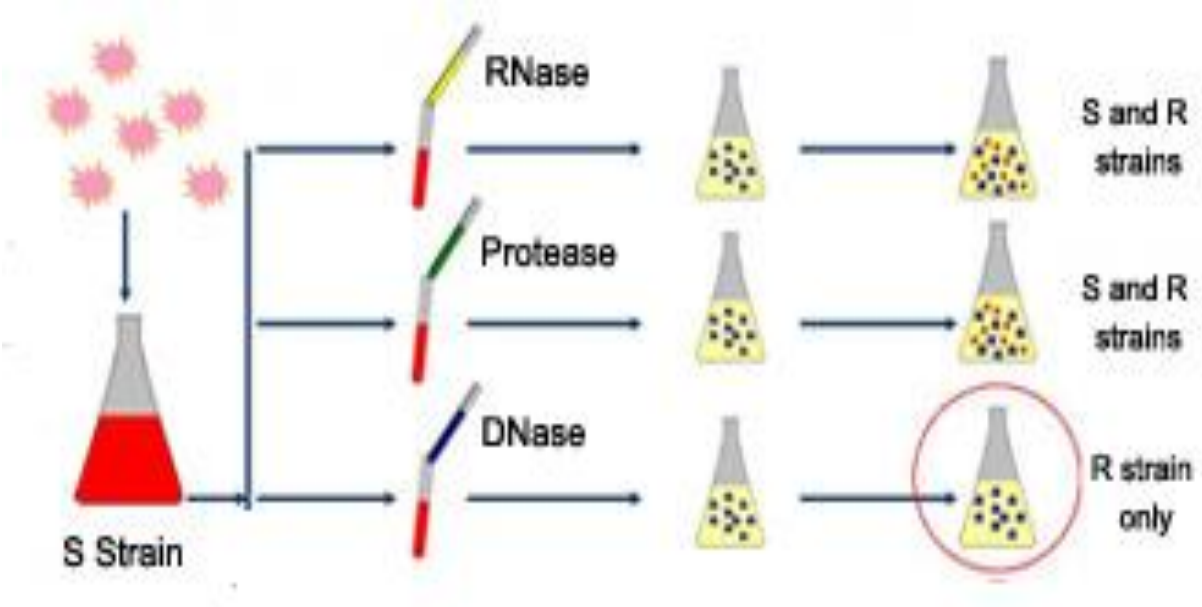
Griffith finally concluded that R-Type bacteria had been *transformed.*

Griffith theorized that some substance of the dead bacteria might be responsible for that transformation . He called that substance as the transforming principle.

19. Avery's Experiments (1944)

In 1944, experiments by Oswald T. Avery showed that DNA is the substance that causes bacterial transformation, in an era when it had been widely believed that it was proteins that served the function of carrying genetic information (with the very word *protein* itself coined to indicate a belief that its function was *primary*). It was the culmination of research in the 1930s and early 1940s at the Rockefeller Institute for Medical Research to purify and characterize the "transforming principle" responsible for the transformation phenomenon first described in Griffith's experiment of 1928: killed *Streptococcus pneumoniae* of the virulent strain type III-S, when injected along with living but non-virulent type II-R

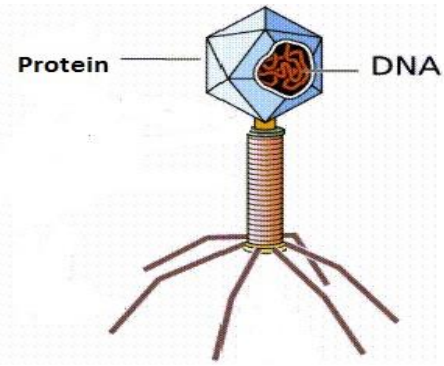
pneumococci, resulted in a deadly infection of type III-S pneumococci.Avery, MacLeod and McCarty succeeded in isolating and purifying the transforming substance in 1944. They showed that it had a chemical composition closely matching that of DNA and quite different from that of proteins. They showed that proteolytic enzymes had no effect on the transforming substance. Ribonuclease also had no effect on it. However, enzymes capable of destroying DNA, destroyed that substance.



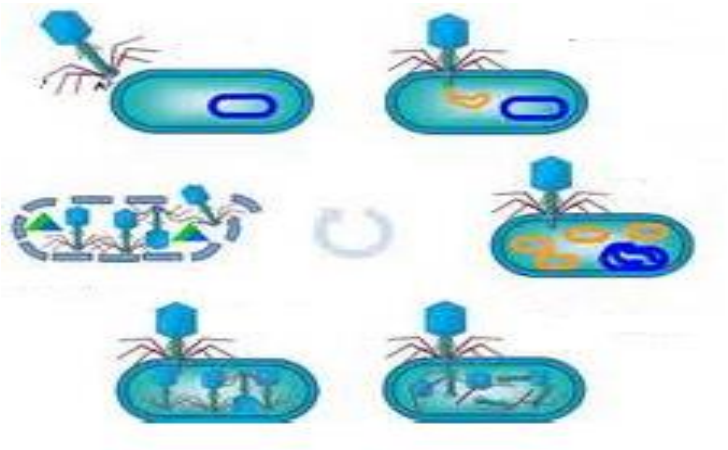
Avery, MacLeod, and McCarty further showed that the purified transforming substance precipitated at about the same rate as purified DNA. It absorbed ultraviolet light at the same wavelengths as does DNA. These findings provided compelling evidence that the transforming principle—and therefore the genetic information—resides in DNA.

20. Hershey-Chase experiments (1952)

The Hershey–Chase experiments were a series ofexperiments conducted in 1952 by Alfred Hershey andMartha Chase that helped to confirm that DNA is genetic material. While DNA had been known to biologists since 1869, many scientists still assumed at the time that proteinscarried the information for inheritance because DNA appeared simpler than proteins. In their experiments, Hershey and Chase showed that when bacteriophages, which are composed of DNA and protein, infect bacteria, their DNA enters the host bacterial cell, but most of their protein does not. Although the results were not conclusive, and Hershey and Chase were cautious in their interpretation, previous, contemporaneous and subsequent discoveries all served to prove that DNA is the hereditary material. Knowledge of DNA gained from these discoveries has applications in forensics,crime investigation and genealogy.



Bacteriophage

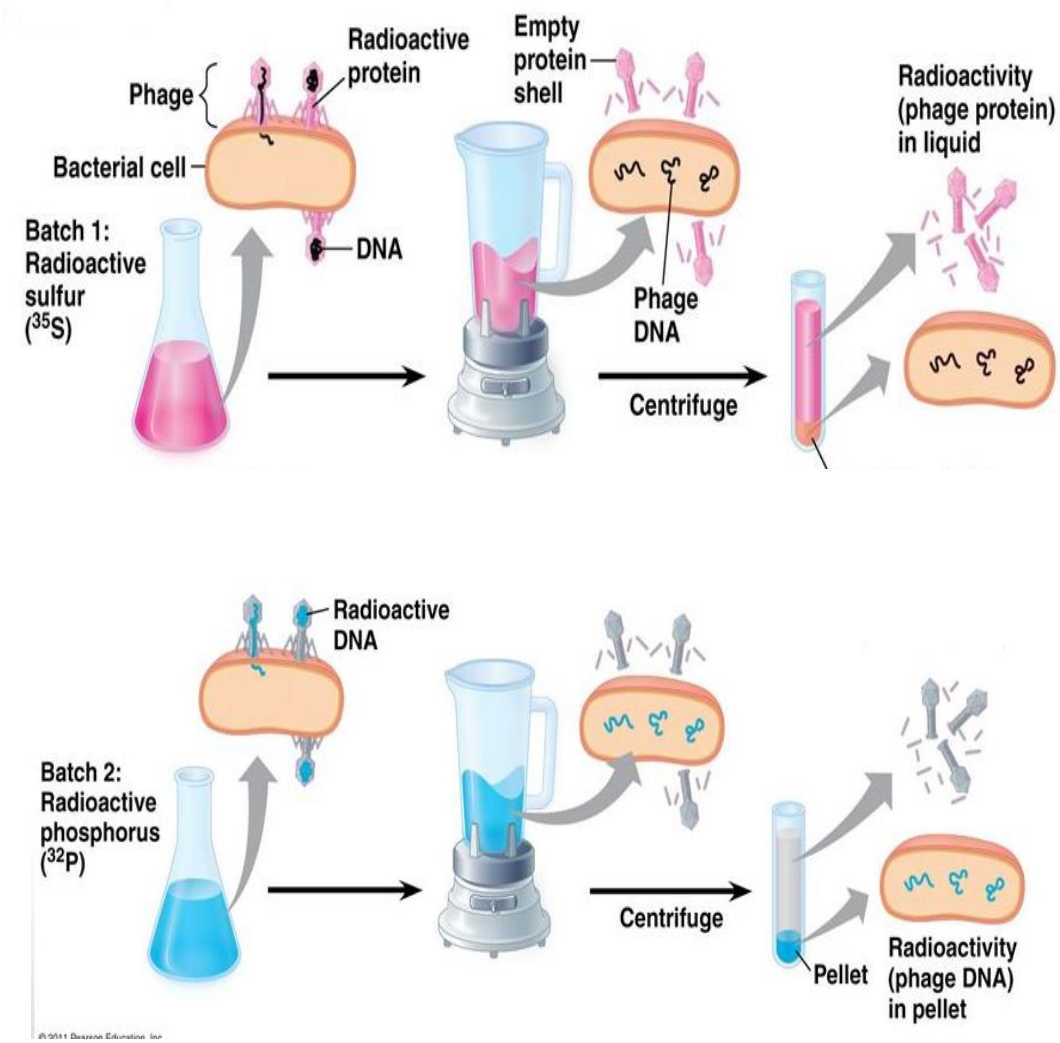


Life Cycle of a Bacteriophage

They used radioactive isotopes of phosphorus and sulfur. DNA contains P but not S; so Hershey and Chase used ^{32}P to follow phage DNA during reproduction. Protein contains sulfur but not phosphorus; so they used ^{35}S to follow the protein. Hershey and Chase first 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 labelled with ^{35}S.

Hershey and Chase then infected separate batches of unlabeled *E. coli* with the ^{35}S- and ^{32}P-labeled phages. Then they placed the *E. coli* cells in a blender and sheared off the empty protein coats 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.

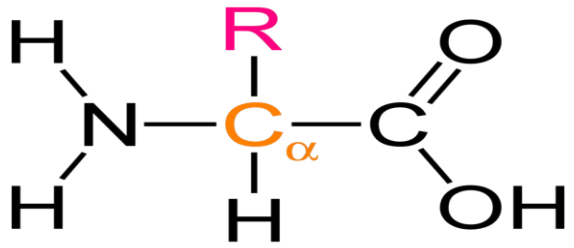
When phages labeled with ^{35}S infected the bacteria, most of the radioactivity separated with the protein coats. When new phages emerged from the cell, they contained almost no radioactivity. When phages labelled with ^{32}P infected the bacteria, and removed the protein coat, radioactivity was present in the cells. When new phages emerged from the cell, they were also radioactive.



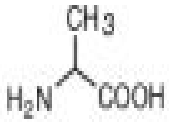
Chapter 5. Protein structure

21. Chemical composition of protein

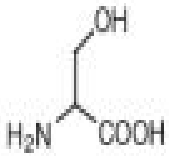
Proteins are polymers of amino acids.They range in size from small to very large. All the proteins are made up of twenty different types of amino acids. So these amino acids are called standard amino acids.



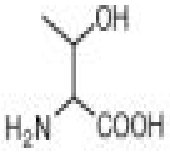
Glycine (Gly, G)



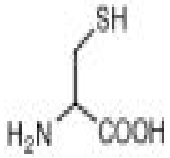
Alanine (Ala, A)



Serine (Ser, S)

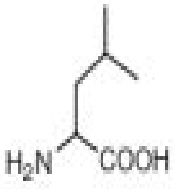


Threonine (Thr, T)

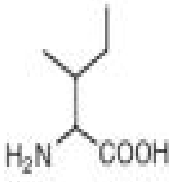


Cysteine (Cys, C)

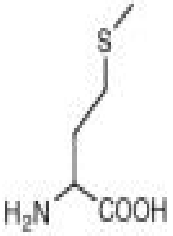
Valine (Val, V)



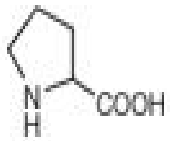
Leucine (Leu, L)



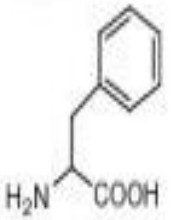
Isoleucine (Ile, I)



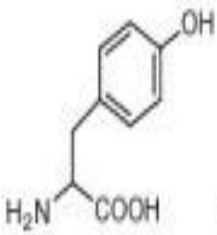
Methionine (Met, M)



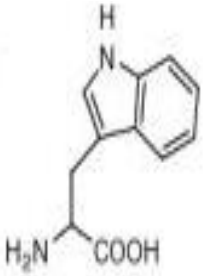
Proline (Pro, P)



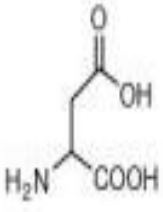
Phenylalanine (Phe, F)



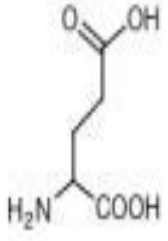
Tyrosine (Tyr, Y)



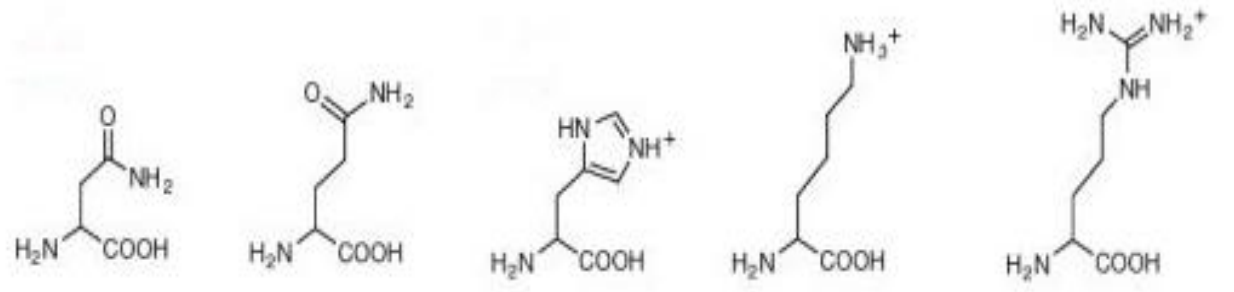
Tryptophan (Trp, W)



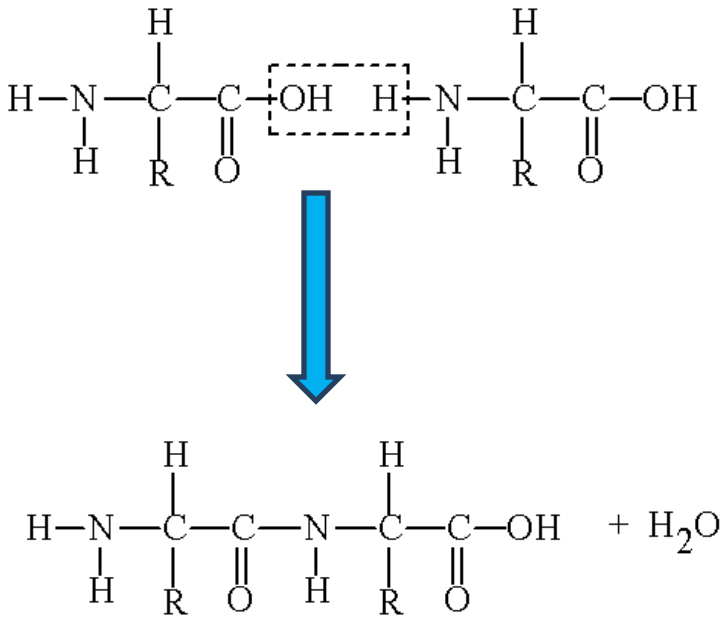
Aspartic Acid (Asp, D)



Glutamic Acid (Glu, E)



In a protein molecule, each amino acid residue is joined to its neighbour by a specific type of covalent bond which is called Peptide Bond.

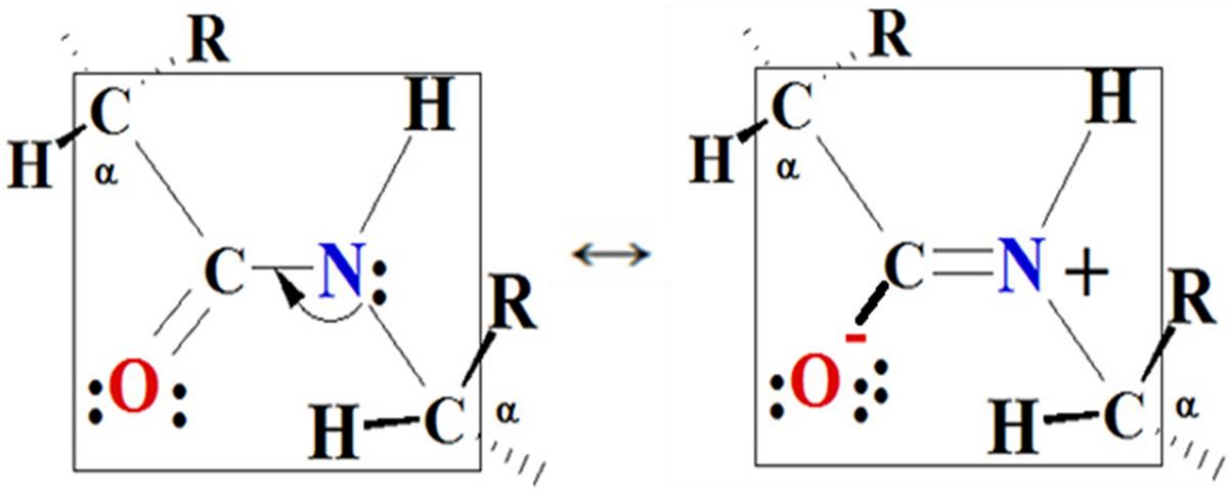


Amino acids can successively join to form dipeptides, tripeptides, tetrapeptides, oligo peptides and polypeptides.

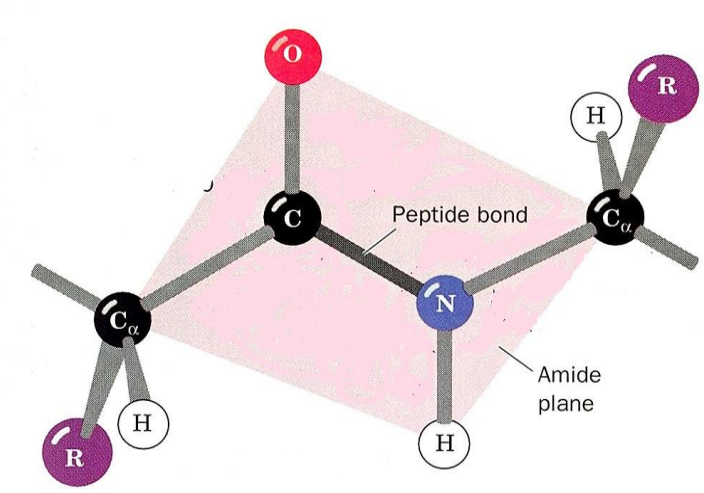
22. Primary structure of proteins

Primary structure or covalent structure of protein refers to the amino acid sequence of its polypeptide chain. Each type of protein has a unique amino acid sequence. In general, polypeptides are unbranched polymers, so their primary structure can often be specified by the sequence of amino acids along their backbone. However, proteins can become cross-linked, most commonly by disulfide bonds, and the primary structure also requires specifying the cross-linking atoms, e.g., specifying the cysteines involved in the protein's disulfide bonds. Other crosslinks include desmosine. The chiral centers of a polypeptide chain can undergo racemization. In particular, the L-amino acids normally found in proteins can spontaneously isomerize at the alpha carbon atom to form D-amino acids, which cannot be cleaved by most proteases.

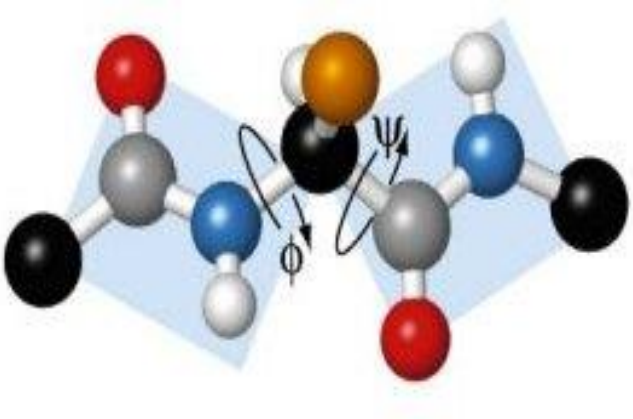
Peptide Bond Is Rigid and Planar.Linus Pauling and Robert Corey carefully analyzed the peptide bond. Their findings laid the foundation for our present understanding of protein structure. They demonstrated that the peptide C - N bond is somewhat shorter than the C - N bond in a simple amine.



The six atoms of the peptide group are co-planar i.e., lie in a single plane, with the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen trans to each other.



Pauling and Corey concluded that the peptide C - N bonds are unable to rotate freely because of their partial double-bond character.Rotation is permitted about the N - αC and the αC - C bonds.



The bond angles resulting from rotations at C are labelled ϕ (phi) for the N - αC bond and ψ (psi) for the αC - C bond.In principle, ϕ and ψ can have any value between +180 & -180.

23. Secondary structure of proteins

In biochemistry and structural biology, protein secondary structure is the general three-dimensional form of *local segments* of proteins. Secondary structure can be formally defined by the pattern of hydrogen bonds of the protein (such as alpha helices and beta sheets) that are observed in an atomic-resolution structure. More specifically, the secondary structure is defined

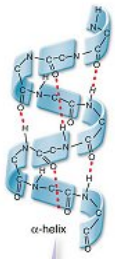
by the patterns of hydrogen bonds formed between amine hydrogen and carbonyloxygen atoms contained in the backbone peptide bonds of the protein. The secondary structure may alternatively be defined based on the regular pattern of backbone dihedral angles in a particular region of the Ramachandran plot; thus, a segment of residues with such dihedral angles may be called a helix, regardless of whether it has the correct hydrogen bonds. The secondary structure may be provided by crystallographers in the corresponding PDB file.

Secondary structure does not describe the specific identity of amino acids in the protein which are defined as theprimary structure, nor the *global* atomic positions in three-dimensional space, which are considered to be tertiary structure. Other types of biopolymers such as nucleic acids also possess characteristic secondary structures.

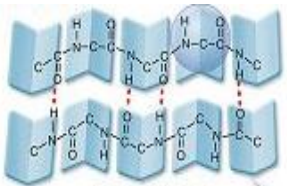
The concept of secondary structure was first introduced by Kaj Ulrik Linderstrøm-Lang at Stanford in 1952.

The most prominent are:-

α-helix



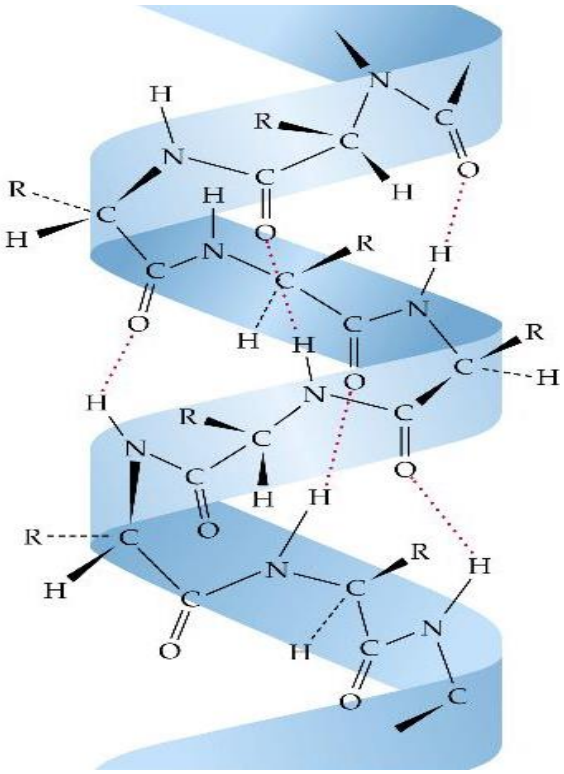
β- conformations.



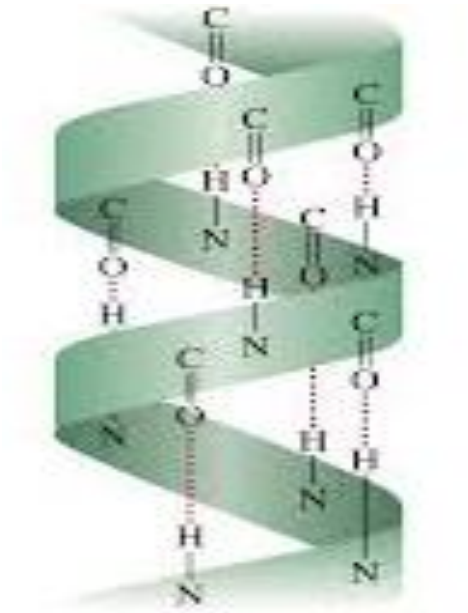
24. α- Helix

The alpha helix (α-helix) is a common secondary structure of proteins and is a righthand-coiled or spiral conformation (helix) in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues earlier (i+4=i, ydrogen bonding). This secondary structure is also sometimes called a classic Pauling–Corey–Branson alpha helix (see below). The name 3.6_{13}-helix is also used for this type of helix, denoting the number of residues

per helical turn, and 13 atoms being involved in the ring formed by the hydrogen bond. Among types of local structure in proteins, the α-helix is the most regular and the most predictable from sequence, as well as the most prevalent.



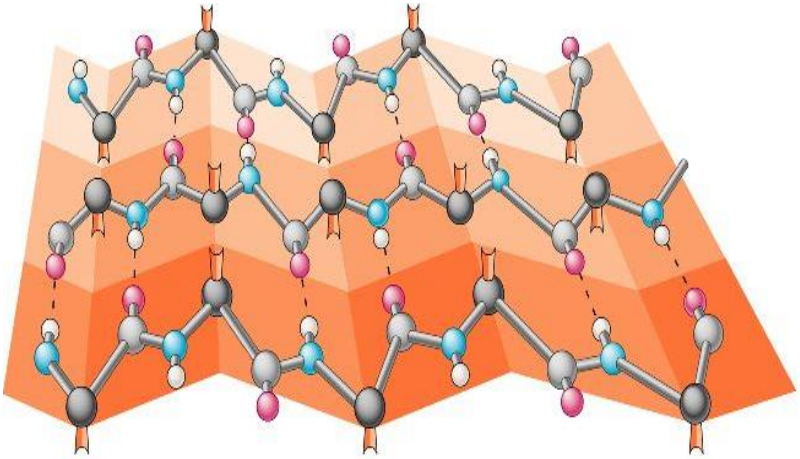
The helical twist of the α-helix found in all proteins is right-handed. The repeating unit is a single turn of the helix, which extends about 5.4 Å (includes 3.6 amino acid residues) along the long axis. The amino acid residues in an helix have conformations with psi = –45 to –50 and phi = – 60. An helix makes optimal use of internal hydrogen bonds.



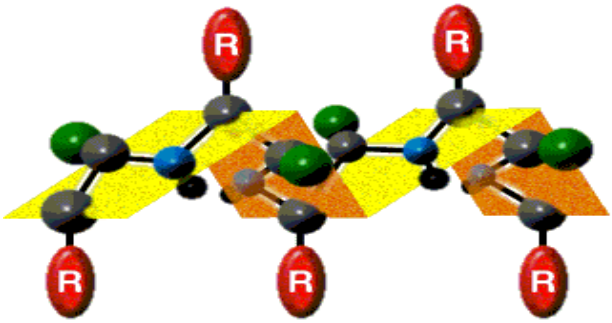
About one-fourth of all amino acid residues in polypeptides are found in α-helices while in some proteins it is the predominant structure.

25. β- Pleated Sheets

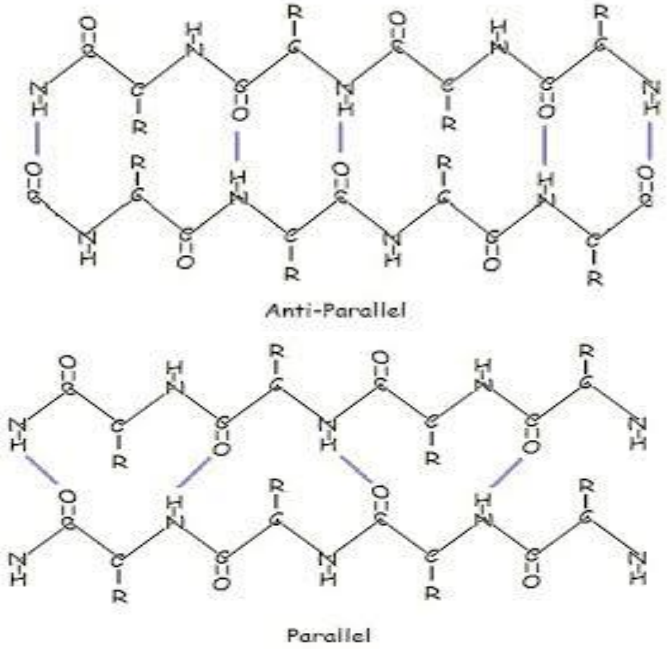
Pauling and Corey predicted a second type of secondary structure which they called **β-sheets.** This is a more extended conformation of polypeptide chains. The β sheet (also β-pleated sheet) is the second form of regular secondary structure in proteins. Beta sheets consist of beta strands connected laterally by at least two or three backbone hydrogen bonds, forming a generally twisted, pleated sheet. A beta strand (also β strand) is a stretch of polypeptide chain typically 3 to 10 amino acids long with backbone in an extended conformation. The higher-level association of β sheets has been implicated in formation of the protein aggregates and fibrils observed in many human diseases, notably theamyloidoses such as Alzheimer's disease.



The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions.



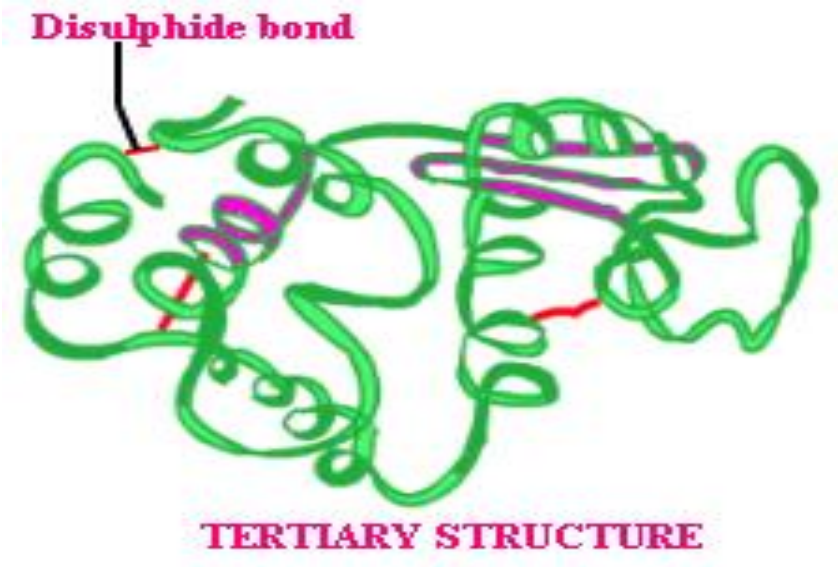
Hydrogen bonds are formed between adjacent segments of polypeptide chain.The adjacent polypeptide chains in a sheet can be either parallel or antiparallel.



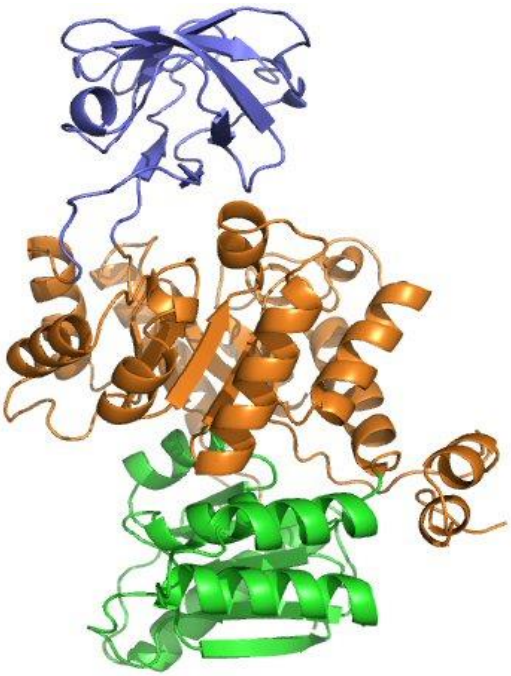
26. Tertiary Structure of Proteins

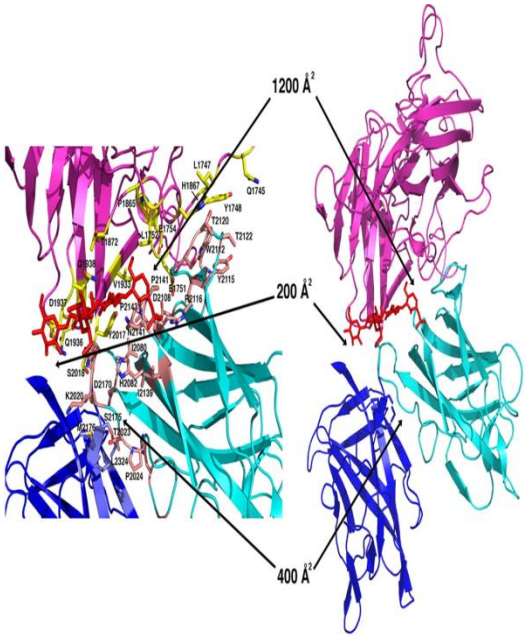
The term protein tertiary structure refers to a protein's geometric shape. The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein’s tertiary structure.

The tertiary structure will have a single polypeptide chain "backbone" with one or more protein secondary structures, the protein domains. Amino acid side chains may interact and bond in a number of ways. The interactions and bonds of side chains within a particular protein determine its tertiary structure. The protein tertiary structure is defined by its atomiccoordinates. These coordinates may refer either to a protein domain or to the entire tertiary structure.[1][2] A number of tertiary structures may fold into a quaternary structure.



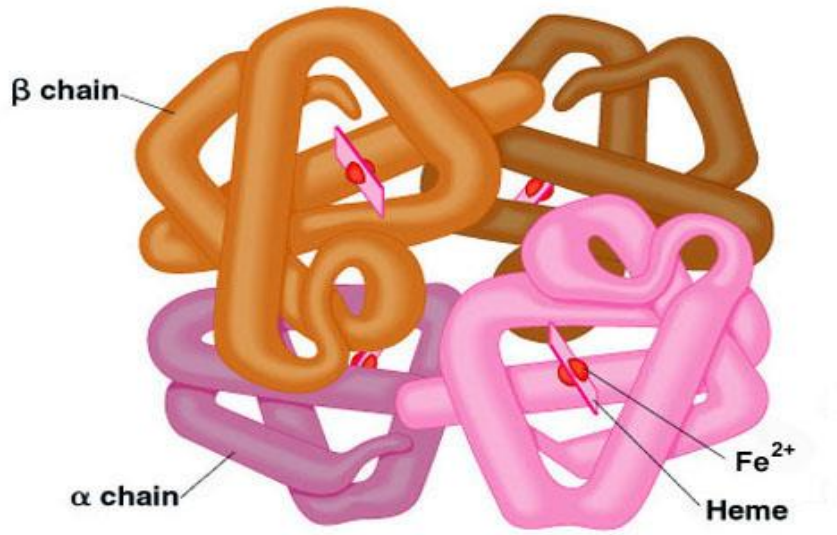
It includes longer-range aspects of amino acid sequence. Amino acids that are far apart in the polypeptide chain may interact within the completely folded structure of a protein. Interacting segments of polypeptide chains are held in their characteristic tertiary positions by different kinds of weak interactions (and sometimes by covalent bonds) between the segments. Large polypeptide chains usually fold into two or more globular clusters known as domains, which often give these proteins a bi- or multilobal appearance.



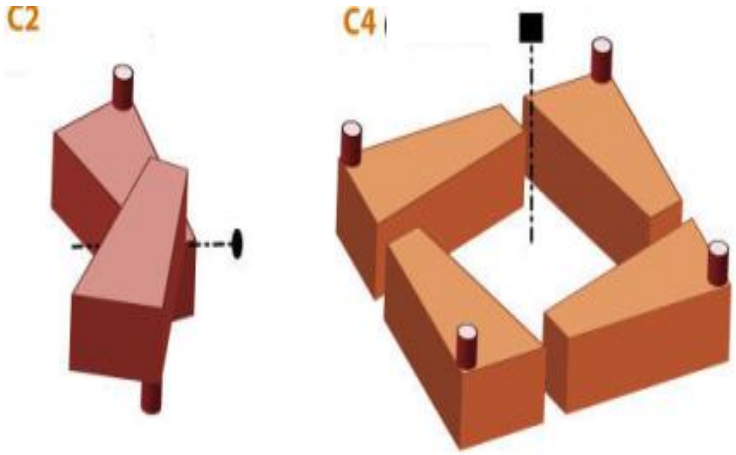


27. Quaternary Structure of Proteins

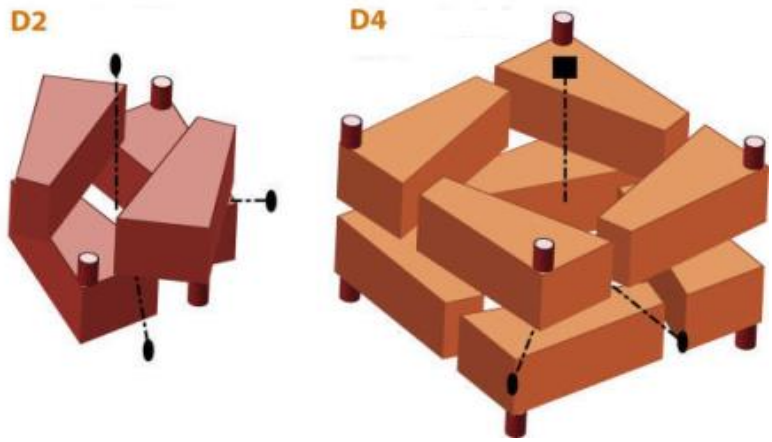
Quaternary structure is the number and arrangement of multiple folded protein subunits in a multi-subunit complex. It includes organisations from simple dimers to large homooligomers and complexes with defined or variable numbers of subunits. Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different. The spatial arrangement of these subunits is known as a protein's quaternary structure. A multi-subunit protein is also referred to as a multimer. A multimer with just a few subunits is called as oligomer and a single subunit or a group of subunits, is called a protomer.



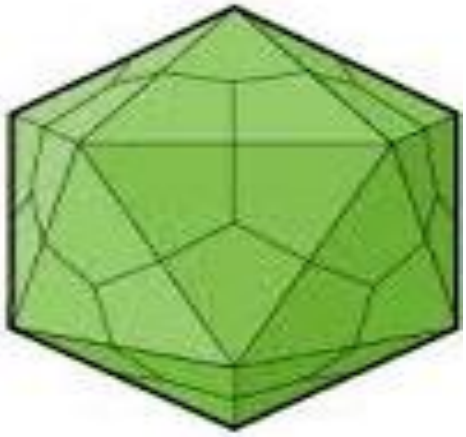
Identical subunits of multimeric proteins are generally arranged in a symmetric patterns. Oligomers can have either rotational symmetry or helical symmetry. There are several forms of rotational symmetry. The simplest is cyclic symmetry, involving rotation about a single axis.



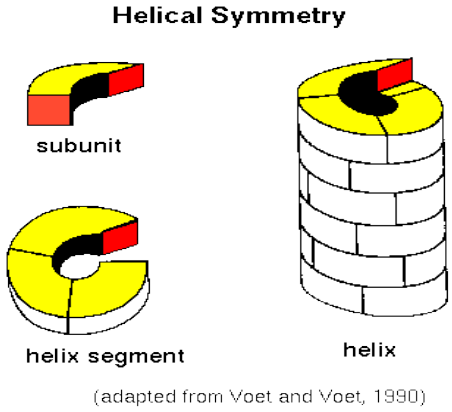
A somewhat more complicated rotational symmetry is dihedral symmetry, in which a twofold rotational axis is present.



More complex rotational symmetries include icosahedral symmetry. An icosahedron is a regular 12-cornered polyhedron having 20 triangular faces.



The other major type of symmetry found in oligomers is helical symmetry.



(adapted from Voet and Voet, 1990)

Chapter 6. Organisation of Genetic Material

28. Genetic Materials in Viruses

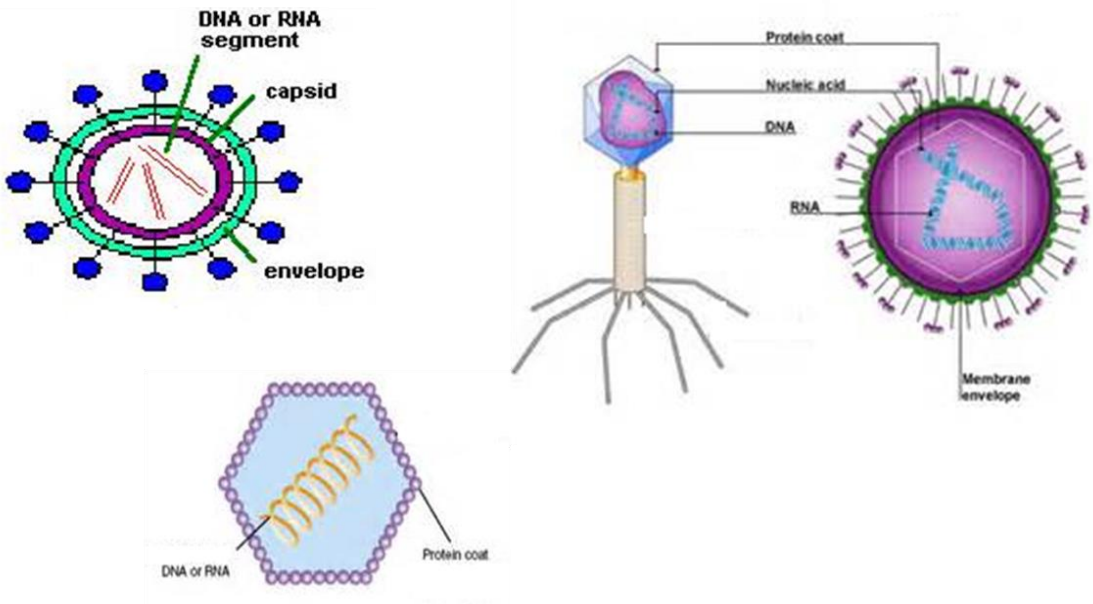
Viruses are exceptionally simple and extremely small microorganisms. An enormous variety of genomic structures can be seen among viral species; as a group, they contain more structural genomic diversity than plants, animals, archaea, or bacteria. There are millions of different types of viruses,[4] although only about 5,000 types have been described in detail.[3] As of September 2015, the NCBI Virus genome database has more than 75,000 complete genome sequences, but there are doubtlessly many more to be discovered.

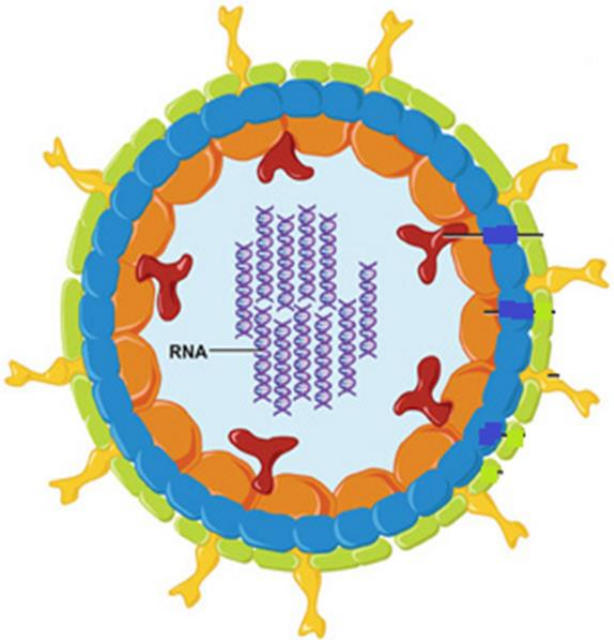
A virus has either a DNA or an RNA genome and is called a DNA virus or an RNA virus, respectively. The vast majority of viruses have RNA genomes. Plant viruses tend to have single-stranded RNA genomes and bacteriophages tend to have double-stranded DNA genomes.

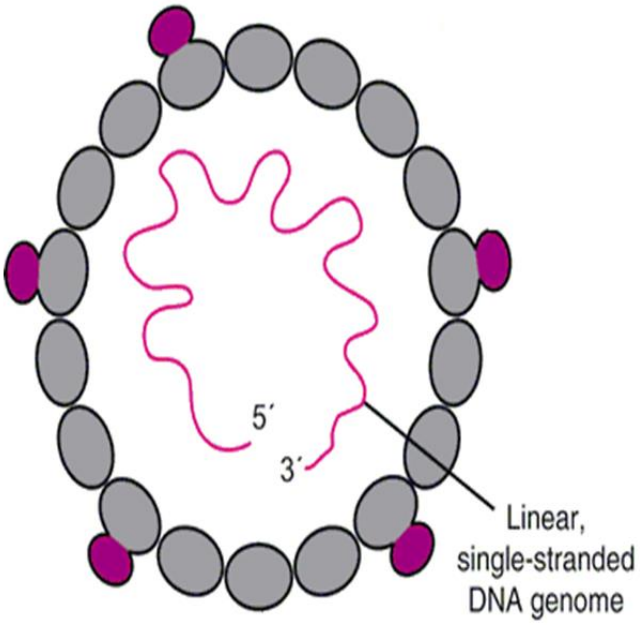
Viral genomes are circular, as in the polyomaviruses, or linear, as in the adenoviruses. The type of nucleic acid is irrelevant to the shape of the genome. Among RNA viruses and certain DNA viruses, the genome is often divided up into separate parts, in which case it is called segmented. For RNA viruses, each segment often codes for only one protein and they are usually found together in one capsid. However, all segments are not required to be in the same virion for the virus to be infectious, as demonstrated by brome mosaic virus and several other plant viruses.

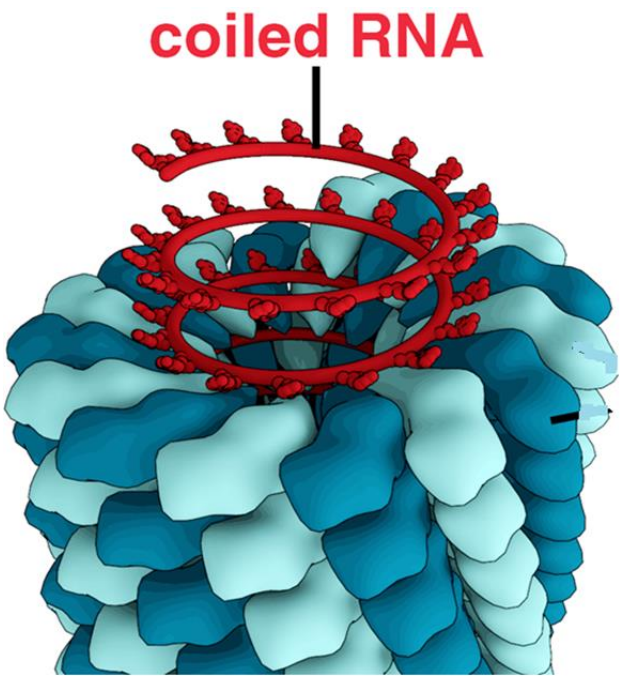
A viral genome, irrespective of nucleic acid type, is almost always either single-stranded or double-stranded. Single-stranded genomes consist of an unpaired nucleic acid, analogous to one-half of a ladder split down the middle. Double-stranded genomes consist of two complementary paired nucleic acids, analogous to a ladder. The virus particles of some virus families, such as those belonging to the Hepadnaviridae, contain a genome that is partially double-stranded and partially single-stranded. For most viruses with RNA genomes and some with single-stranded DNA genomes, the single strands are said to be eitherpositive-sense (called the plus-strand) or negative-sense (called the minus-strand), depending on if they are complementary to the viral messenger RNA (mRNA). Positive-sense viral RNA is in the same sense as viral mRNA and thus at least a part of it can be immediately translated by the host cell. Negative-sense viral RNA is complementary to mRNA and thus must be converted to positive-sense RNA by an RNA-dependent RNA polymerase before translation. DNA nomenclature for viruses with single-sense genomic ssDNA is similar to RNA nomenclature,

They have a very simple structural organization consisting of a molecule of nucleic acid and a protein coat.





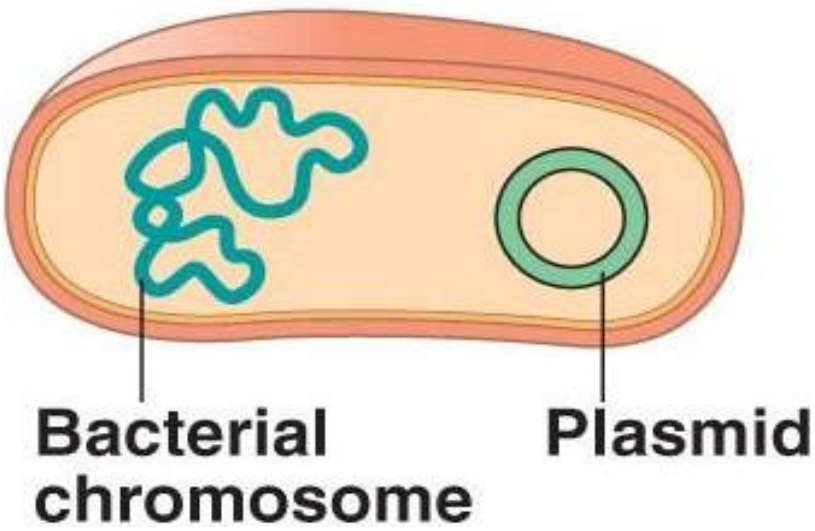




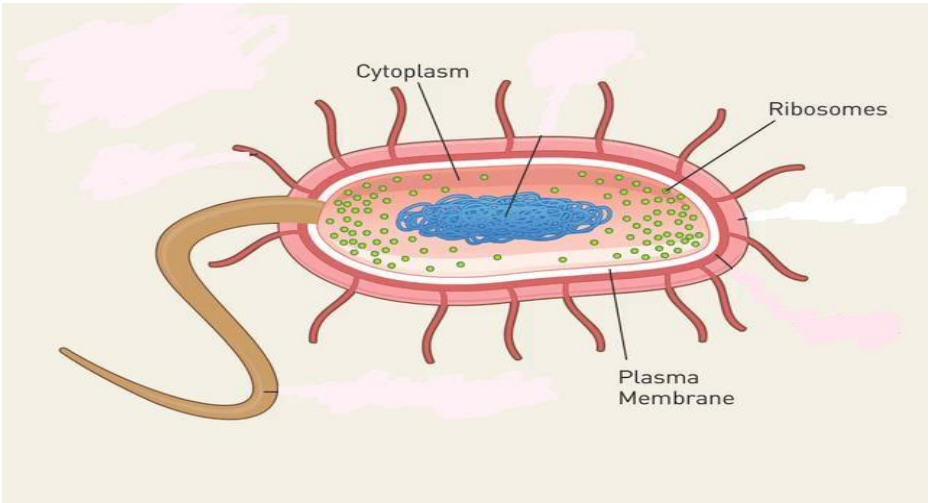
The percentage of nucleic acid in relation to protein is about 1% for the influenza virus and about 50% for some bacteriophages.The total amount of nucleic acid varies from a few thousand nucleotides to as many as 250,000 nucleotides. *E. coli's* chromosome consists of approx. 4 million nucleotide pairs.

29 30. Organization of Genetic Material in Bacteria

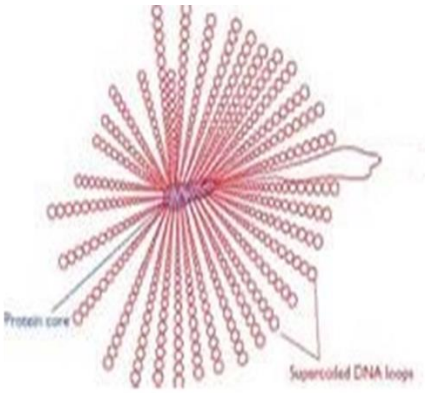
Bacterial genetics is the subfield of genetics devoted to the study of bacteria. Bacterial genetics are subtly different from eukaryotic genetics, however bacteria still serve as a good model for animal genetic studies. One of the major distinctions between bacterial and eukaryotic genetics stems from the bacteria's lack of membrane-bound organelles. Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins. The bacterial chromosome is a very long (up to 1mm). It is looped and folded and attached at one or several points to the plasma membrane.

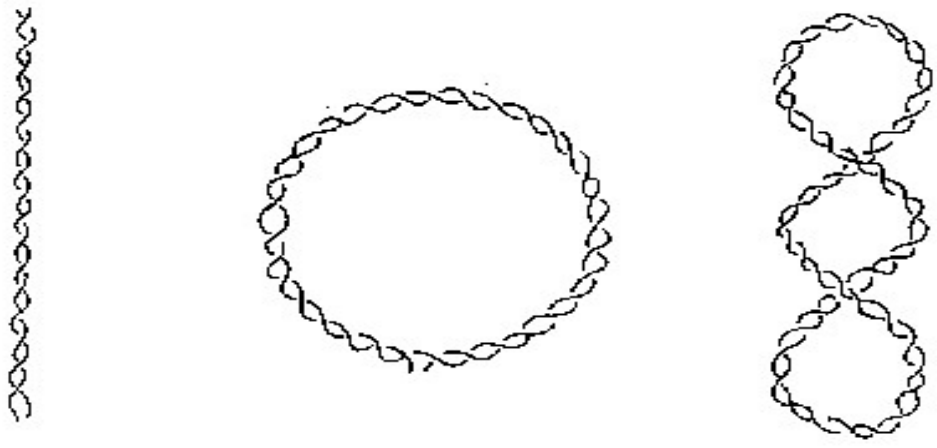


Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the nucleoid.

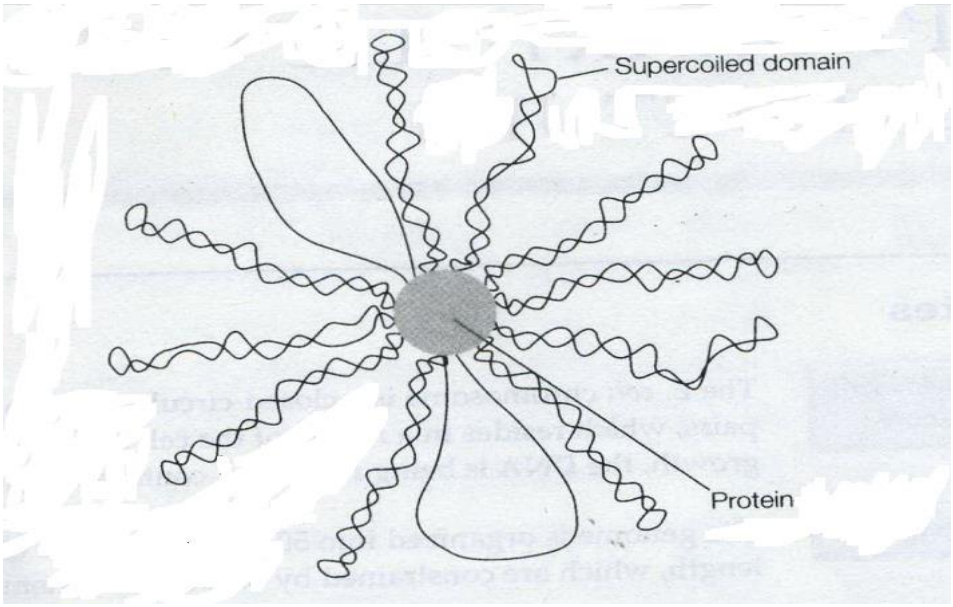


Bacterial chromatin can be released from the cell by gentle lysis of the cell. Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin. Some of the loops are super-coiled while some are relaxed. Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.





If a super-coiled DNA molecule receives a nick, the strain of under-winding is immediately removed, and all the super-coiling is lost. Studies confirm that continued nuclease treatment increases number of relaxed loops.

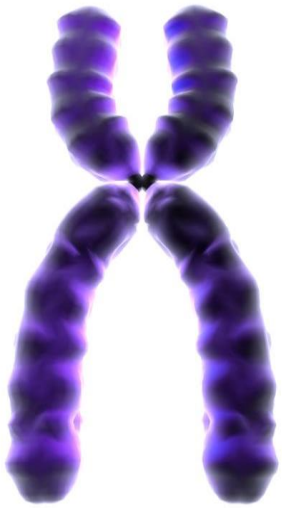


The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others. So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops. The super-coiled loops are dynamic structures which change during cell growth & division. An *E. coli* chromosome is estimated to have about 400 super-coiled loops. Each loop has an average length of about 10-20 kbp. The DNA compaction in a bacterial cell is contributed by super-coiling of loops, macromolecular crowding and DNA-binding proteins.

31. Organization of Genetic Material in Eukaryotes

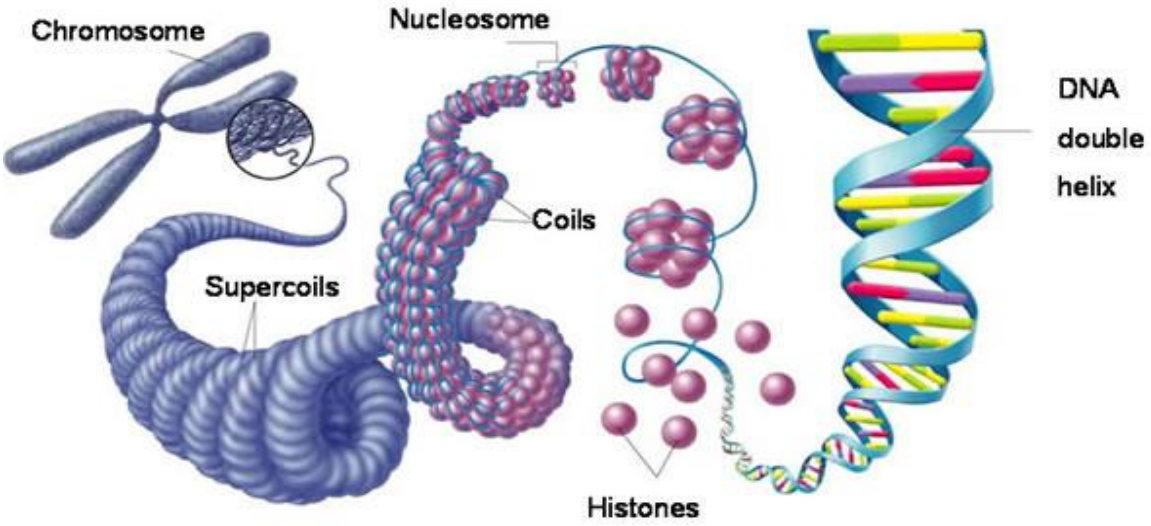
The genome is segmented in eukaryotes. It is made up of a number of linear chromosomes. A genome is the entire set of unique chromosomes that segregate to a gamete or to a haploid life stage in organisms with an alternation of generations like plants.
A diploid has two copies 2n = 2x of their genome while a tetraploid has 4 copies 2n = 4x of the basic monoploid genome X. A diploid then has homologous paired chromosomes.
The strands are organized by coiling the strand of DNA about an octomer of histones into nucleosomes. Each Nucleosome is an assembly of histone proteins with the DNA making two turns around the group and clamped by another histone, H1. This condenses the strand length.
The strand is super coiled for further compaction.
DNA packaging is increased during mitosis or meiosis with proteins like condensin & cohesin linking replicated chromatids dyads at the centromere and holding the supercoils together in each arm. The centromere permits spindle attachments to each pole so daughter cells each get a copy inmitosis.

Located along the DNA strand are sequence patterns the regulate gene expression as well as the open reading frames of the gene loci. Chromosomes contain hundreds to thousands of gene loci depending on the length of the chromosome.



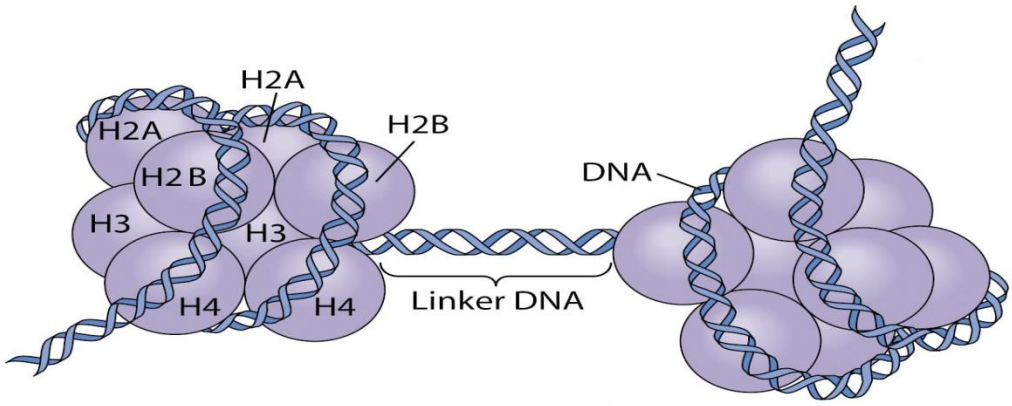
Chromosome

Each un-replicated chromosome consists of a single molecule of DNA. If stretched out, some human chromosomes would be several centimetres long. To package such a tremendous length of DNA into this small volume, each DNA molecule is coiled again and again and tightly packed around histone proteins. As eukaryotic chromosomes are not circular, so instead of super-coiling, the mechanism of packaging involves winding the DNA around special proteins, the histones. DNA with bound histones in the eukaryotes is called as chromatin. Chromatin consists of roughly spherical subunits, the nucleosomes, each containing approx. 200 bp of DNA and nine histones. A condensed mitotic chromosome is about 50,000 times shorter than fully extended DNA. Highly condensed chromatin is known as heterochromatin. The more extended form is known as euchromatin.



32. Histone Proteins

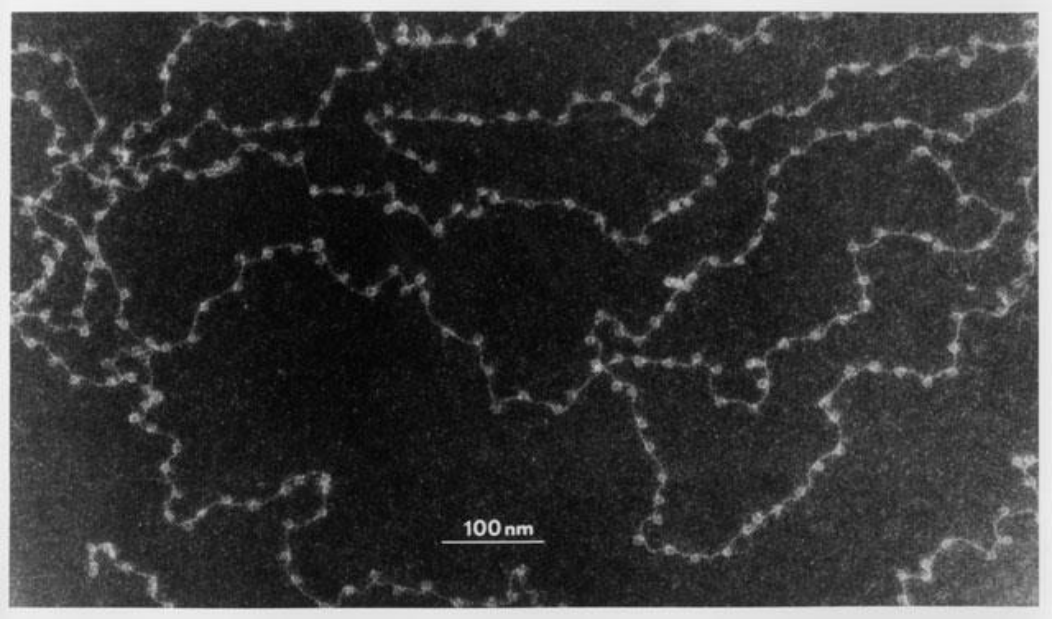
Most abundant proteins in the chromatin are histones. Histones are highly alkaline proteins found in eukaryotic cell nuclei that package and order the DNAinto structural units called nucleosomes. They are the chief protein components of chromatin, acting as spools around which DNA winds, and playing a role in gene regulation. Without histones, the unwound DNA in chromosomes would be very long (a length to width ratio of more than 10 million to 1 in human DNA). For example, each human cell has about 1.8 meters of DNA, (~6 ft) but wound on the histones it has about 90 micrometers (0.09 mm) of chromatin, which, when duplicated and condensed during mitosis, result in about 120 micrometers of chromosomes. There are nine types of histones including two each of H2A, H2B, H3 and H4 and one of H1. These histones fall in five major classes i.e., H1, H2A, H2B, H3 and H4.A typical human cell contains about 60 million copies of each kind of histone.



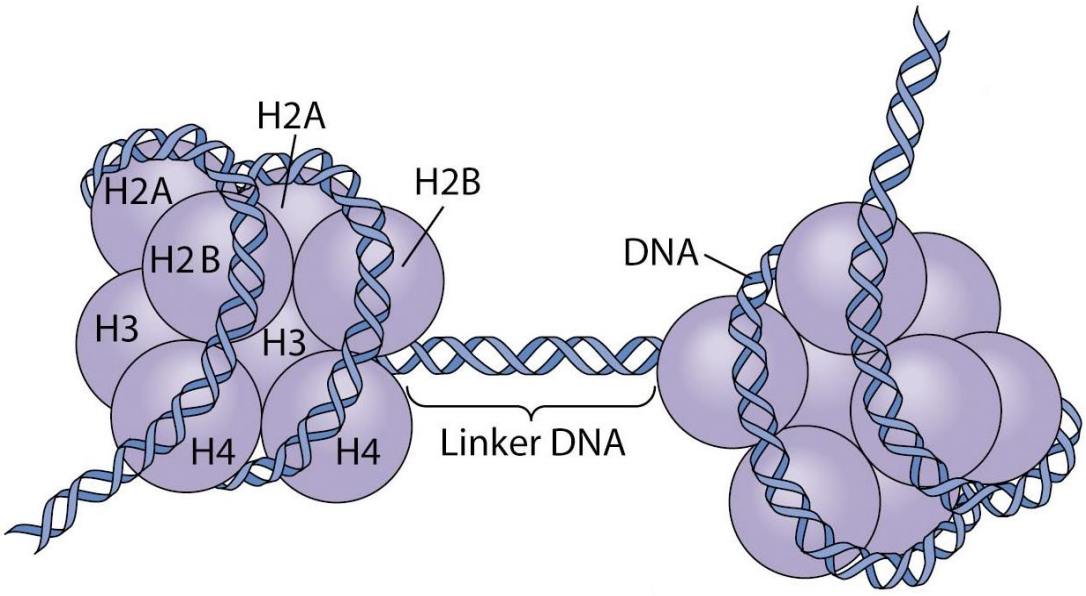
All histones have a high percentage of arginine and lysine but the lysine-to-arginine ratio differs in each type of histone. The positively charged side chains of lysine and arginine enable histones to bind to the negatively charged phosphate groups of the DNA. The electrostatic attraction is an important stabilizing force in the chromatin.

33. The Nucleosome

The uncondensed chromatin resembles beads on a string when viewed under the electron microscope.Each bead is a nucleoprotein complex called nucleosome. A nucleosome is a basic unit of DNA packaging in eukaryotes, consisting of a segment of DNA wound in sequence around eight histone protein cores.[2] This structure is often compared to thread wrapped around a spool.

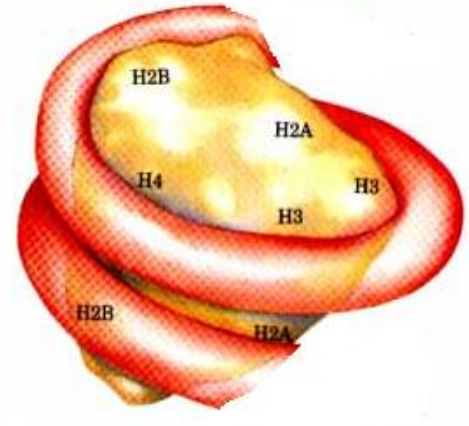


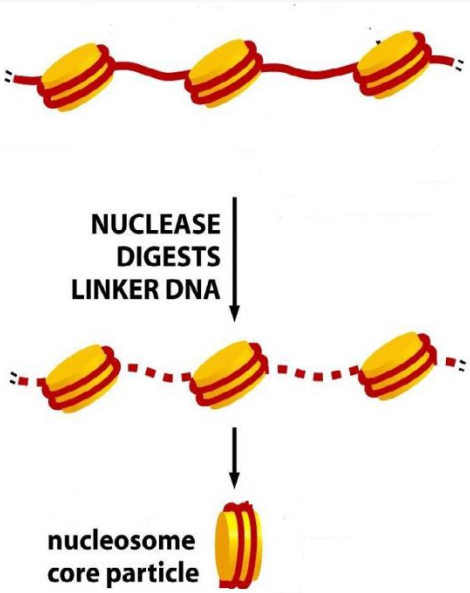
Nucleosomes form the fundamental repeating units of eukaryotic chromatin, which is used to pack the large eukaryotic genomes into the nucleus while still ensuring appropriate access to it (in mammalian cells approximately 2 m of linear DNA have to be packed into a nucleus of roughly 10 µm diameter). Nucleosomes are folded through a series of successively higher order structures to eventually form a chromosome; this both compacts DNA and creates an added layer of regulatory control, which ensures correct gene expression. Nucleosomes are thought to carry epigenetically inherited information in the form of covalent modifications of their core histones. Nucleosomes were observed as particles in the electron microscope by Don and Ada Olins and their existence and structure (as histone octamers surrounded by approximately 200 base pairs of DNA) were proposed by Roger Kornberg. The role of the nucleosome as a general gene repressor was demonstrated by Lorch et al. in vitro [8] and by Han and Grunstein in vivo.



The size of the linker DNA between the nucleosomes varies among different organisms and even different organs of the same organism. The length of DNA wrapped around nucleosomes also varies from one organism to the other ranging from about 170-240 bp. Prolonged nuclease digestion of chromatin cleaves additional nucleotides. The structure that remains is the nucleosome core particle. The nucleosome core particle consists of an octameric protein complex (two copies of each H2A, H2B, H3 & H4) with a 146 bp DNA fragment wound around it.

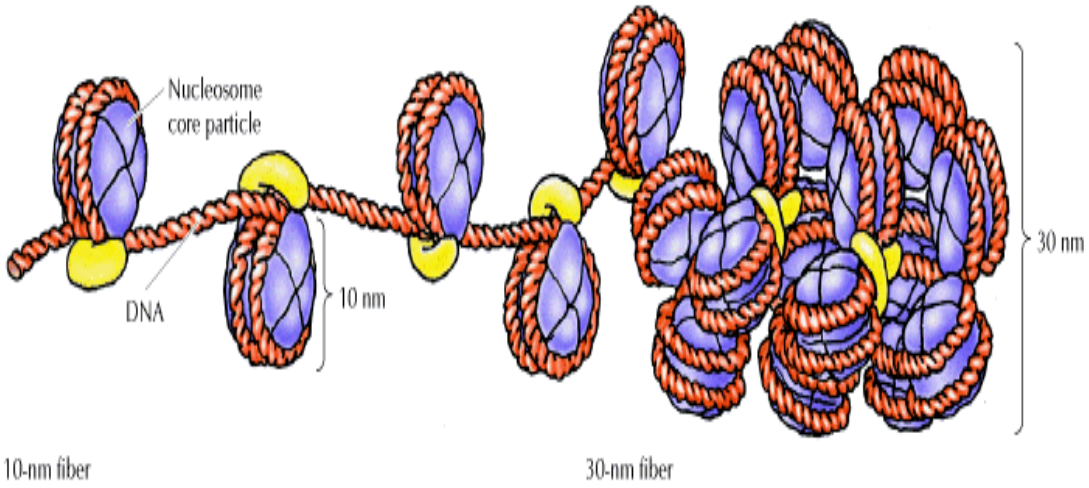
The crystal structure of the nucleosome core particle (PDB ID:1EQZ . Histones H2A,H2B, H3 and H4 are coloured, DNA is gray.The nucleosome core particle consists of approximately 147 base pairs of DNA wrapped in 1.67 left-handed superhelical turns around a histone octamer consisting of 2 copies each of the core histones H2A, H2B, H3, and H4. Core particles are connected by stretches of "linker DNA", which can be up to about 80 bp long. Technically, a nucleosome is defined as the core particle plus one of these linker regions; however the word is often synonymous with the core particle. Genome-wide nucleosome positioning maps are now available for many model organisms including mouse liver and brain.



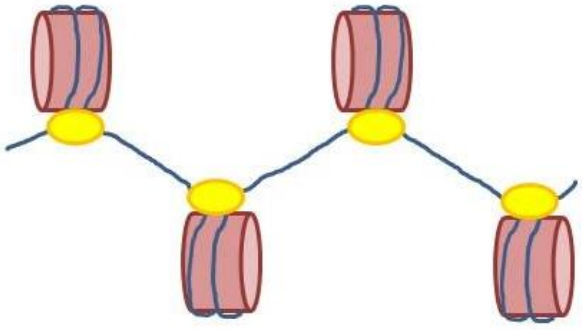


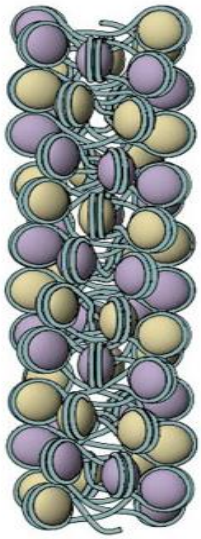
34 &35. The 30-nm Fiber

It is still unclear that how a chain of nucleosomes folds into higher order structures.The next level of organization is a 30-nm fiber.

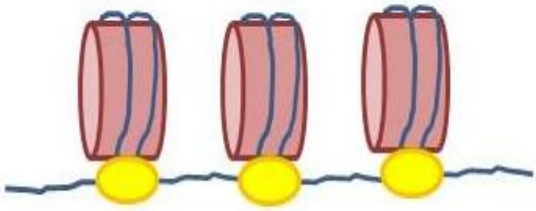


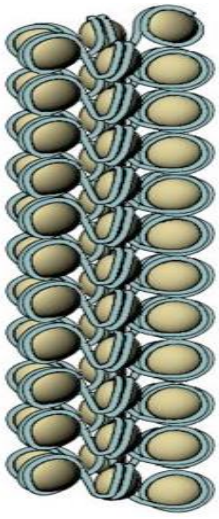
Various models have been proposed to explain how nucleosomes fold to form the 30-nm fiber. However, two models gained the most support:-
Zigzag Model
Solenoid Model
Zigzag model predicts that the linker DNA forms a straight path between successive nucleosomes.The nucleosomes lie on opposite sides of the fiber.





The solenoid model predicts that the nucleosome chain forms a helical structure with about 6 nucleosomes per turn.Linker DNA is bent to connect neighbouring nucleosomes. Some of the reconstitution experiments appear to support Zigzag models while some others support Solenoid model.

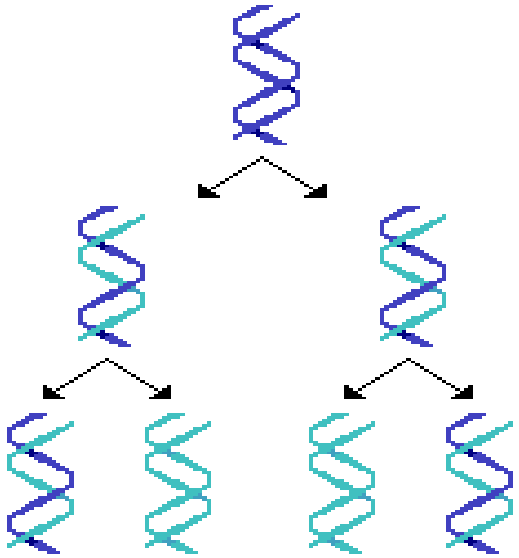




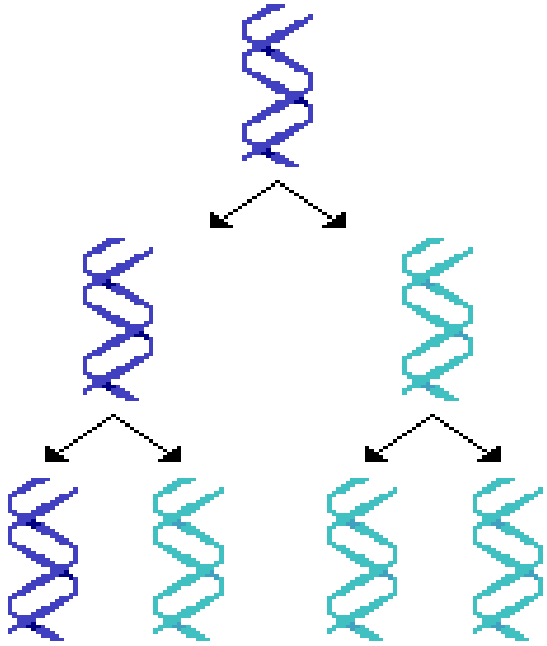
Chapter 7. DNA Replication

36. Replication of DNA

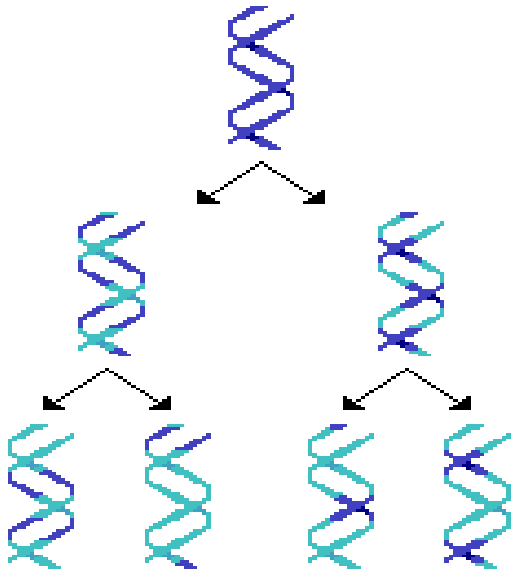
DNA replication is the process of producing two identical replicas from one original DNA molecule. This biological process occurs in all living organisms and is the basis for biological inheritance. DNA is made up of two strands and each strand of the original DNA molecule serves as a template for the production of the complementary strand. The double-helical model for DNA includes the concept that the two strands are complementary.
Thus, each strand can in principle serve as the template for making its own partner. A number of models were proposed to explain the mode of replication of DNA.
But the semiconservative model for DNA replication is the correct one. The Watson–Crick model for DNA replication proposed that the two parental strands separate and that each then serves as a template for a new progeny strand. This is called semiconservative replication because each daughter duplex has one parental strand and one new strand which means that one of the parental strands is “conserved” in each daughter duplex.



Another potential mechanism is conservative replication, in which the two parental strands stay together and somehow produce another daughter helix with two completely new strands.

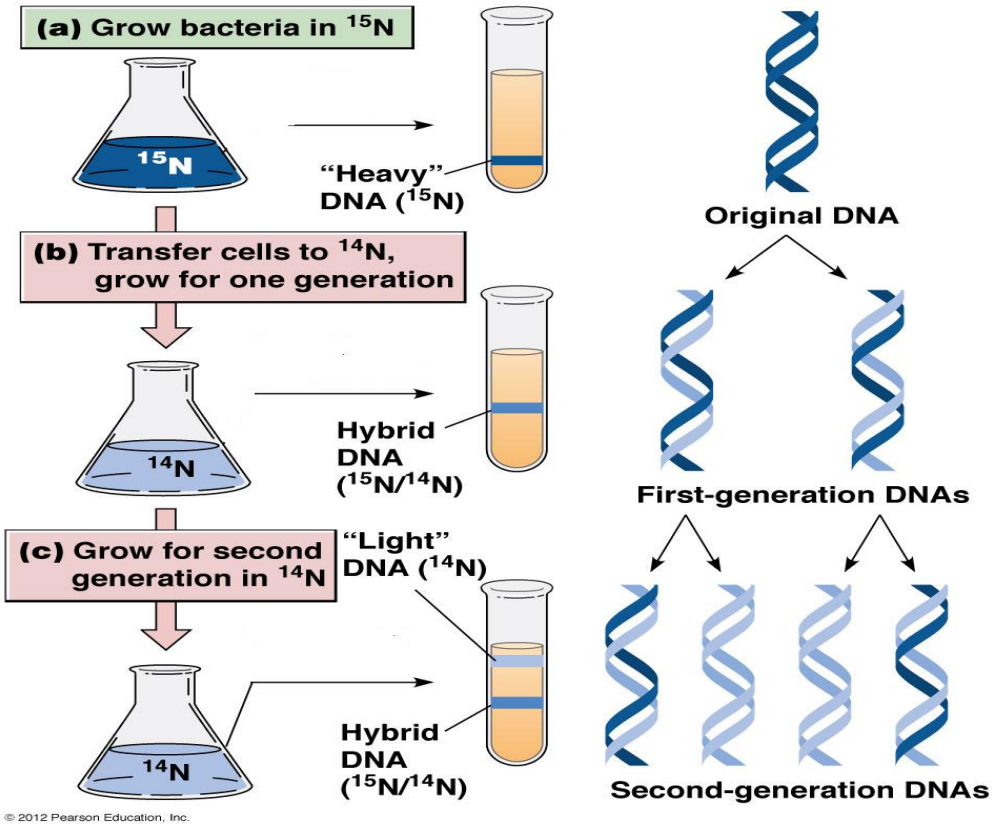


Yet another possibility is dispersive replication, in which the DNA becomes fragmented so that new and old DNAs coexist in the same strand after replication.



37. Experiment of Meselson & Stahl

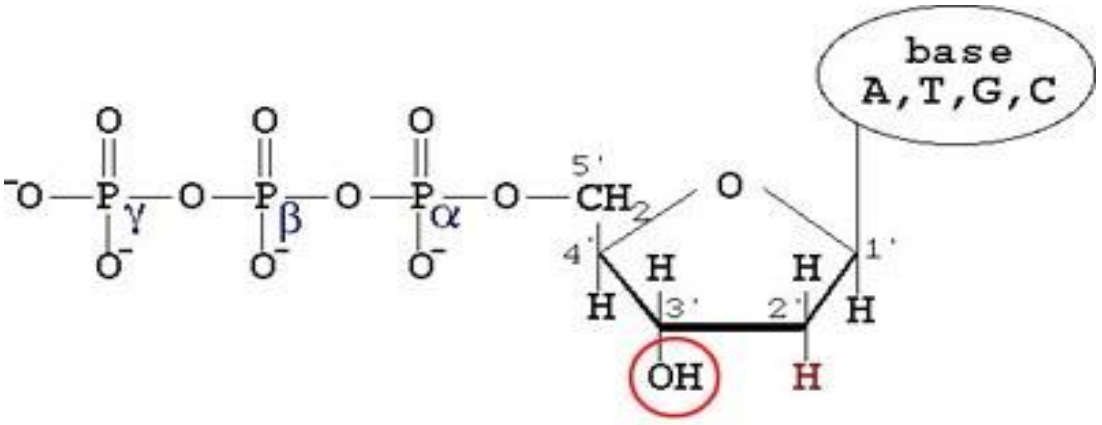
The Meselson–Stahl experiment was an experiment by Matthew Meselson and Franklin Stahl in 1958 which supported the hypothesis that DNA replication wassemiconservative. In semiconservative replication, when the double stranded DNA helix is replicated, each of the two new double-stranded DNA helixes consisted of one strand from the original helix and one newly synthesized. It has been called "the most beautiful experiment in biology. Meselson and Stahl decided the best way to tag the parent DNA would be to change one of the atoms in the parent DNA molecule. Since nitrogen is found in the nitrogenous bases of each nucleotide, they decided to use an isotope of nitrogen to distinguish between parent and newly-copied DNA. The isotope of nitrogen had an extra neutron in the nucleus, which made it heavier. They labeled *E. coli* DNA with heavy nitrogen (15N) by growing cells in a medium enriched in this nitrogen isotope. This made the DNA denser than normal. Then they switched the cells to an ordinary medium containing primarily ^{14}N, for various lengths of time.
Finally, they subjected the DNA to density gradient centrifugation to determine the density of the DNA.



38 & 39. Chemistry of DNA Synthesis

Oligonucleotide synthesis is the chemical synthesis of relatively short fragments of nucleic acids with defined chemical structure (sequence). Two key substrates are required for the synthesis of DNA to proceed.
Deoxynucleoside triphosphates
Primer:template junction

Four deoxynucleoside triphospahtes namely dGTP, dCTP, dATP & dTTP are required.
Nucleoside triphosphates have three phosphoryl groups attached to the 5' hydroxyl of deoxyribose. The innermost phosphoryl group is called the α-phosphate whereas the middle and outermost groups are called β- and ɣ- phosphates.



dNTP

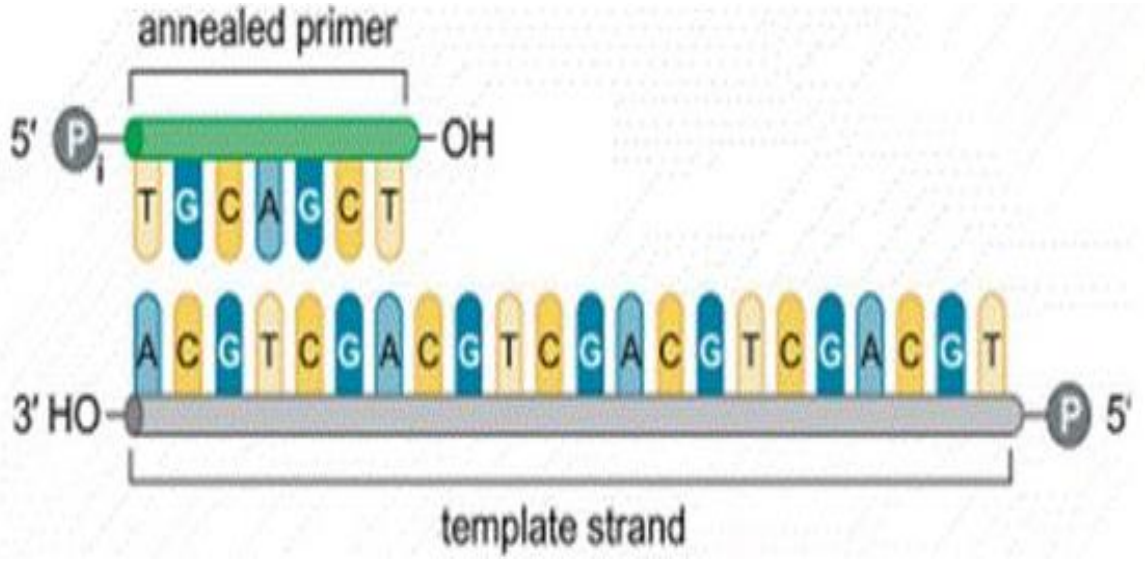
deoxyribonucleotide triphosphate

The second important substrate for DNA synthesis is a particular arrangement of single stranded DNA (ssDNA) and double stranded DNA (dsDNA).This particular arrangement is called a primer:template junction.

It has two components:-

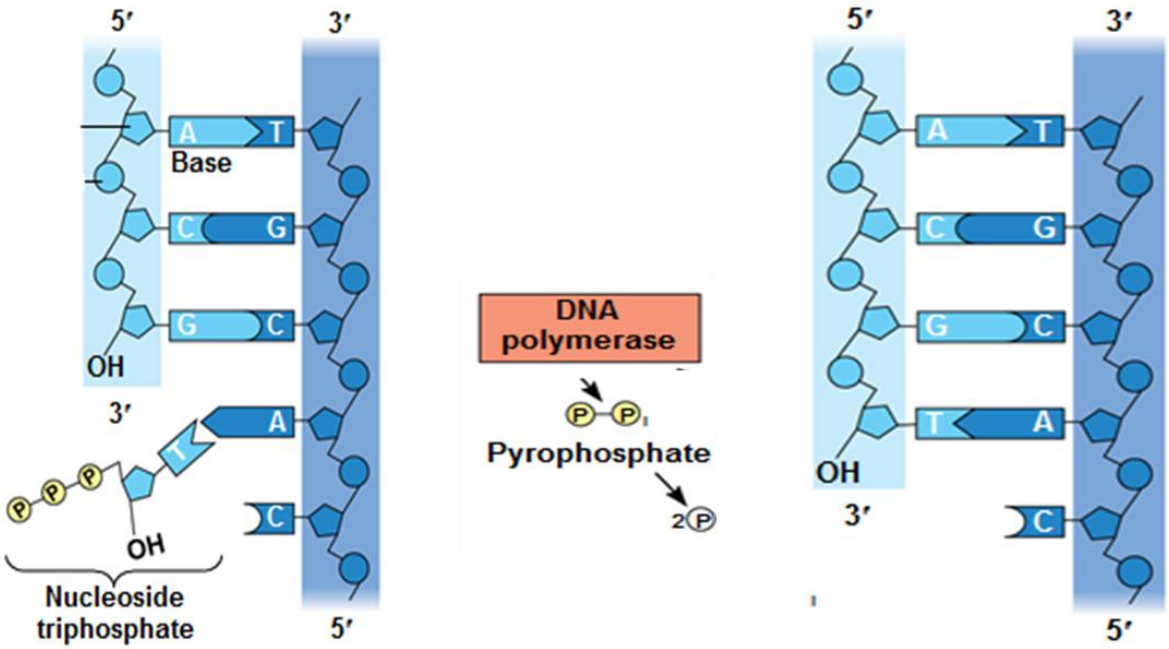
The Template

The Primer



The new chain of DNA grows by extending the 3' end of the primer. The phosphodiester bond is formed in an SN2 reaction. In this reaction, the hydroxyl group of the 3' end of the primer attacks the α-phosphoryl group of the incoming nucleoside triphosphate. The leaving group of

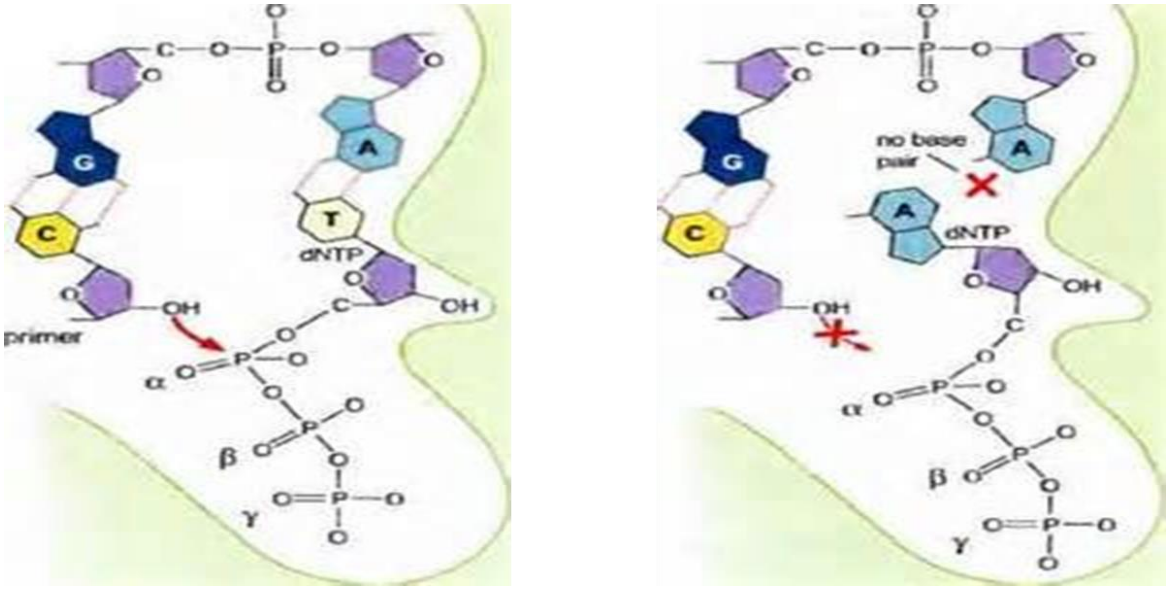
the reaction is pyrophosphate which arises from the release of β- and ɣ- phosphates of the nucleoside.



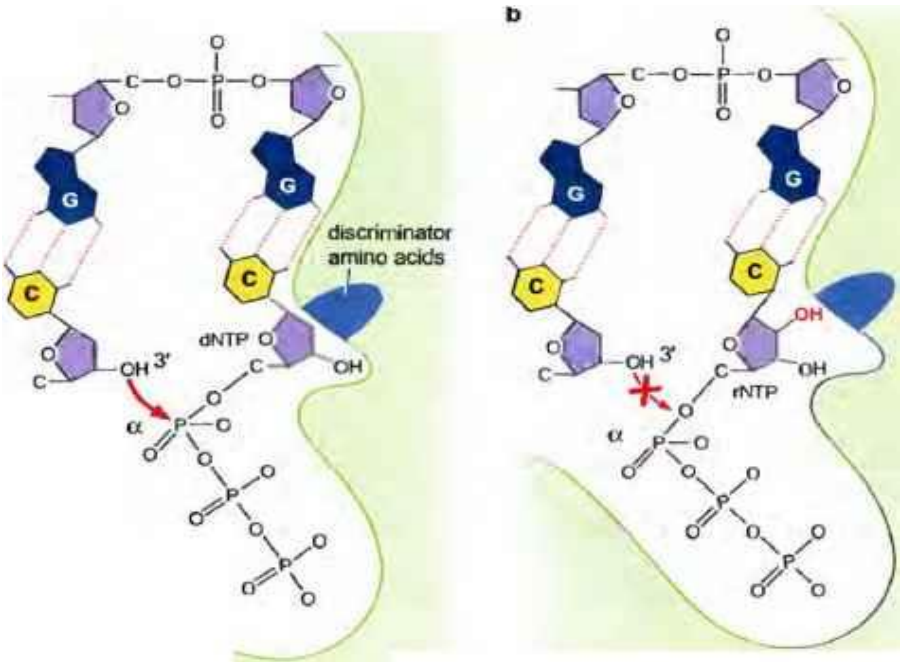
The template strand directs which of the four nucleoside triphosphates is added. The incoming nucleoside triphosphate base pairs with the template strand. The free energy for this reaction is provided by the rapid hydrolysis of the pyrophosphate into two phosphate groups by an enzyme known as pyrophosphatase. The net result of nucleotide addition and pyrophosphate hydrolysis is the simultaneous breaking of two high energy phosphate bonds. Therefore, DNA synthesis is a coupled process. This reaction is highly favourable with high value of K_{eq} which means that its an irreversible reaction.

40. Mechanism of DNA Polymerase

The DNA polymerases are enzymes that create DNA molecules by assembling nucleotides, the building blocks of DNA. These enzymes are essential to DNA replication and usually work in pairs to create two identical DNA strands from a single original DNA molecule. Every time a cell divides, DNA polymerase is required to help duplicate the cell's DNA, so that a copy of the original DNA molecule can be passed to each of the daughter cells. It uses a single active site to catalyze the addition of any of four deoxynucleoside triphosphates. DNA polymerase monitors the ability of the incoming nucleotide to form an A:T or G:C base pair, rather than detecting the exact nucleotide that enters the active site. Only when a correct nucleotide comes, the 3'-OH of the primer and the α-phosphate of the nucleotide align in optimum position for catalysis to take place.

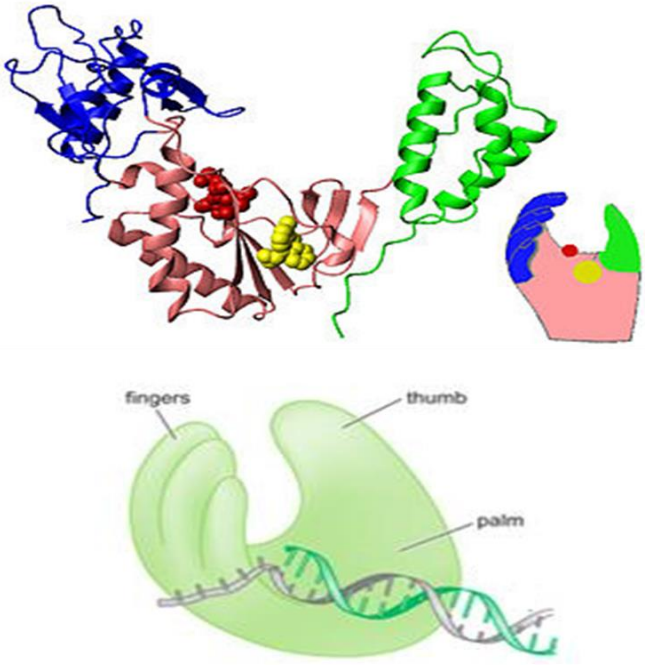


Incorrect base pairing leads to dramatically lower rate of nucleotide addition as a result of catalytically unfavourable alignment of these substrates. DNA polymerase shows an impressive ability to distinguish between ribonucleoside (rNTPs) and deoxyribonucleoside triphosphates (dNTPs). Although rNTPs are present at approx. ten-fold higher concentration in the cell, yet their incorporation rate is 1000-folds lower than dNTPs. This discrimination is mediated by the steric exclusion of rNTPs from the active site of DNA polymerase. In DNA polymerase, the nucleotide-binding pocket cannot accommodate a 2'-OH on the in-coming nucleotide. This space is occupied by two amino acids that make van der Waals contacts with the deoxyribose ring. These amino acids are called discriminator amino acids.



41 & 42. DNA Polymerases Resemble a Hand

The structural studies on DNA polymerases reveal that the DNA substrate sits in a large cleft that resembles a partially closed right hand. Based on the hand analogy, the three domains of the DNA polymerase are called the thumb, fingers and palm.



The palm domain is composed of a β-sheet and contains the primary elements of the catalytic site.

This region of DNA polymerase binds two divalent metal ions (Mg^{2+} or Zn^{2+}). One metal ion reduces the affinity of the 3'-OH for its hydrogen.

This generates a 3'-O_2 that is primed for the nucleophilic attack of the α-phosphate of the incoming dNTP. The second metal ion coordinates the negative charges of the β- and γ-phosphates of the dNTP and stabilizes the pyrophosphate produced by joining the primer and the incoming nucleotide. In addition to its role in catalysis, the palm domain also monitors the base pairing of the most recently added nucleotides. The fingers of the polymerase are also important for catalysis.

Several residues located within the fingers bind to the incoming dNTP. More importantly, once a correct base pair is formed between the incoming dNTP and the template, the finger domain moves to enclose the dNTP.

This closed form of the polymerase "hand" stimulates catalysis. In contrast to the fingers and the palm, the thumb domain is not intimately involved in catalysis.

Instead, it interacts with the DNA that has been most recently synthesized. This serves two purposes:-

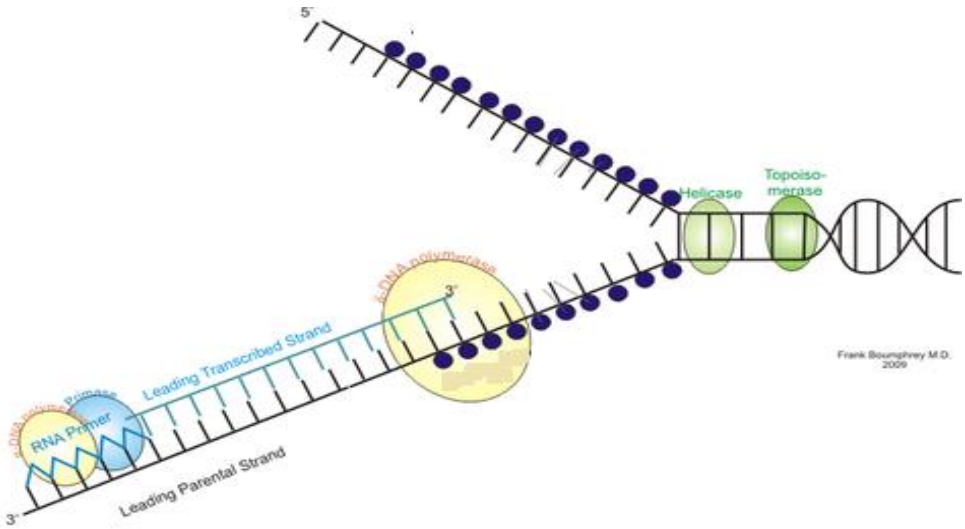
First, it maintains the correct position of the primer and the active site. Second, the thumb helps to maintain a strong association between the DNA polymerase and its substrate.

This association contributes to the ability of the DNA polymerase to add many dNTPs.

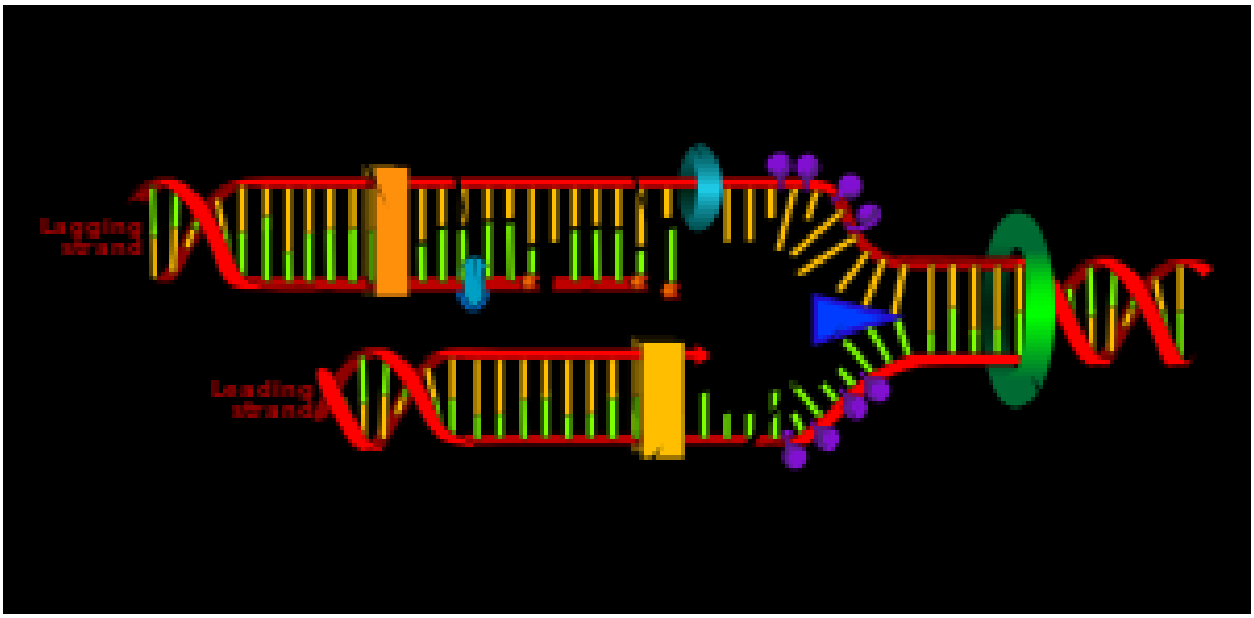
43. THE REPLICATION FORK

In the cell, both strands of the DNA duplex are replicated at the same time. So it requires separation of the two strands of the double helix to create two template DNAs. The junction between the newly separated template strands and the unreplicated duplex DNA is known as the Replication Fork.

The replication fork moves continuously towards the duplex region of unreplicated DNA. As the fork moves, it creates two ssDNA templates that each directs the synthesis of a complementary DNA strand. The antiparallel nature of DNA creates a complication for the simultaneous replication of the two exposed templates at the replication fork. Because DNA is synthesized only by elongating a 3' end, only one of the two exposed templates can be replicated continuously as the replication fork moves. The newly synthesized DNA strand directed by this template is known as the leading strand.



Synthesis of the newDNA strand directed by the other ssDNA template is more complicated.
This template directs the DNA polymerase to move in the opposite direction of the replication fork. The new DNA strand directed by this template is known as the lagging strand.
This strand of DNA must be synthesized in a discontinuous fashion. Synthesis of the lagging strand must wait for movement of the replication fork to expose a substantial length of template before it can be replicated. Each time a substantial length of the template is exposed, DNA synthesis is initiated and continues until it reaches the 5' end of the previous newly synthesized fragment of lagging strand DNA. The resulting short fragments of new DNA formed on the lagging strand are called Okazaki fragments. They vary in length from 1000 to 2000 nucleotides in bacteria and from 100 to 400 nucleotides in eukaryotes.

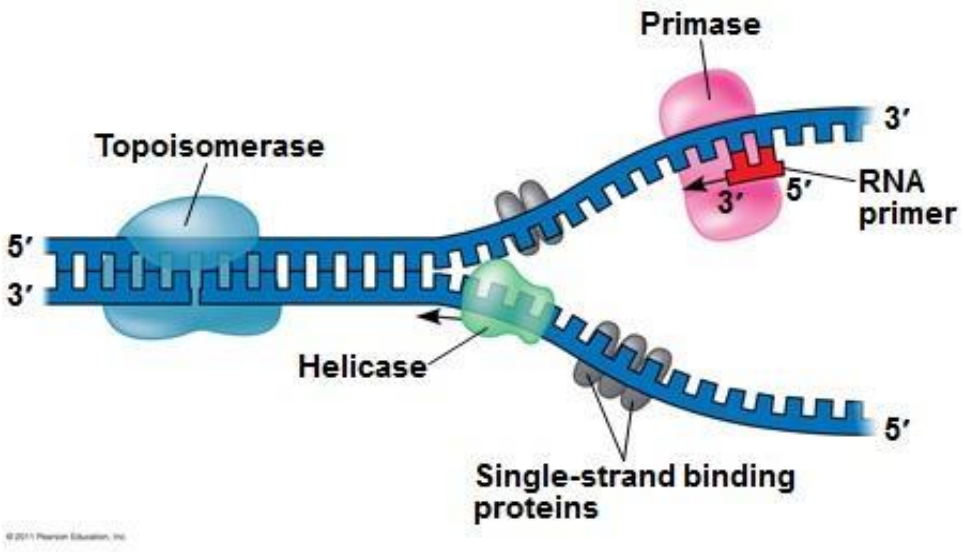


Shortly after being synthesized, Okazaki fragments are covalently joined together to generate a continuous, intact strand of new DNA. Okazaki fragments are therefore transient intermediates in DNA replication.

44. THE RNA PRIMER

All DNA polymerases require a primer with a free 3'-OH. They cannot initiate the synthesis of new DNA strand de novo. To accomplish this, the cell takes advantage of the ability of RNA polymerases to do what DNA polymerases cannot: start new RNA chains de novo. Primase is a specialized RNA polymerase dedicated to making short RNA primers (5–10 nucleotides long) on a ssDNA template. These primers are then extended by DNA polymerase.

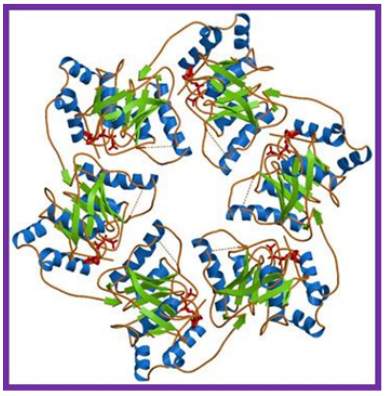
Figure 16.13

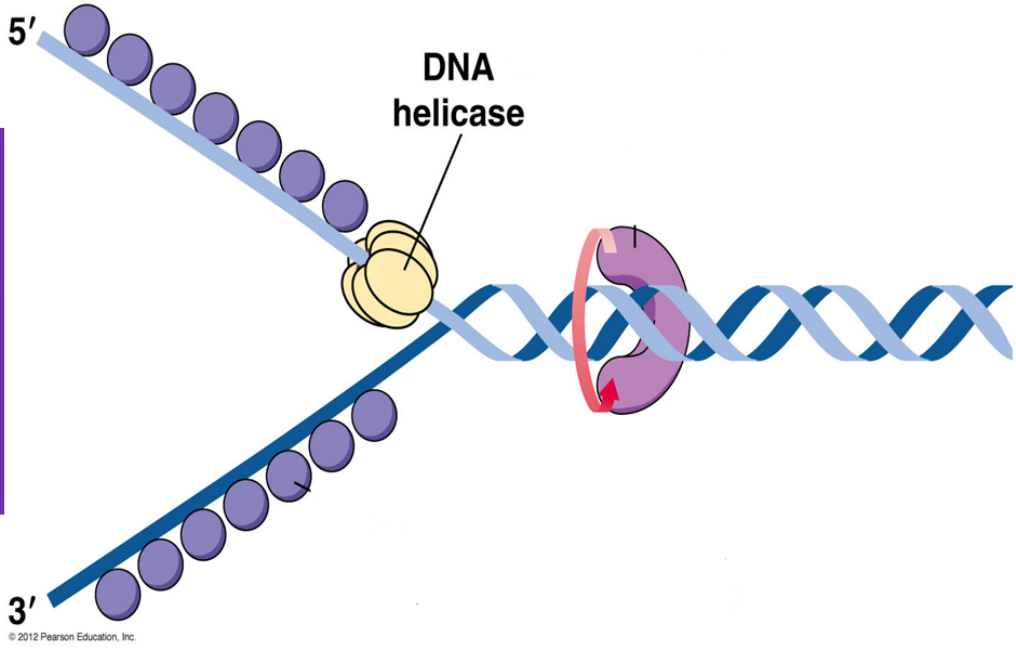


Both the leading and lagging strands require primase to initiate DNA synthesis. Each leading strand requires only a single RNA primer. The discontinuous synthesis of the lagging strand means that new primers are needed for each Okazaki fragment.
Synthesis of the lagging strand can require hundreds of Okazaki fragments and their associated RNA primers. Primase activity is dramatically increased when it associates with another protein that acts at the replication fork called DNA Helicase. This protein unwinds the DNA at the replication fork, creating an ssDNA template that can be acted on by primase.

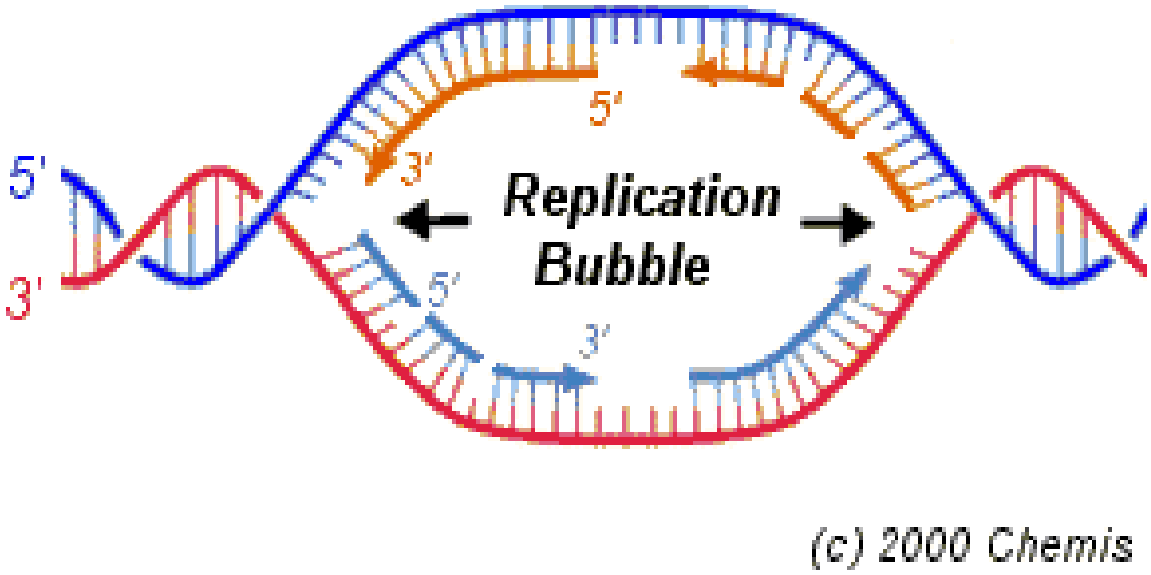
45. THE DNA HELICASE

DNA polymerases are unable to separate the two strands of duplex DNA. Therefore a third class of enzymes, called DNA Helicases catalyze the separation of the two strands of duplex DNA at the replication fork. DNA helicases are hexameric proteins that assume the shape of a ring. This ring encircles one of the two single strands at the replication fork adjacent to the single-stranded:double-stranded junction.





DNA helicases found at replication forks exhibit high processivity because they encircle the DNA. They associate with the DNA and unwind multiple base pairs of DNA. Release of the helicase from the DNA therefore requires the opening of the hexameric protein ring, which is a rare event. However, the helicase can dissociate when it reaches the end of the DNA strand. Of course, this arrangement of enzyme and DNA poses problems for the binding of the DNA helicase to the DNA strand in the first place. Thus, there are specialized mechanisms that open the DNA helicase (hexameric) ring and place it around the DNA before re-forming the ring. Each DNA helicase moves along ssDNA in a defined direction This property is referred to as the polarity of the DNA helicase. DNA helicases can have a polarity of either 5’ 3’ or 3’ 5’. This direction is always defined according to the strand of DNA bound rather than the strand that is displaced.

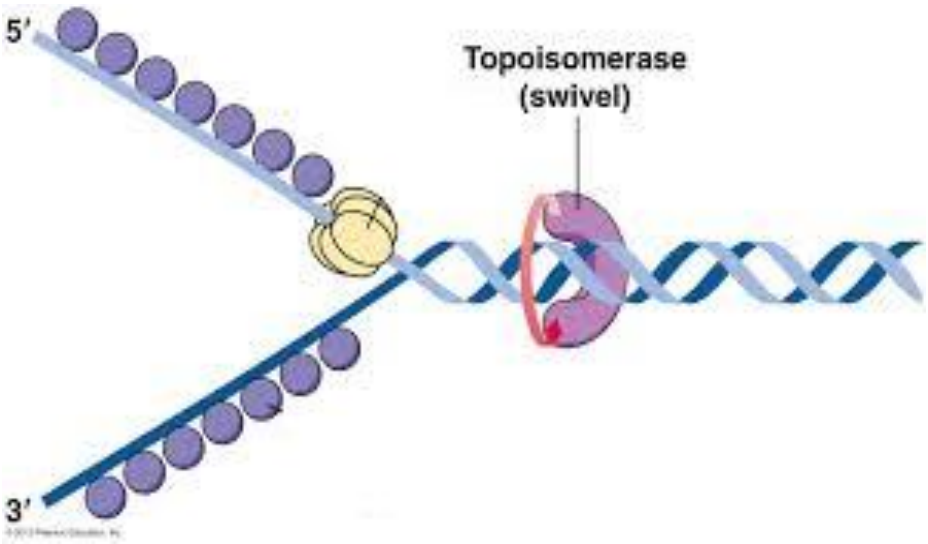


46. TOPOISOMERASES

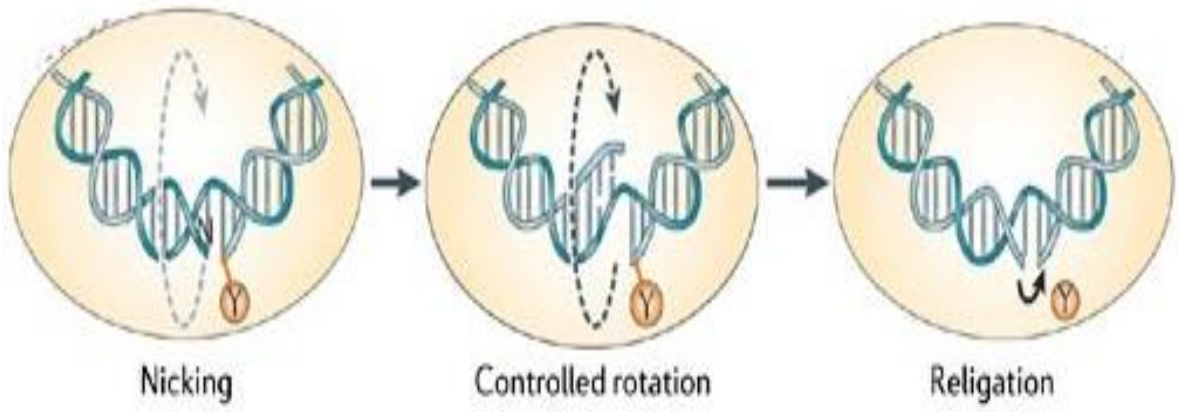
Topoisomerases are enzymes that regulate the overwinding or underwinding of DNA. The winding problem of DNA arises due to the intertwined nature of its double-helical structure. During DNA replication and transcription, DNA becomes overwound ahead of a replication fork. If left unabated, this tension would eventually stop the ability of RNA & DNA polymerase involved in these processes to continue down the DNA strand.

In order to prevent and correct these types of topological problems caused by the double helix, topoisomerases bind to either single-stranded or double-stranded DNA and cut the phosphate backbone of the DNA. This intermediate break allows the DNA to be untangled or unwound, and, at the end of these processes, the DNA backbone is resealed again. Since the overall chemical composition and connectivity of the DNA do not change, the tangled and untangled DNAs are chemical isomers, differing only in their global topology, thus their name. Topoisomerases are isomerase enzymes that act on thetopology of DNA.

Bacterial topoisomerase and human topoisomerase proceed via the same mechanism for replication and transcription.



These enzymes do this by breaking either one or both strands of the DNA without letting go of the DNA and passing the same number of DNA strands through the break.



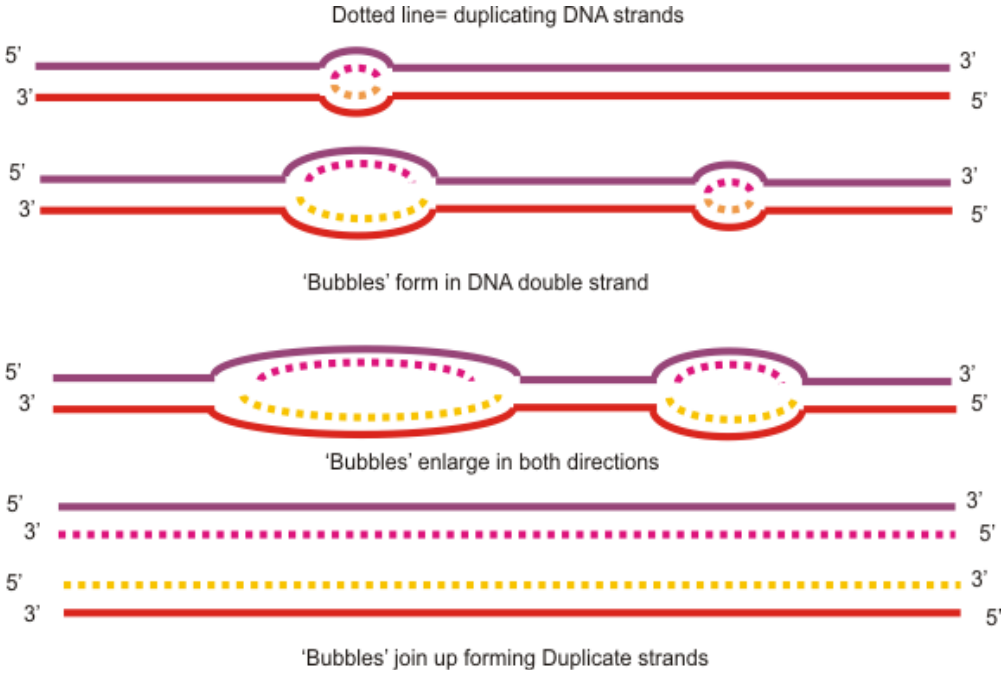
This action relieves the accumulation of supercoils. In this way, topoisomerases act as a "swivelase" that prevents the accumulation of supercoils ahead of the replication fork.

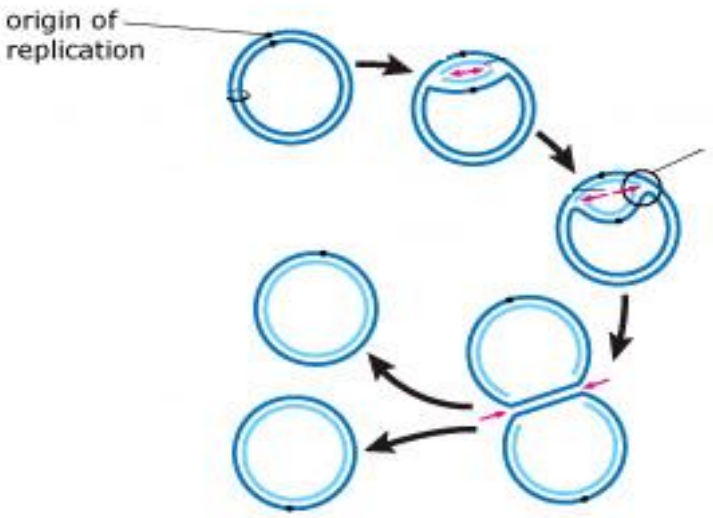
47. INITIATION OF REPLICATION

For a cell to divide, it must first replicate its DNA. This process is initiated at particular points in the DNA, known as "origins", which are targeted by initiator proteins. In *E. coli* this protein is DnaA; in yeast, this is the origin recognition complex. Sequences used by initiator proteins

tend to be "AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) which are easier to unzip. Once the origin has been located, these initiators recruit other proteins and form the pre-replication complex, which unzips the double-stranded DNA.

The initial formation of a replication fork requires the separation of the two strands of the DNA duplex to provide the ssDNA. ssDNA is required for DNA helicase binding and to act as a template for the synthesis of both the RNA primer and new DNA. Although DNA strand separation (DNA unwinding) is most easily accomplished at chromosome ends, but DNA synthesis generally initiates at internal regions. As the circular chromosomes lack the chromosome ends so it makes internal DNA unwinding essential for the replication initiation. The specific sites at which DNA unwinding and initiation of replication occur are called Origins of Replication. Depending on the organism, there may be as few as one or as many as thousands of origins per chromosome.





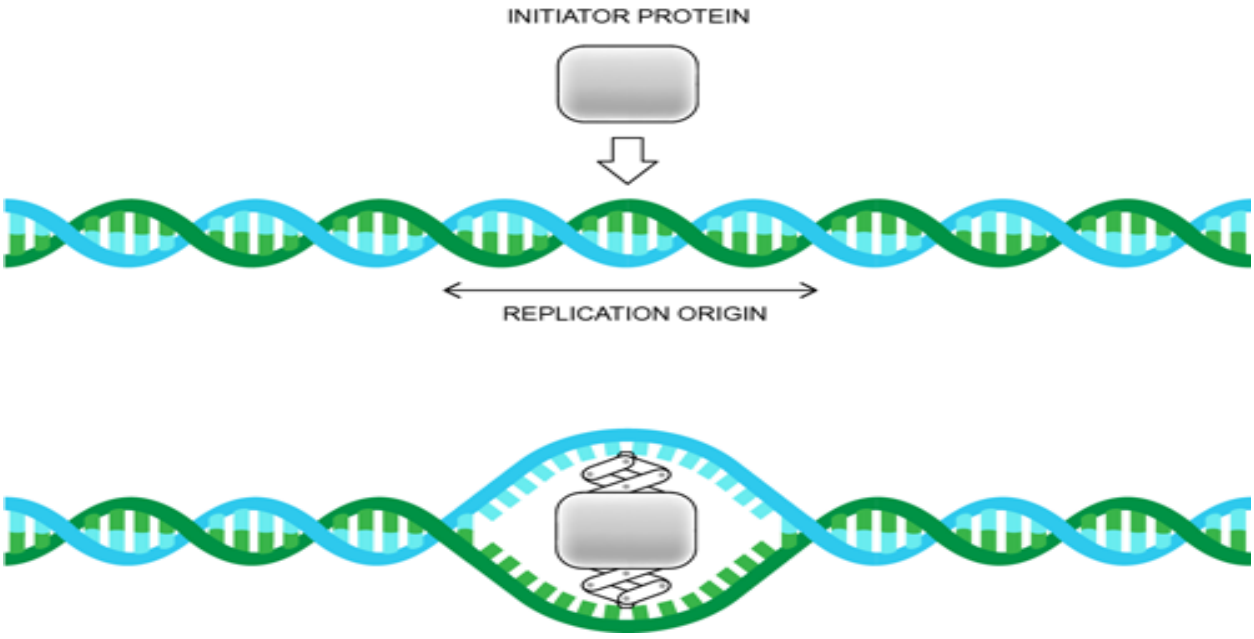
From *Biology* by Campbell and Reece

48. THE REPLICON MODEL

Initiation of replication was explained by Francois Jacob, Sydney Brenner and Jacques Cuzin in 1963, which is called as The Replicon Model of replication Initiation. They defined all of the DNA replicated from a particular origin of replication as a replicon. As the single chromosome found in *E. coli* cells has only one origin of replication, the entire chromosome is a single replicon. In contrast, the presence of multiple origins of replication divides each eukaryotic chromosome into multiple replicons; one for each origin of replication.

The replicon model proposed two components that controlled the initiation of replication; the replicator and the initiator. The replicator is defined as the cis-acting DNA sequences that are sufficient to direct the initiation of DNA replication. This is in contrast to the origin of replication, which is the physical site on the DNA where the DNA is unwound and DNA synthesis initiates. Although the origin of replication is always part of the replicator, sometimes the origin of replication is only a fraction of the DNA sequences required to direct the initiation of replication (the replicator).

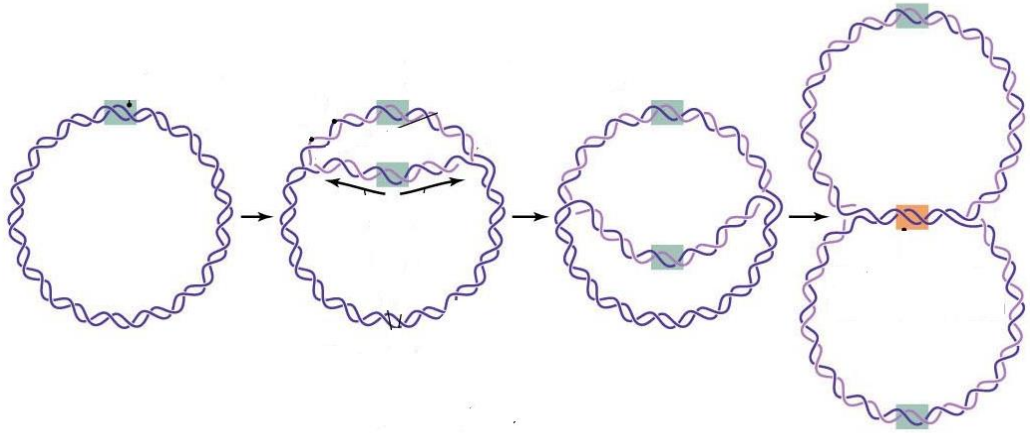
The second component of the replicon model is the initiator protein. This protein specifically recognizes a DNA element in the replicator and activates the initiation of replication. All initiator proteins select the sites that will become origins of replication. The initiator protein is the only sequence-specific DNA binding protein involved in the initiation of replication.



All the remaining proteins other than initiator protein, required for replication initiation do not bind to a DNA sequence specifically.

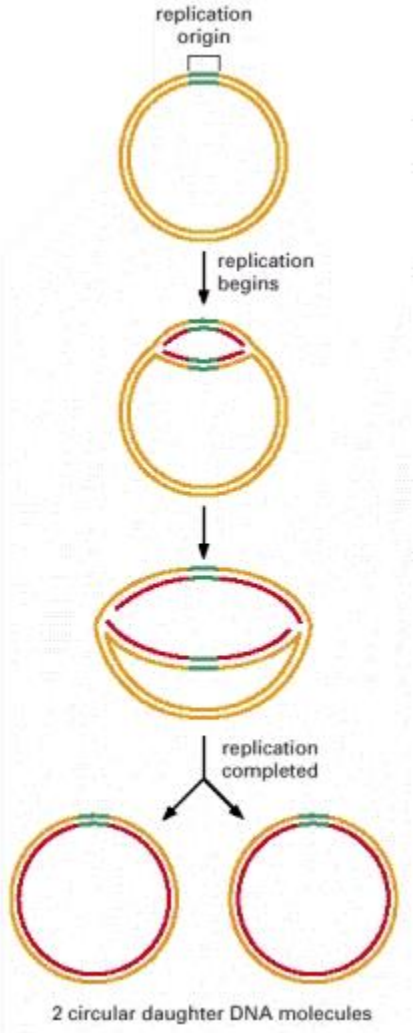
49. FINISHING REPLICATION

Completion of DNA replication requires a set of specific events. These events are different for circular chromosomes and linear chromosomes. In case of circular chromosome, the conventional replication fork machinery replicates the entire molecule, but the resulting daughter molecules are topologically linked to each other.



While in case of linear chromosome, the replication fork machinery cannot complete replication of the very ends of linear chromosomes. Therefore, organisms containing linear chromosomes have developed novel strategies to replicate their chromosome ends. After replication of a circular chromosome is complete, the resulting daughter DNA molecules remain linked together as catenanes.

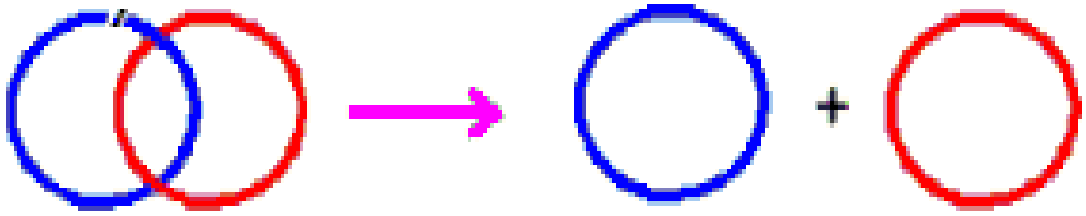
While in case of linear chromosome, the replication fork machinery cannot complete replication of the very ends of linear chromosomes.



50. TYPE II TOPOISOMERASES

Type IIA topoisomerases are essential in the separation of entangled daughter strands during replication. This function is believed to be performed by topoisomerase II in eukaryotes and by topoisomerase IV in prokaryotes. Failure to separate these strands leads to cell death. Type IIA topoisomerases have the special ability to relax DNA to a state below that of thermodynamic equilibrium, a feature unlike type IA, IB, and IIB topoisomerases. This ability, known as topology simplification, was first identified by Rybenkov et al. (Science 1997). The hydrolysis of ATP drives this simplification, but a clear molecular mechanism for this simplification is still lacking.

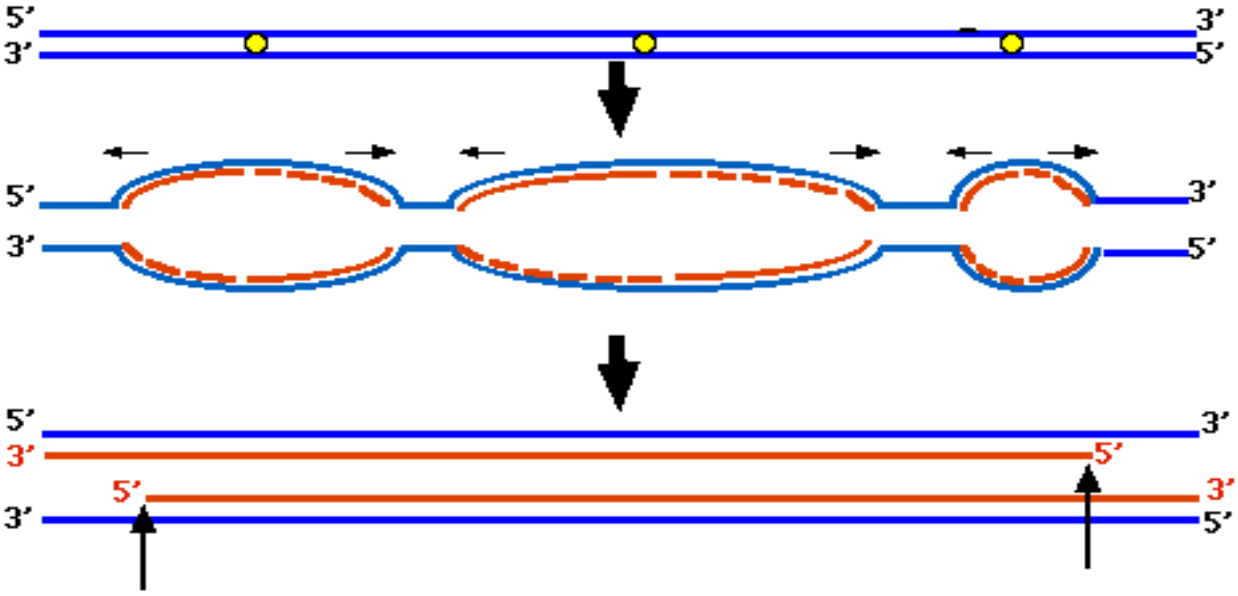
Type II Topoisomerases are the enzymes which have the ability to break a dsDNA molecule and pass a second dsDNA molecule through this break.



So this reaction can easily decatenate the two circular daughter chromosomes and allow their segregation into separate cells. The activity of type II topoisomerases is also critical to the segregation of large linear molecules. Although there is no inherent topological linkage after the replication of a linear molecule, the large sized chromosomes necessitates the intricate folding of the DNA into loops which are attached to a protein scaffold. These attachments lead to many of the same problems that circular chromosomes have after replication.
So type II topoisomerases allow these linked DNAs to be separated. So as in the case of circular chromosomes, type II topoisomerases also allow these linked DNAs to be separated.

51. TELOMERASE

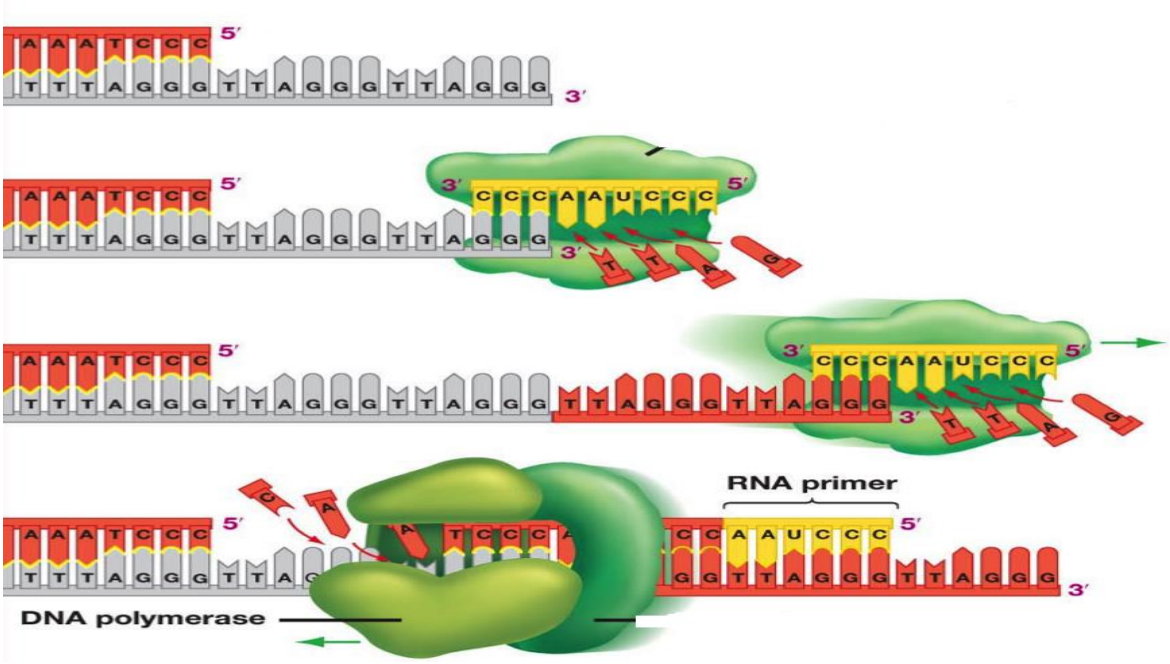
The requirement for an RNA primer to initiate all new DNA synthesis creates a dilemma for the replication of the ends of linear chromosomes. This is called as the end replication problem. This difficulty is not observed during the duplication of the leading-strand template. Because it requires only one RNA primer which completes the DNA synthesis up to extreme terminus of the strand. In contrast, the requirement for multiple primers to complete lagging-strand synthesis means that a complete copy of its template cannot be made. Even if the end of the last RNA primer for Okazaki fragment synthesis anneals to the final base of the lagging-strand template. Once this RNA molecule is removed, there will remain a short region (the size of the RNA primer) of un-replicated ssDNA at the end of the chromosome.



Telomerase, also called telomere terminal transferase,[1] is aribonucleoprotein that adds the polynucleotide "TTAGGG" to the 3' end oftelomeres, which are found at the ends of eukaryotic chromosomes. A telomere is a region of repetitive sequences at each end of a chromatid, which protects the end of the chromosome from deterioration or from fusion with neighbouring chromosomes.

Telomerase is a reverse transcriptase enzyme that carries its own RNAmolecule (with the pattern of "CCCAAUCCC" in vertebrates), which is used as a template when it elongates telomeres.

Like all other DNA polymerases, telomerase acts to extend the 3’ end of its DNA substrate. But unlike most DNA polymerases, telomerase does not need an exogenous DNA template to direct the addition of new dNTPs. Instead, the RNA component of telomerase serves as the template for adding the telomeric sequence to the 3’ terminus at the end of the chromosome. Telomerase specifically elongates the 3’-OH of telomeric ssDNA sequences using its own RNA as a template.
As a result, the newly synthesized DNA is single-stranded. So when telomerase acts on the 3’ end of the telomere, it extends only one of the two strands of the chromosome. This is accomplished by the lagging-strand DNA replication machinery.
By providing an extended 3’ end, telomerase provides additional template for the lagging-strand replication machinery. By synthesizing and extending RNA primers using the telomerase extended 3’ end as a template, the cell can effectively increase the length of the 5’ end of the chromosome as well.



Chapter 8. DNA Mutation

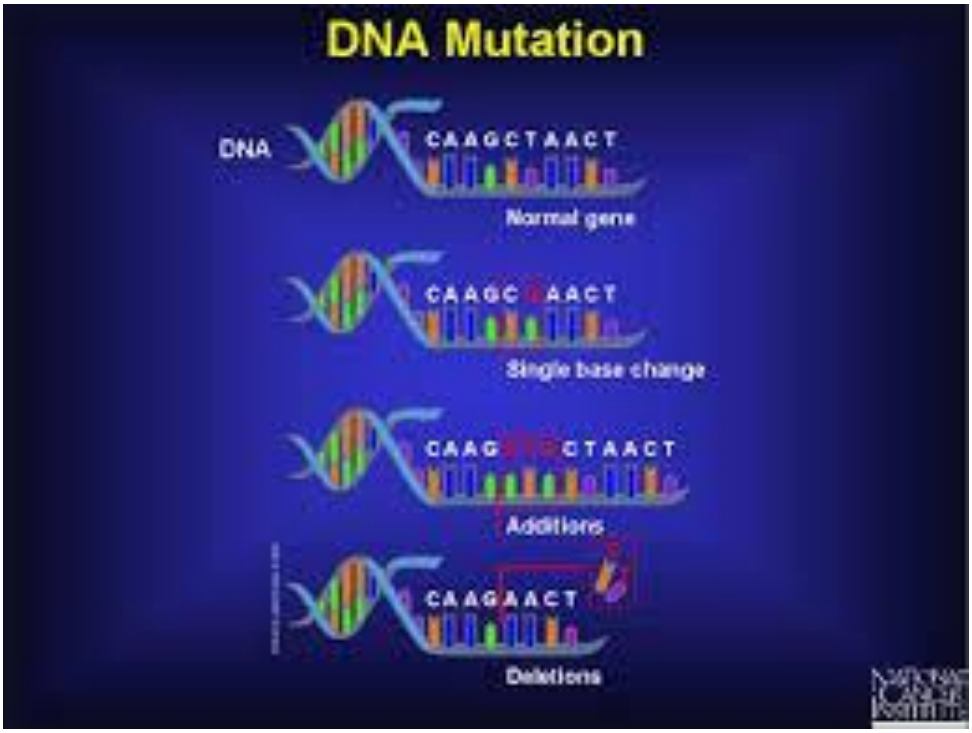
52. DNA MUTATIONS

Mutation is a permanent alteration of the nucleotide sequence of the genome of an organism, virus, or extrachromosomal DNA or other genetic elements. Mutations result from damage to DNA which is not repaired, errors in the process of replication, or from the insertion or deletion of segments of DNA by mobile genetic elements. Mutations may or may not produce discernible changes in the observable characteristics (phenotype) of an organism. Mutations play a part in both normal and abnormal biological processes including: evolution, cancer, and the development of the immune system, including junctional diversity.

Mutation can result in many different types of change in sequences. Mutations in genes can either have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely. Mutations can also occur in nongenic regions.

DNA can be easily damaged even under normal physiological conditions. Many different kinds of chemical and physical agents can damage DNA. Some of these agents are endogenous which are produced inside the cells as a result of normal metabolic pathways. While some others are exogenous agents which come from the surrounding environment. On one hand, DNA stability is required to ensure that the genetic information may pass accurately from one generation to the next. It is also require for the correct functioning of thousands of genes. On the other hand the

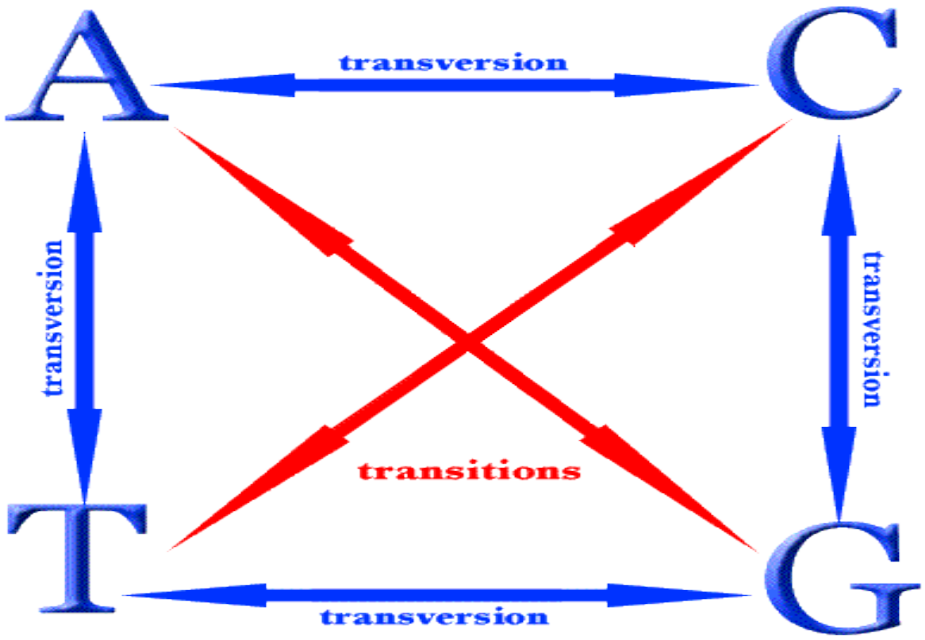
genetic variation is needed to drive evolution. If this variation would be lacking, the new species, including humans, would have not arisen. So the life and biodiversity depend on a happy balance between DNA damage (mutation) and its repair.



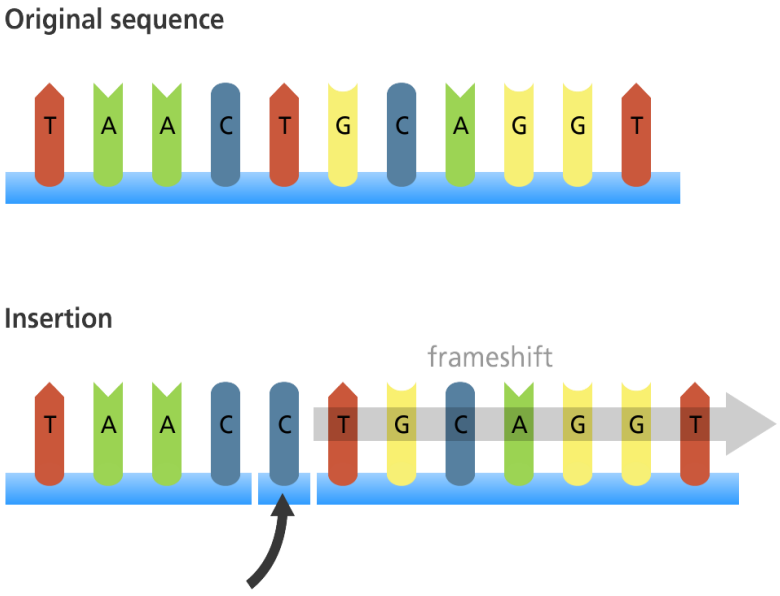
53. NATURE OF MUTATIONS

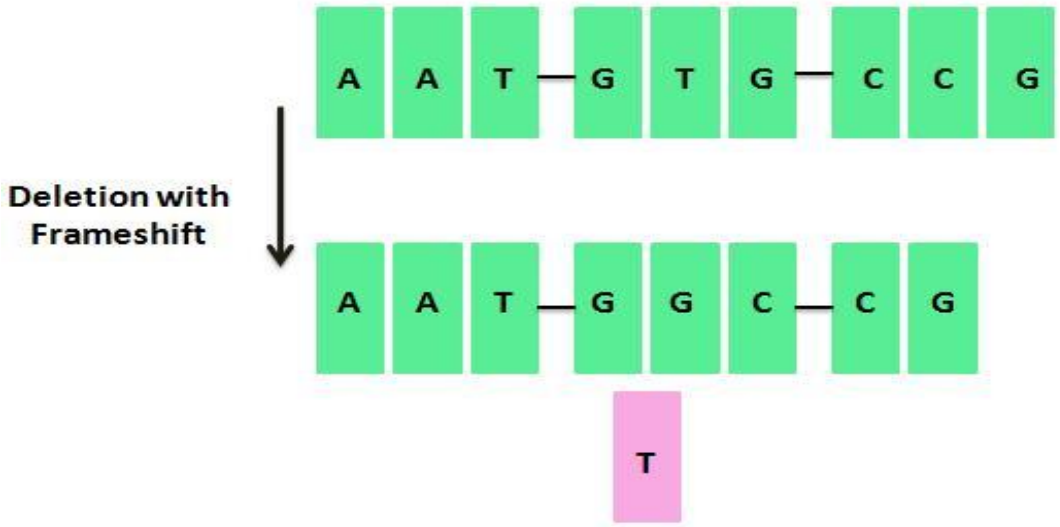
DNA mutations may be very simple (single base change) or very complex and including several thousands of nucleotides. The simplest mutations are switches of one base for another.
There are two kinds of such mutations which include:-
Transitions
Transversions

Transitions are pyrimidine-to-pyrimidine and purine-to-purine substitutions, such as thymine (T) to cytosine (C) and adenine (A) to guanine (G). Transversions are pyrimidine-to-purine and purine-to-pyrimidine substitutions, such as T to G or A and A to C or T.

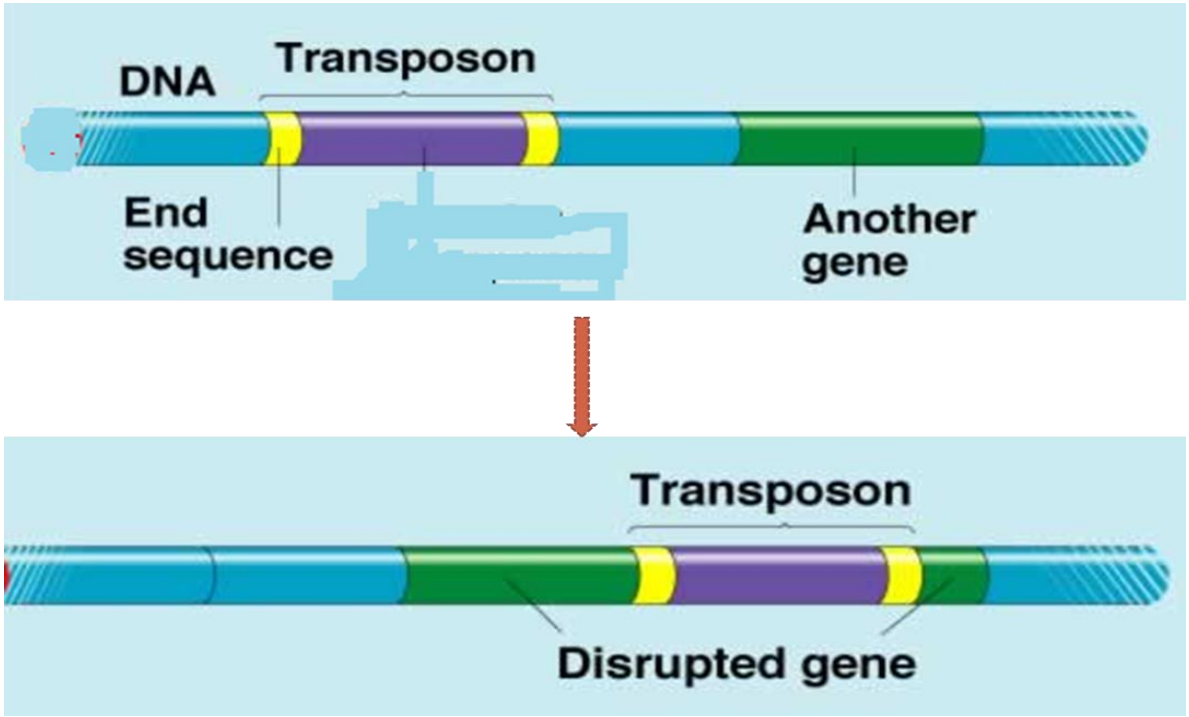


Other simple mutations are insertions or deletions of a nucleotide or a small number of nucleotides.

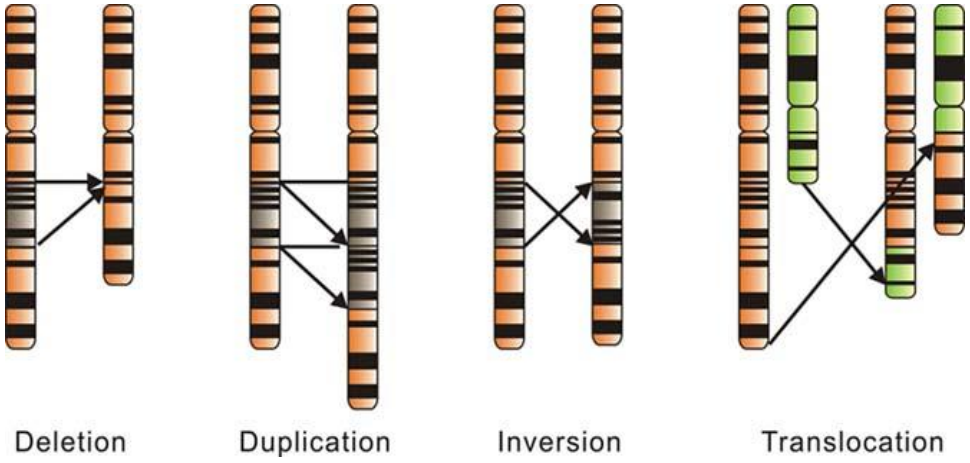




All such mutations that alter a single nucleotide are called point mutations. Other kinds of mutations cause more drastic changes in DNA, such as extensive insertions and deletions and gross rearrangements of chromosome structure. Such changes might be caused, for example, by the insertion of a transposon, which typically places many thousands of nucleotides of foreign DNA in the coding or regulatory sequences of a gene.



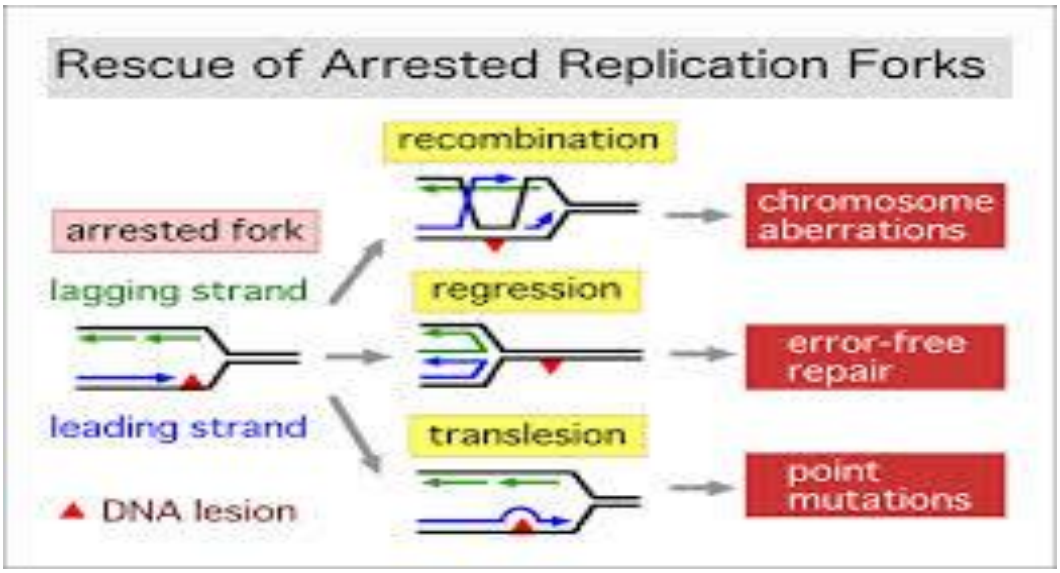
Another type of mutations which are more drastic occur at chromosomal levels. These are changes in appearance of the individual chromosomes through mutation-induced rearrangements.

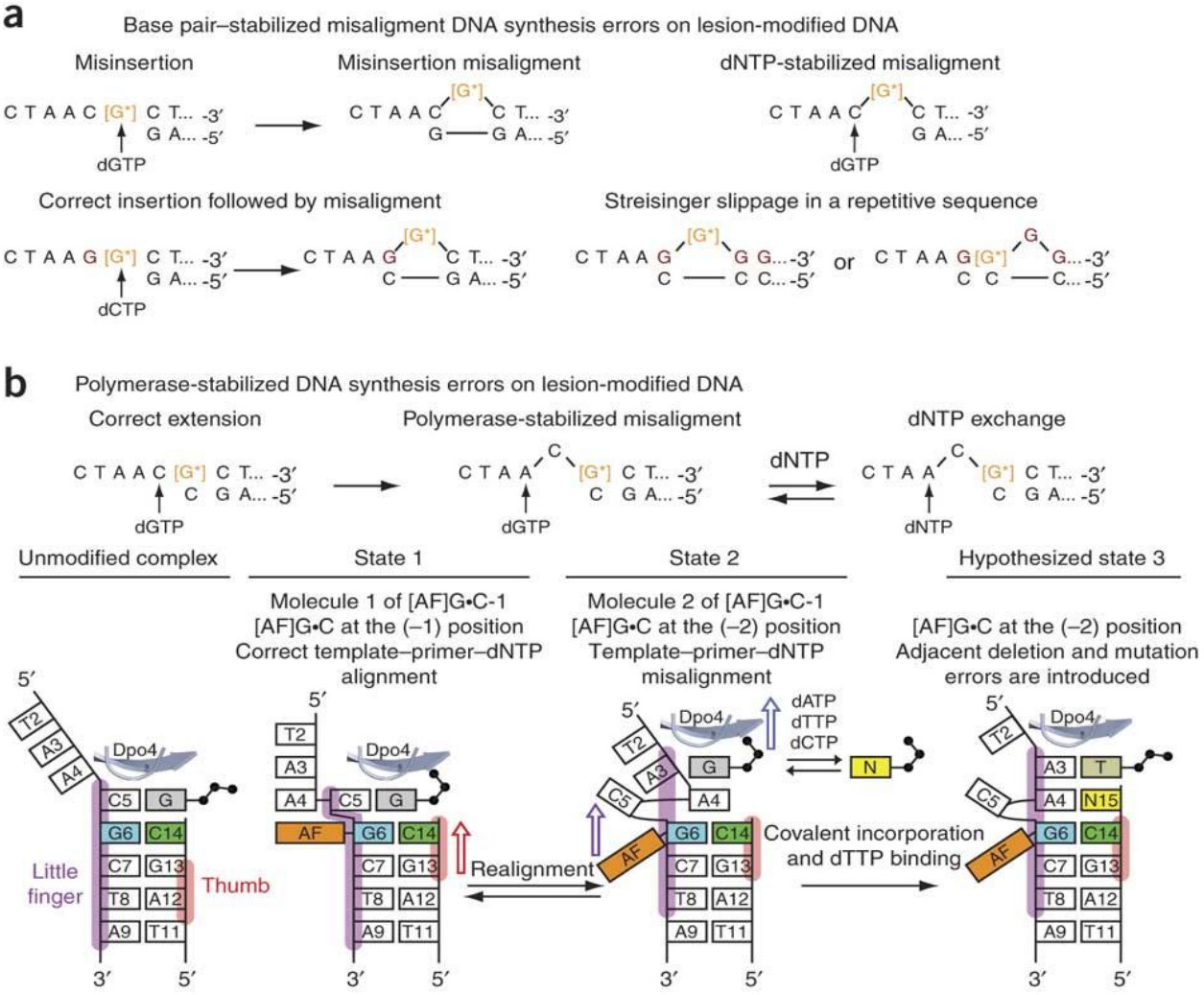


54. REPLICATION ERRORS

The replication machinery achieves a remarkably high degree of accuracy using a proofreading mechanism, which removes wrongly incorporated nucleotides. However, this proofreading is not foolproof.

Some mis-incorporated nucleotides escape detection and become a mismatch between the newly synthesized strand and the template strand. If the misincorporated nucleotide is not subsequently detected and replaced, the sequence change will become permanent in the genome. During a second round of replication, the mis-incorporated nucleotide will direct the incorporation of its complementary nucleotide into the newly synthesized strand. At this point, the mismatch will no longer exist; instead, it will have resulted in a permanent change (a mutation) in the DNA sequence.


Rescue of Arrested Replication Forks
recombination
arrested fork
chromosome aberrations
lagging strand
regression
error-free repair
leading strand
translesion
point mutations
DNA lesion

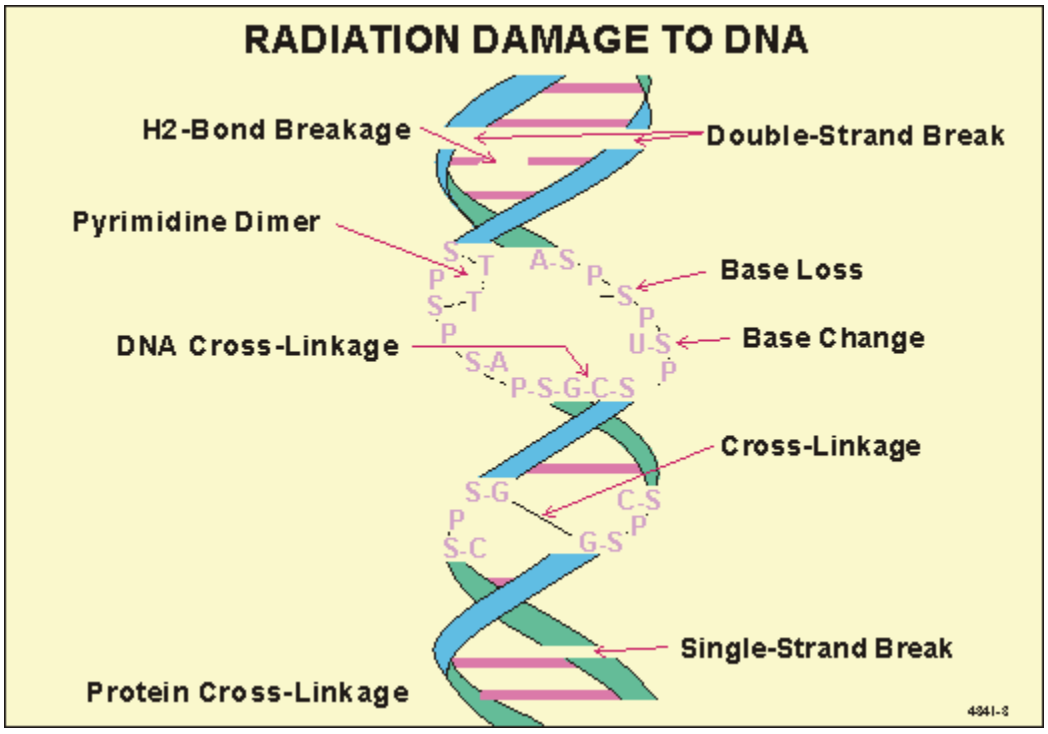

a
Base pair–stabilized misalignment DNA synthesis errors on lesion-modified DNA
Misinsertion
Misinsertion misalignment
dNTP-stabilized misalignment
dGTP
Correct insertion followed by misalignment
Streisinger slippage in a repetitive sequence
dCTP
or
b
Polymerase-stabilized DNA synthesis errors on lesion-modified DNA
Correct extension
Polymerase-stabilized misalignment
dNTP exchange
dNTP
Unmodified complex
State 1
State 2
Hypothesized state 3
Molecule 1 of [AF]G•C-1
[AF]G•C at the (–1) position
Correct template–primer–dNTP alignment
Molecule 2 of [AF]G•C-1
[AF]G•C at the (–2) position
Template–primer–dNTP misalignment
[AF]G•C at the (–2) position
Adjacent deletion and mutation errors are introduced
Dpo4
Little finger
Thumb
Realignment
dATP
dTTP
dCTP
Covalent incorporation and dTTP binding

55. RADIATION DAMAGE

Cells are exposed to three types of high energy electromagnetic radiations which can damage their DNA.
These radiations include:-
Ultraviolet light (wavelength 100-400nm)
X-rays (wavelength 0.01-100nm)
Gamma rays (wavelength <0.01nm)
Later two are ionizing radiations.
Ultraviolet light is divided into three bands:

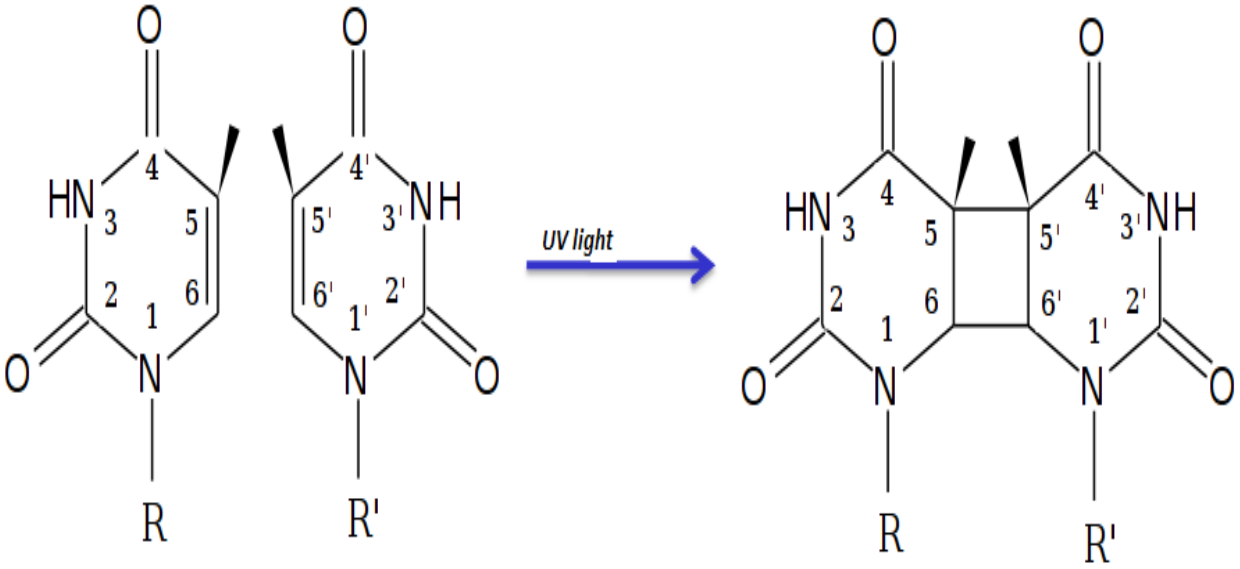
- UV –A (321-400nm)
- UV –B (296-320nm)
- UV –C (100-295nm)

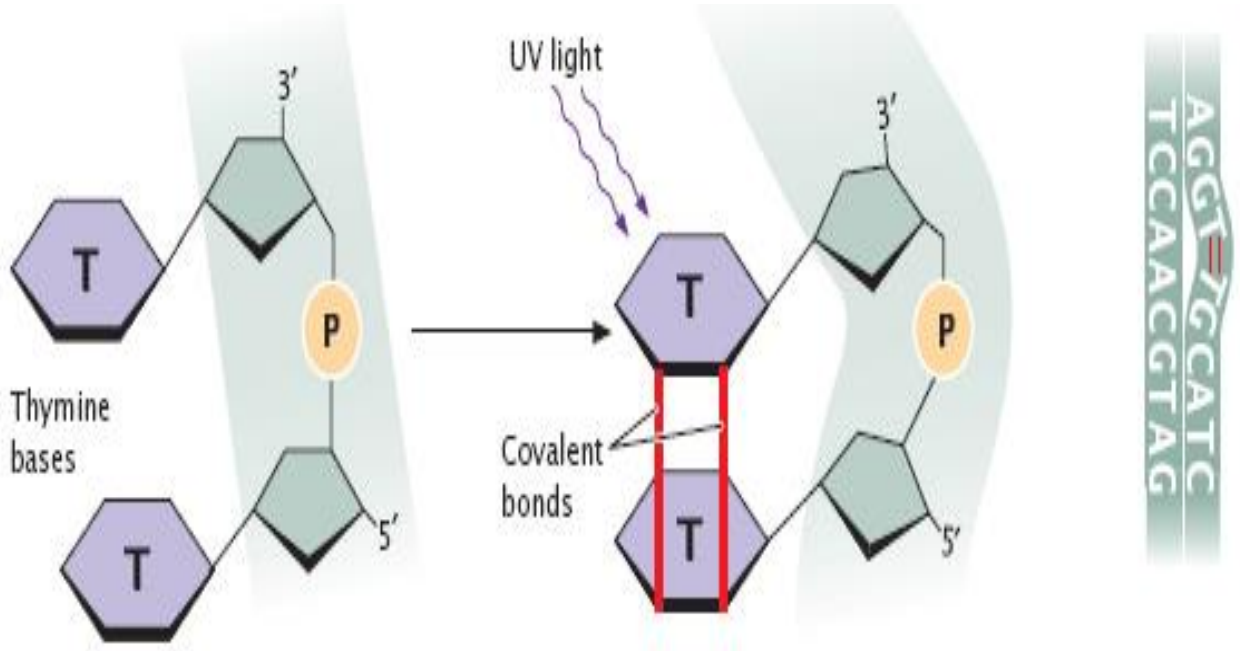
The majority of UV light reaching on earth is UV – A. This is least energetic band and so does little damage to DNA. UV – B accounts for about 10% of the UV radiation reaching the earth's surface. It is responsible for most of the DNA damage in the skin. UV – C includes the wavelength of maximum DNA absorbance (260nm). So it would cause a great deal of DNA damage to exposed organisms if it were able to penetrate the earth's surface. Fortunately, very little UV - C reaches the earth's surface because the ozone layer prevents it from penetration. However, during lab studies, the germicidal lamps that produce UV - C light are used.



56. Cyclobutane pyrimidine Dimer

Pyrimidine dimers are molecular lesions formed from thymine orcytosine bases in DNA via photochemical reactions.[1][2] Ultraviolet light induces the formation of covalent linkages by reactions localized on the C=C double bonds. Two major photoproducts account for nearly all of the UV - induced DNA damage. Which involve the dimer formation between adjacent pyrimidine bases on the same DNA strand. The first pathway includes the formation of cyclobutane pyrimidine dimer (CPD) which accounts for about 75% of all the UV - induced damage. The cyclobutane ring is generated by forming one bond between C-5 atoms and another between C-6 atoms on adjacent pyrimidine rings.





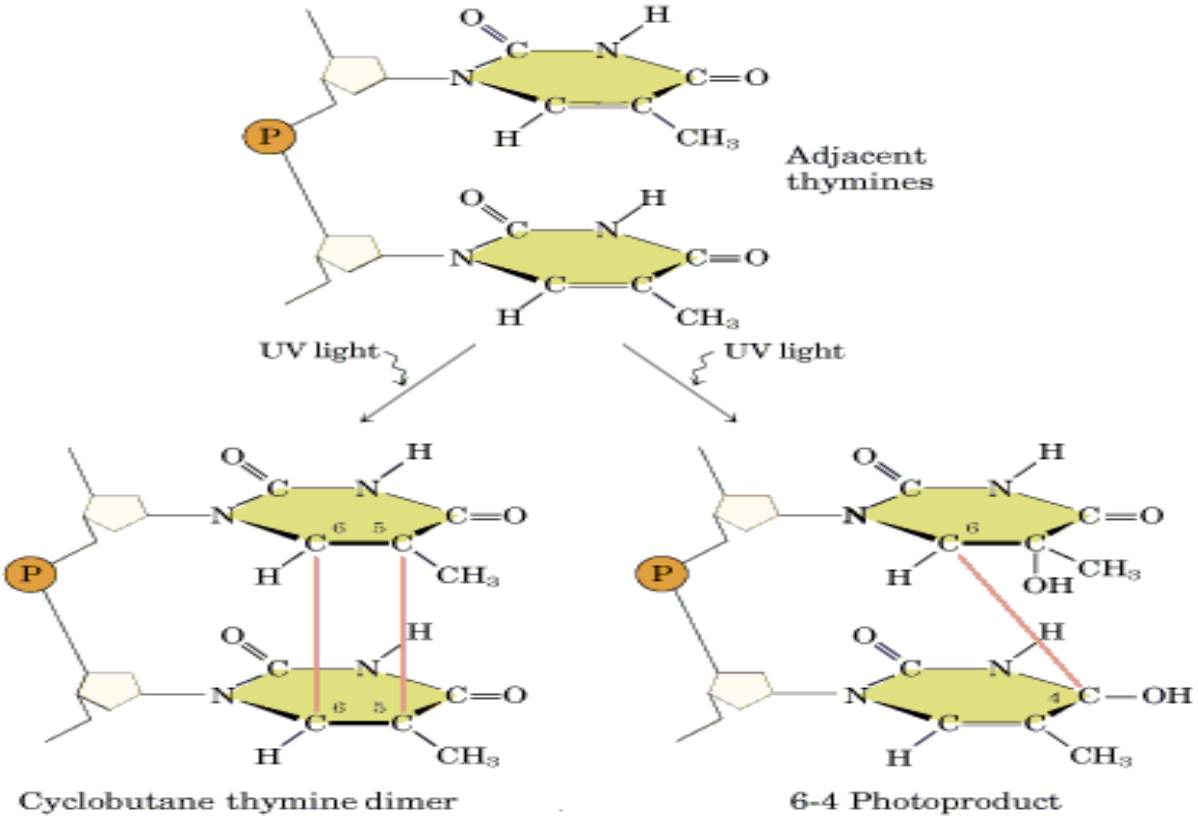
The most common cyclobutane pyrimidine dimer is the thymine-thymine (T< >T) dimer. Cytosine-thymine (C< >T) and cytosine- cytosine (C< >C) dimers are also form but at slower rates. Structural studies show that:-

- 1) B-DNA can accommodate a single T< >T dimer forcing the helical axis to bend by about 30˚ towards the major groove.
- 2) The dimer's 3'-thymine can form a normal base pairing with its adenine partner on the complementary strand.
- 3) The interaction between the 5'-thymine and its complementary adenine partner will be weaker than normal because a single hydrogen bond can be formed here.

Thymine-thymine dimers are often used to study DNA repair systems because they are stable, easy to form and easy to detect.

57. (6-4) Photoproducts

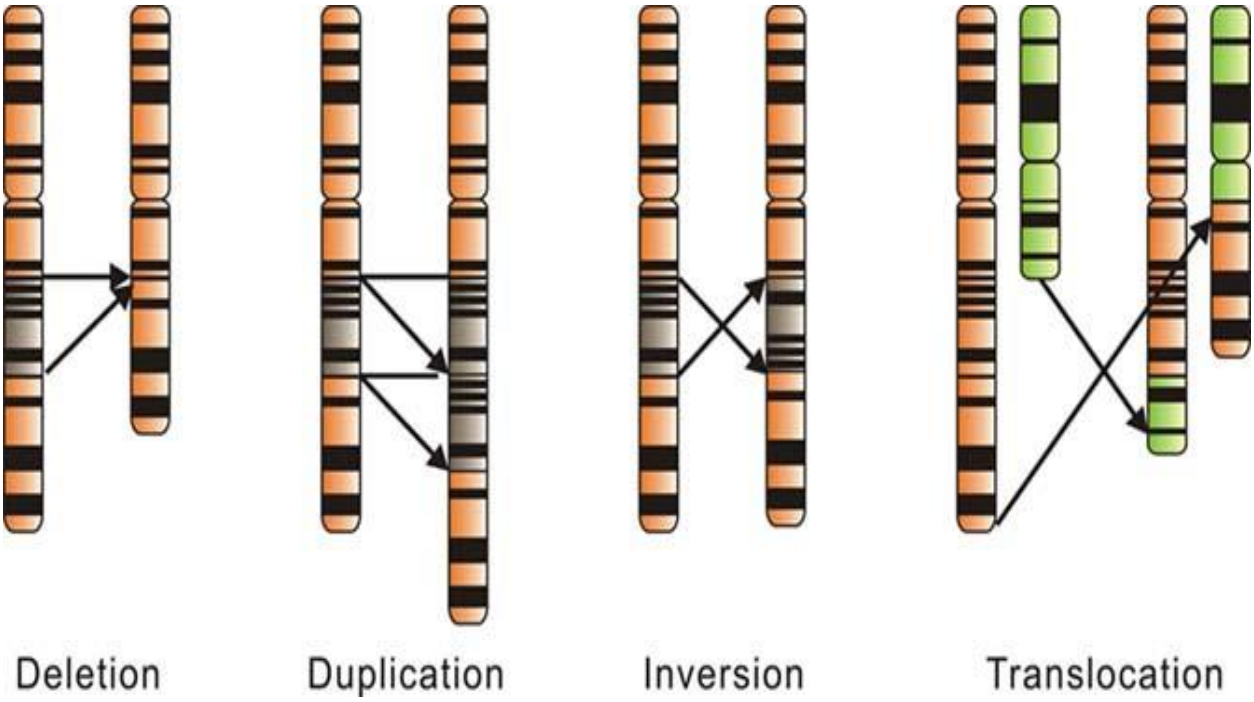
The second pathway, which accounts for most of the remaining UV-induced DNA damage, produces (6-4) photoproducts. A bond is formed between the C-6 atom of the 3'-pyrimidine (either thymine or cytosine) and the C-4 atom of the 5'-pyrimidine (usually cytosine).



The (6-4) photoproduct causes a major distortion in B-DNA because the two pyrimidine rings are perpendicular to each other. If not removed, a pyrimidine dimer or (6-4) photoproduct can interfere with the normal operation of the replication and transcription machinery. This interference results in mutations and cell death. Even if the lesion is removed, the result can be a mutation.

58. X-rays & gamma rays damage

Gamma radiations and X-rays (ionizing radiation) are particularly hazardous because they cause double-strand breaks in the DNA, which are difficult to repair. Ionizing radiations directly or indirectly generate many different kinds of DNA lesions.
Direct damage takes place when DNA or water bound to it absorbs the radiation. Indirect damage takes place when water molecule or any other molecule surrounding the DNA absorb the radiation and form reactive species that then damage the DNA. The DNA lesions caused due to these radiations may be isolated or clustered.
Many lesions within a few helical turns are called clustered lesions. One type of clustered lesion, the double-strand break, is generally thought to be the primary reason that ionization radiation is so lethal to cells. Double-strand breaks are also responsible for various chromosomal aberrations such as deletions, duplications, inversions and translocations.



About 65% of the DNA damage caused by these radiations is due to indirect effects, primarily due to transfer of photons to water. The photon transfer activates the water and causing it to undergo two types of primary reactions.
In the first of these reactions, the water molecule is ionized.

$H_2O \rightarrow H_2O^{\bullet+} + e^-$

The $H_2O^{\bullet+}$ that is formed readily dissociates and release a proton and a hydroxyl radical ($^{\bullet}OH$).

$H_2O^{\bullet+} \rightarrow H^+ + {}^{\bullet}OH$

The electron generated by the first reaction can combine with molecular oxygen to form a super-oxide radical ($O_2^{\bullet-}$).

$e^- + O_2 \rightarrow O_2^{\bullet-}$

In the second type of primary reaction, excited water molecule (H_2O^*) splits into a hydrogen atom and a hydroxyl radical.

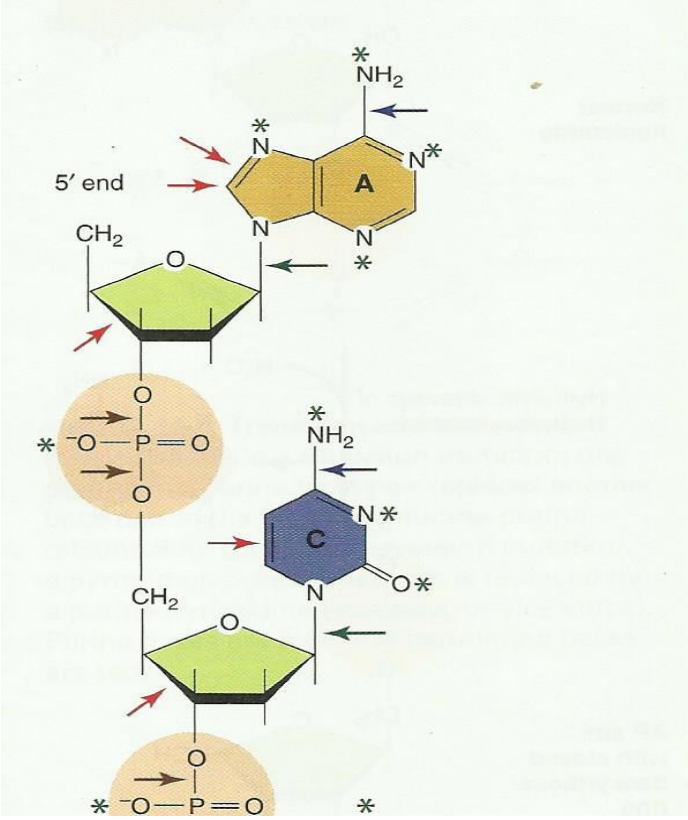
$H_2O^* \rightarrow {}^{\bullet}H + {}^{\bullet}OH$

So the three highly reactive chemical species i,e., H, OH and O_2^- are produced by the two primary pathways. Each of these attacks and damages whatever biomolecule they encounter. A wide variety of changes take place when that molecule happens to be DNA.

59. DNA INSTABILITY IN WATER

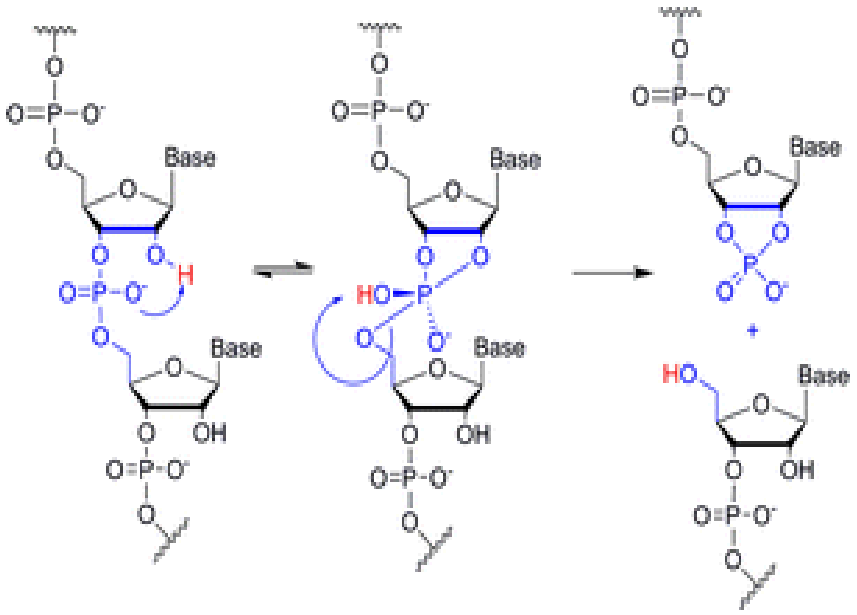
DNA has three kinds of bonds with the potential for hydrolytic cleavage, namely:

- Phosphodiester bond
- N-glycosyl bond
- Bonds linking exocyclic amine groups to bases



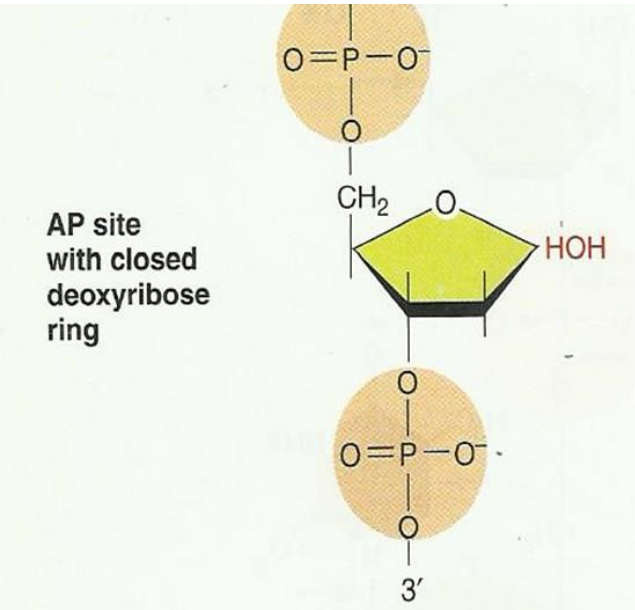
Hydrolytic bond Cleavage

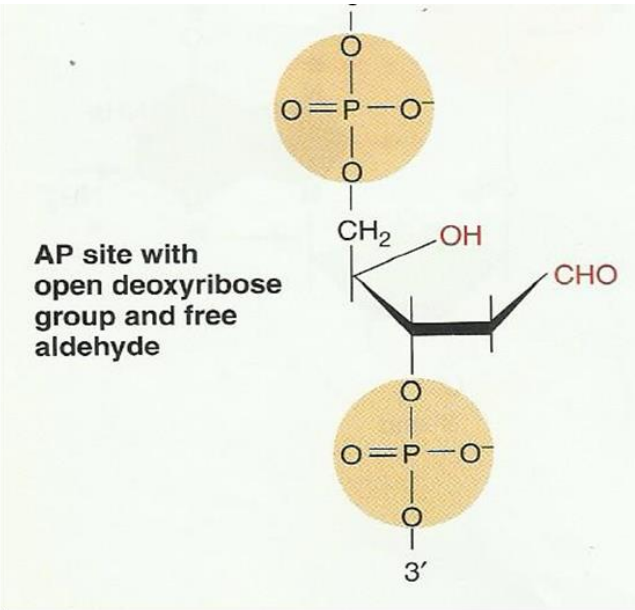
Spontaneous phosphodiester bond cleavage, which introduces a nick in the DNA strand, is a very rare occurrence and doesn't make a significant contribution to DNA damage.



N-glycosyl bond cleavage leads to the formation of an abasic site, which is also known as an AP site (AP for apurinic & apyrimidinic).

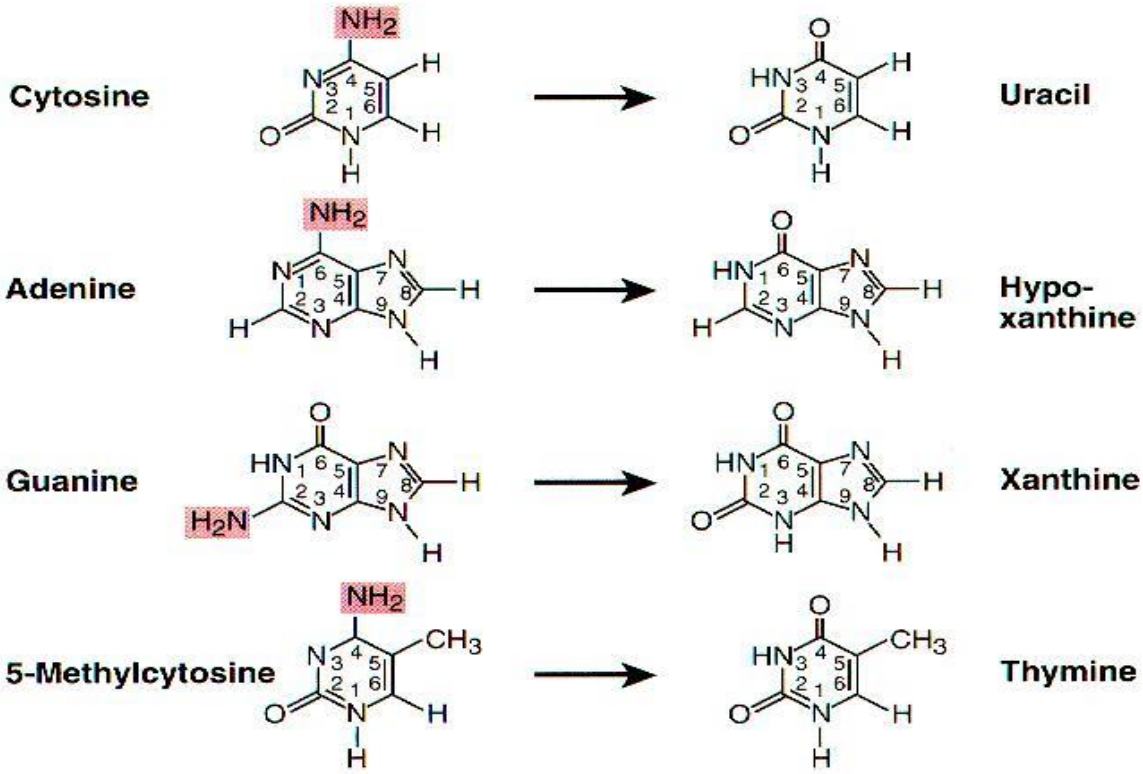
According to current etimates, about 10,000 purine and 500 pyrimidine bases are lost from DNA in a mammalian cell nucleus each day. Experiments are also showing that purine N-glycosyl bonds are more easily hydrolyzed than pyrimidine N-glycosyl bonds. AP site formation sensitizes the neighbouring 3'-phosphodiester bond to cleavage which can be attributed to the formation of a free aldehyde group. A DNA strand with one or more AP sites makes a poor template because it lacks the information required to direct accurate replication and transcription.



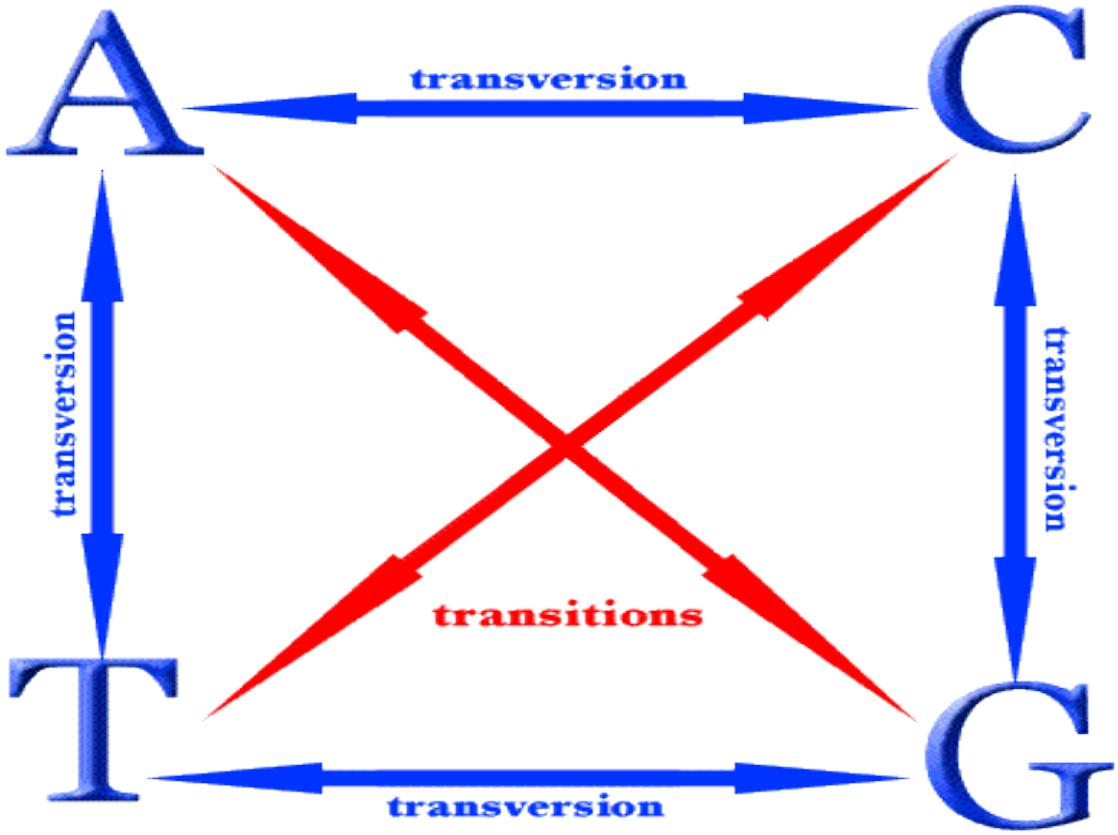


60 &61. Water-mediated Deamination

Water-mediated deamination converts cytosine, guanine and adenine to uracil, xanthine and hypoxanthine, respectively. Hydrolytic deamination of cytosine is estimated to take place about 100 to 500 times a day in a mammalian cell. Whereas, combined guanine and adenine deaminations are estimated to occur at about 1 or 2% of that of cytosine deamination.



The conversion of guanine to xanthine may result in mutations or arrested DNA synthesis because xanthine does not stable base pairs with either cytosine or thymine. While the conversion of adenine to hypoxanthine will cause a T – A base pair to be replaced by C – G base pair. Likewise, an uncorrected deamination that converts cytosine to uracil will cause a C – G base pair to be replaced with a T – A base pair. Mutations of this type in which a pyrimidine on one strand is replaced by a different pyrimidine and a purine on the other strand is replaced by a different purine are called transitions. Another type of replacement mutation which is termed as transversion mutation involves replacing a pyrimidine on one strand with purine and a purine on the other strand with a pyrimidine.



A few cytosine bases in eukaryotic DNA are converted into modified base 5-methylcytosine.
This modified base is concentrated in so called CpG islands. CpG islands are small segments of DNA often present in regulatory elements called promoters that are located just before the transcription unit they regulate. The frequency of spontaneous deamination of 5-methylcytosine bases in CpG islands is even greater than those of cytosine. The product of this deamination is thymine and not uracil.
So it results in the conversion of a C – G base pair to a T – A base pair. Nitrous acid (HNO2) is formed from nitrites used as preservatives in processed meat.
It reacts with amine groups attached to the ring structure in C, A and G and greatly increase their rate of deamination. Bisulfite (HSO3) used as an additive in fruit juices and dry fruits also greatly increases the rate of cytosine deamination but does'nt affect purine or 5-methylcytosine deamination.

62 & 63. Oxidative Damage to DNA

The reactive oxygen species damage DNA.
So the reactive oxygen species produced during cellular respiration (ETC) may have the potential to damage DNA. But it is unlikely that they do so because:-

1) the respiratory chain doesn't normally release these reactive oxygen species.

2) cells contain superoxide dismutase to convert superoxide radicals into molecular oxygen and hydrogen peroxide. Then catalase to convert hydrogen peroxide to oxygen and water.

3) superoxides and hydroxyl radicals are so reactive that, if released from respiratory chain, they would react with the nearby biomolecules before they had chance to reach nuclear DNA.

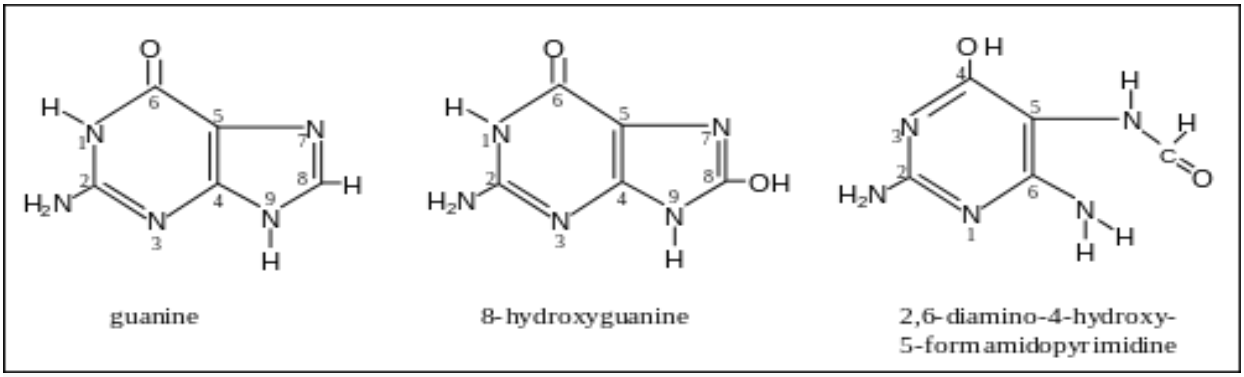
So the reactive oxygen species produced in this way do not damage the nuclear DNA under normal physiological conditions. The primary culprit appears to be the hydroxyl radical, which is produced by ionizing radiations. Hydroxyl radicals can also be produced chemically from hydrogen peroxide. Hydrogen peroxide is not as reactive as superoxide and hydroxyl radicals, so it has a much longer half-life in the cell, provided it escapes catalase and peroxidase.

If it does escape, hydrogen peroxide can be converted into hydroxide radical by the following reaction:-

$$Fe^{2+} + H2O2 \longrightarrow Fe^{3+} + \bullet OH + OH^{-}$$

This reaction is called the Fenton reaction and in this reaction copper, manganese and cobalt can replace iron. Hydroxyl radicals generated by any means are known to cause more than 80 different kinds of base damage.

Two of the oxidized base products are 8-oxoguanine (oxoG) and thymine glycol. 8-Oxoguanine can base pair with adenine or cytosine and if uncorrected, this 8-oxoG – A base pair will be replicated to form a T – A base pair, thus causing a transversion mutation. On the other hand, thymine glycol inhibits DNA replication and is therefore cytotoxic. Hydroxyl radicals produced by the Fenton reaction are tend to be more widely dispersed than those produced by ionizing radiations and therefore, much less likely to produce double-stranded breaks. Cells can repair single-stranded breaks much more easily than they can repair double-stranded breaks. Reactive oxygen species can also convert various biomolecules into reactive species that can then damage DNA. For example, polyunsaturated fatty acid oxidation produces two aldehyde products, malondialdehyde and 4- hydroxynonenal, which contribute to base damage.



oxidative damage of DNA, two of the most common changes guanine to 8-hydroxyguanine or to 2,6-diamino-4-hydroxy-5-formamidopyrimidine

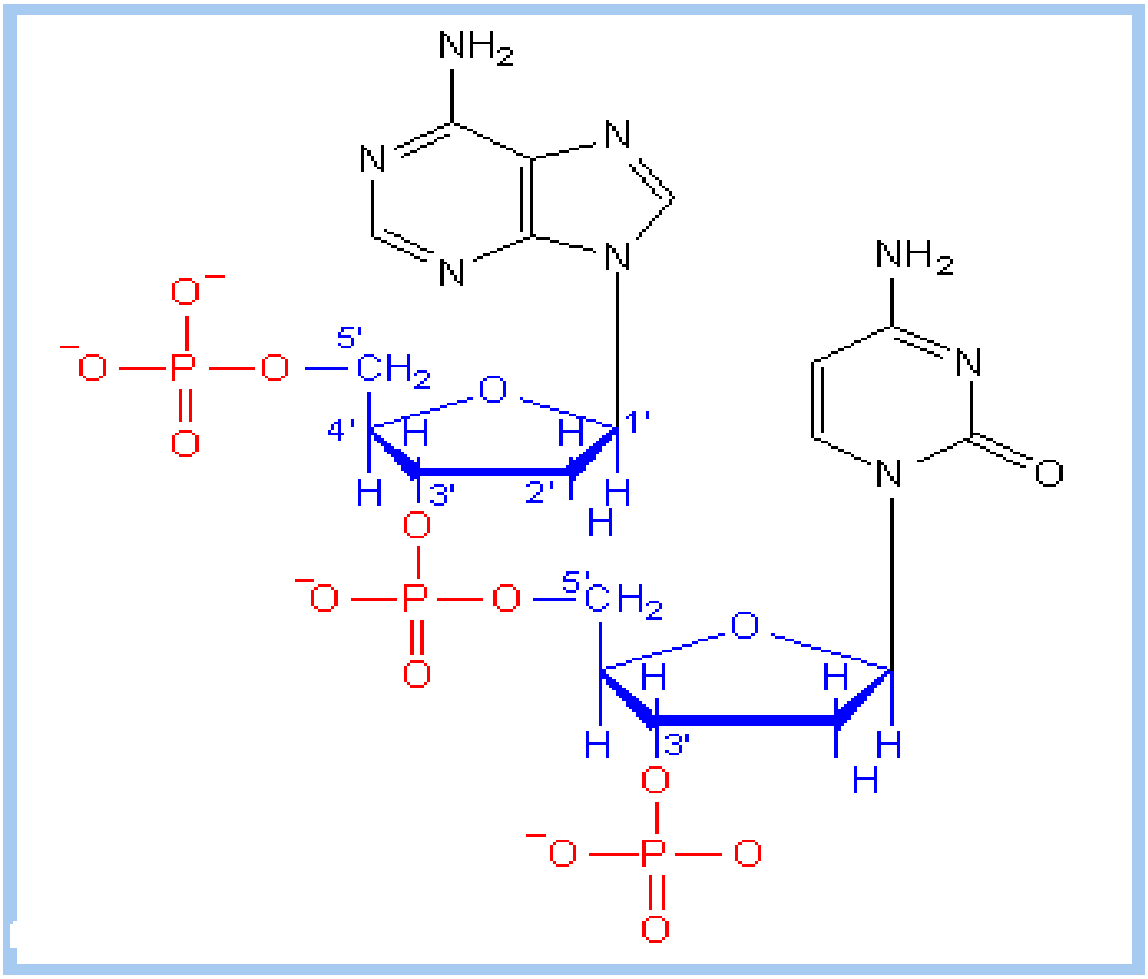
64 & 65. Alkylation Damage to DNA

DNA has electron-rich atoms that are readily attacked by electron-seeking chemicals called electrophiles or electrophilic agents. Alkylation is the transfer of an alkyl group from one molecule to another. The alkyl group may be transferred as an alkylcarbocation, a free radical, a carbanion or a carbene (or their equivalents).[1] An alkyl group is a piece of a molecule with the general formula CnH2n+1, where n is the integer depicting the number of carbons linked together. For example, a methyl group (CH3) is a fragment of a methane molecule (CH4); n = 1 in this case. Alkylating agents utilize selective alkylation by adding the desired aliphatic carbon chain to the previously chosen starting molecule. This is one of many known chemical syntheses. Alkylating agents are a highly reactive group of electrophiles which transfer methyl, ethyl, or alkyl groups to the electron-rich atoms in the DNA and damage it. Alkylation in DNA takes place at:-

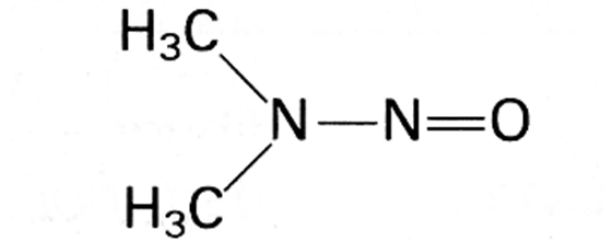
A) nitrogen and oxygen atoms external to the base ring systems;

B) non-bridging oxygen atoms in phosphate groups. C) nitrogen atoms in the base ring systems except those linked to deoxyribose.

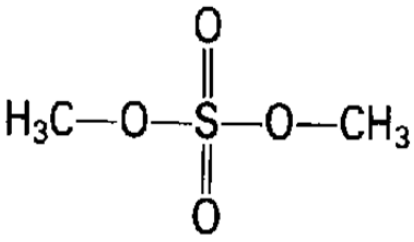
Many different kinds of naturally occurring and synthetic chemical agents are known to transfer alkylating agents to DNA.



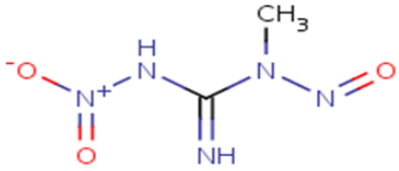
The product formed by attaching a chemical group to DNA is called an adduct. If the chemical group attaches to a single site on the DNA then the product is termed as monoadduct. The exposure of DNA to dimethylnitrosamine leads to the production of a monoadduct in which a single methyl group attaches to DNA.



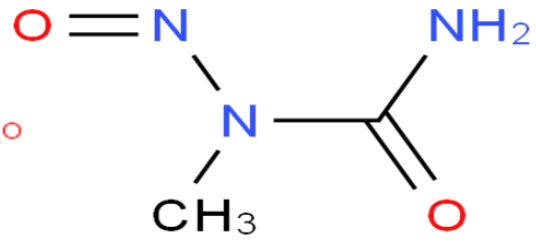
Dimethylnitrosamine



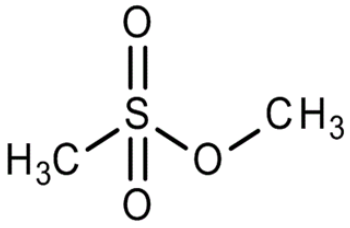
Dimethyl sulfate (DMS)



NTG



MNU

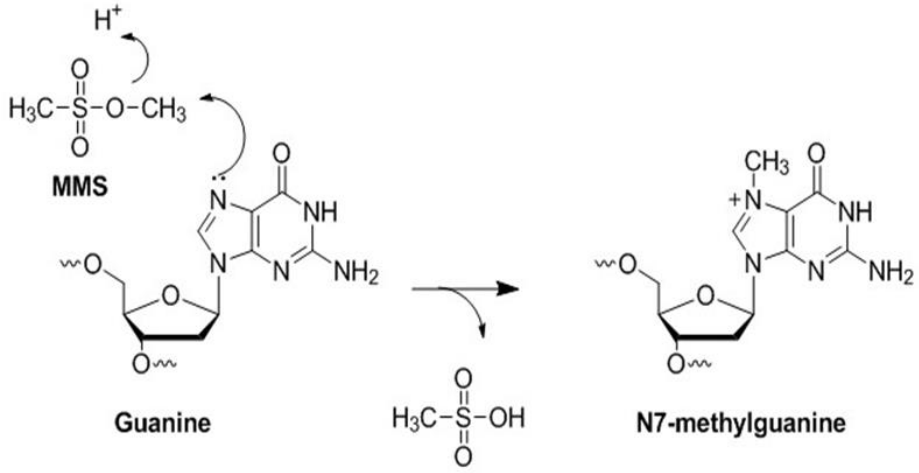


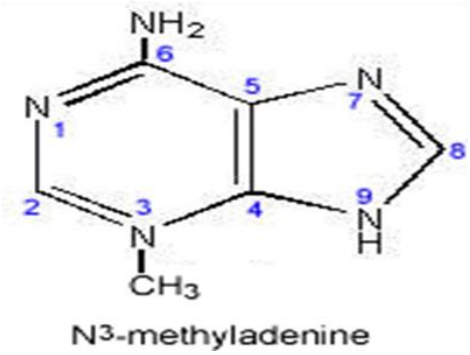
MMS

When DNA is exposed to methyl methane sulfonate (MMS) or N-methyl-N-nitrosourea (MNU), methylation takes place most frequently at:-

i) N-7 position in guanine and

ii) next most frequently at N-3 in adenine.



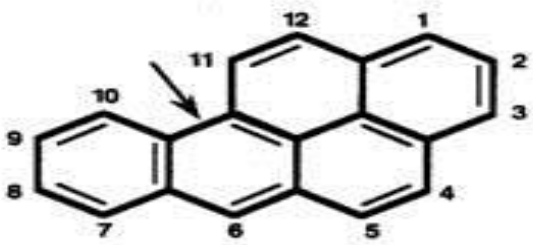


N-methylguanine forms a base pair with cytosine, but it is readily removed from the DNA with the resultant formation of an abasic site. Methylation at N-3 in adenine is of great practical significance because N-3 methyl adenine formation blocks DNA replication but does not appear to lead to mutations. Therefore, a methylating agent that can transfer methyl group exclusively to N-3 position in adenine would have the potential to kill cancer cells without causing mutations. Methylation at O-6 in guanine and O-4 in thymine are much less frequent events than either of the above described methylations. O^6-methylguanine and O^4-methylthymine formation are quite important because the methylated bases mispair during DNA replication, resulting in transition mutations. The phosphate groups in DNA backbone can also be methylated.
The resulting neutral phosphodiester is easily cleaved by water to produce single strand breaks.

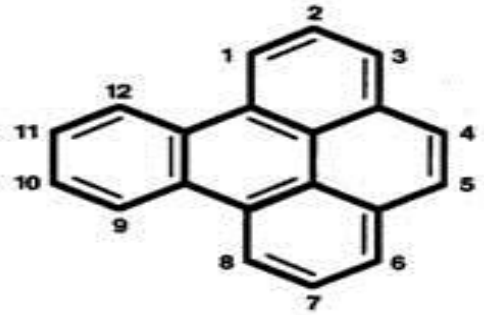
66. DNA Damage by PAHs

Many environmental agents become active alkylating agents only after they are metabolized in the cells. One such agent is a mixture of polycyclic aromatic hydrocarbons (PAHs) formed by the

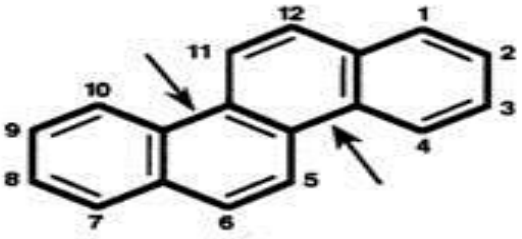
incomplete combustion of the burning wood or coal used as fuel. Similar type of polycyclic aromatic hydrocarbons are also present in tobacco smoke and charbroiled meats.
There are more than 100 different types of PAH compounds.



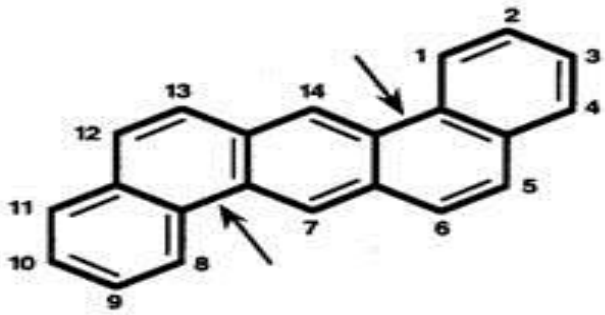
Benzo[*a*]pyrene



Benzo[*e*]pyrene

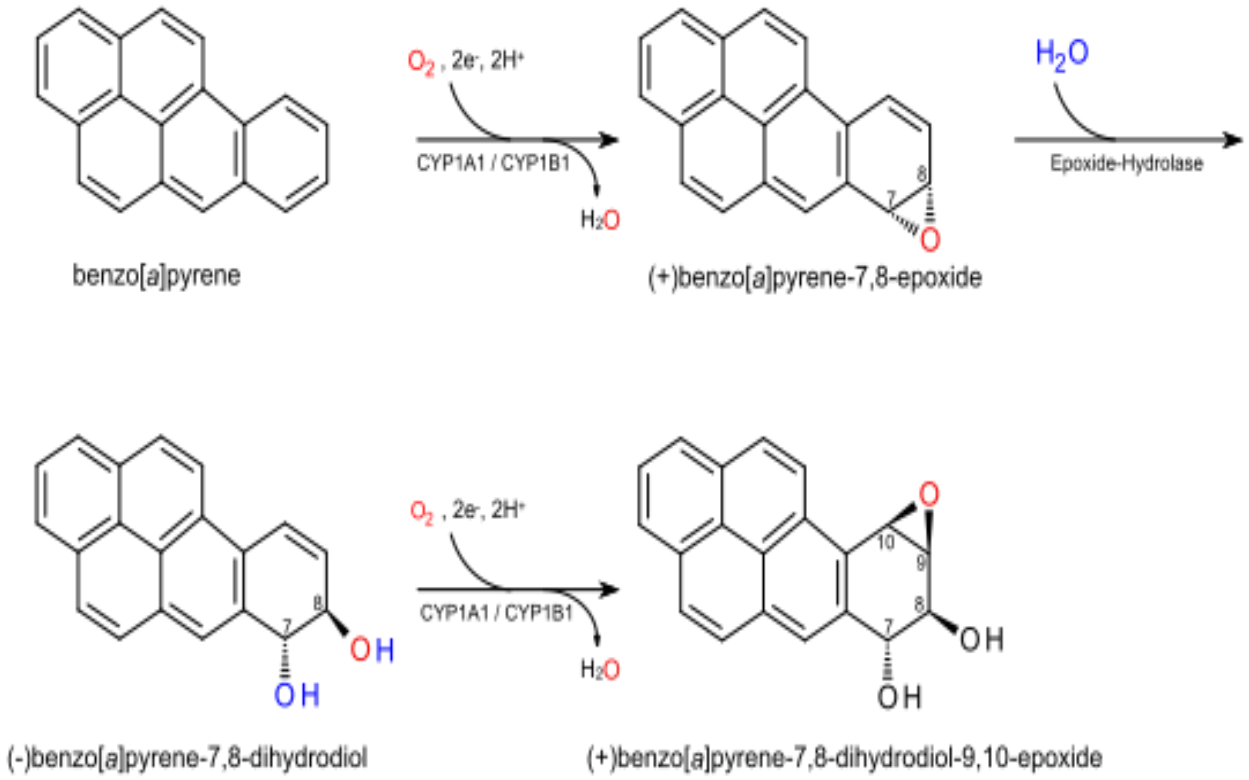


Chrysene

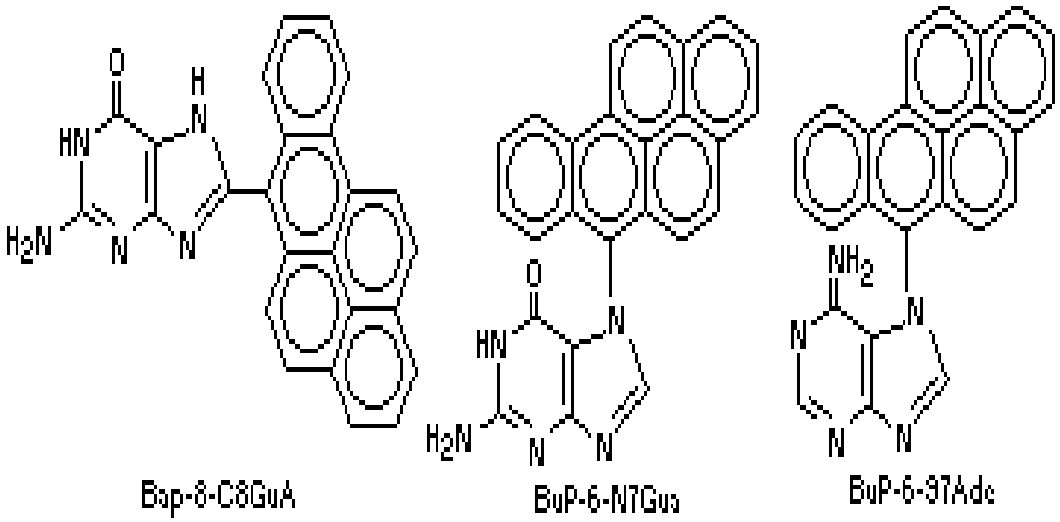


Dibenz[*a,h*]anthracene

The common structural feature in all PAHs is two or more fused aromatic rings. These are not able to damage DNA unless they are metabolically activated in the cell. One of the polycyclic aromatic hydrocarbon Benzo[a]pyrene is converted into an active epoxide alkylating agent through a metabolic pathway.

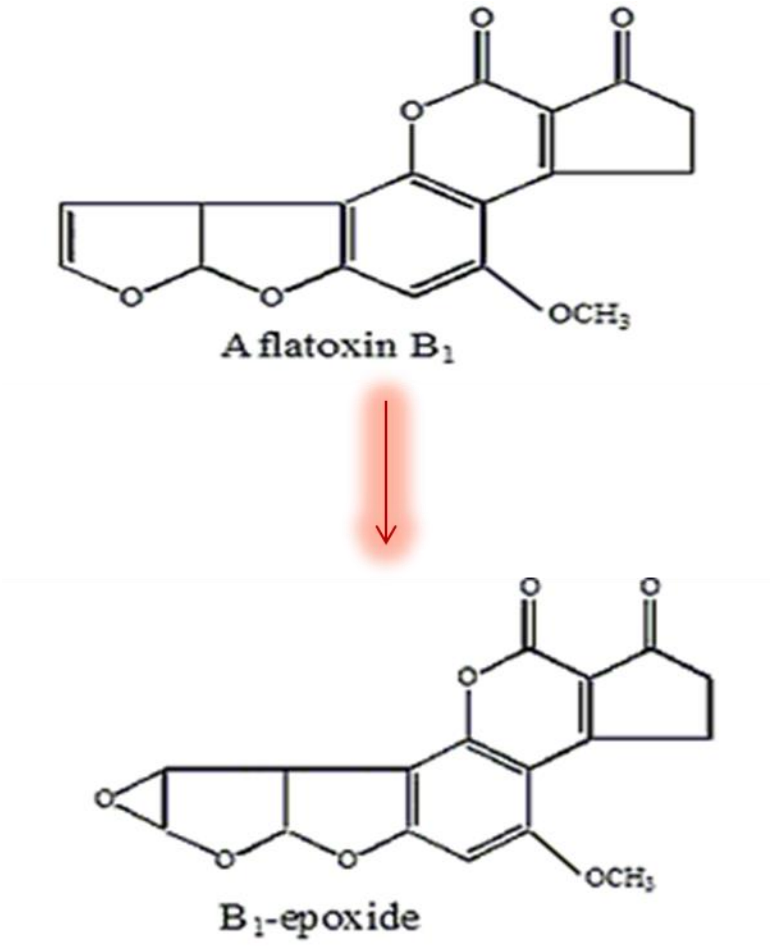


Cytochrome P450 enzymes being not highly specific can act on PAHs such as benzo[a]pyrene and add oxygen atoms to form reactive three- membered epoxide rings. These epoxides then alkylate DNA, causing replication errors that result in mutations, which ultimately convert a normal cell into a cancer cell.

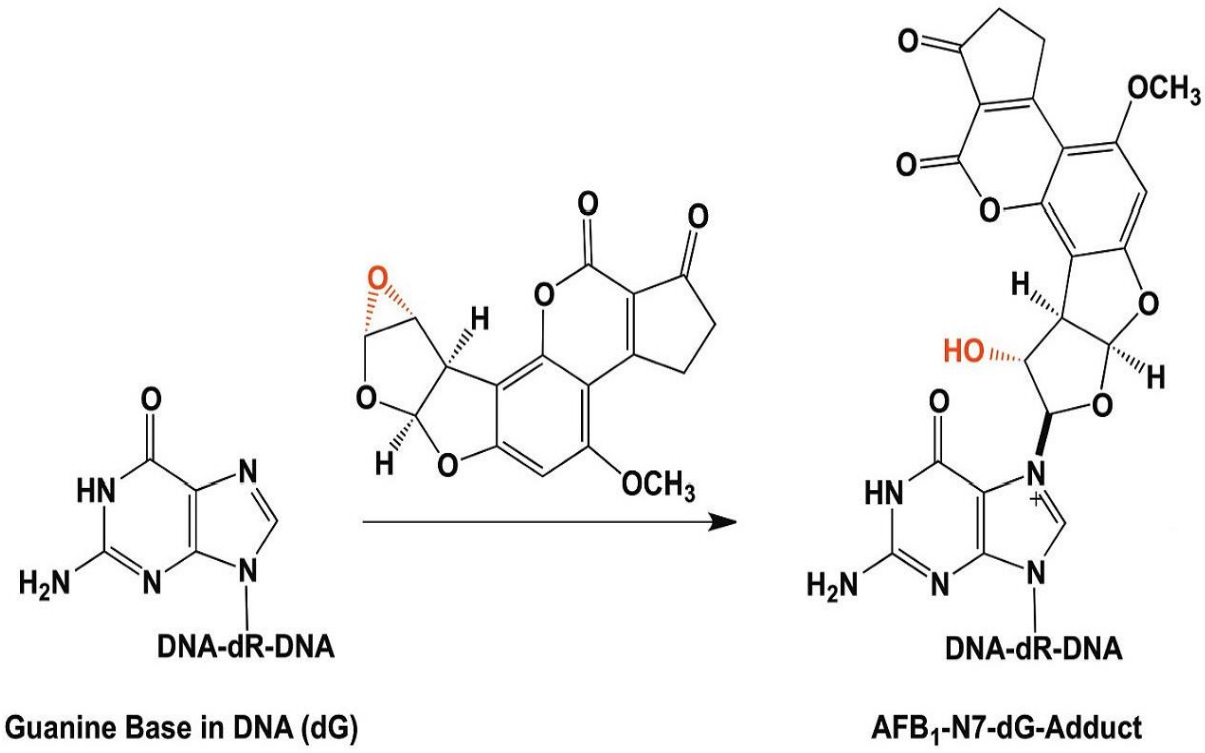


67. DNA Damage by Aflatoxins

Another class of chemical carcinogens that must be activated before damaging DNA are called aflatoxins. They are produced from *Aspergillus flavus* and *A. parasiticus*, fungi that grow on peanut and other grains such as rice and corn. Animals feeding on contaminated peanuts or grains containing aflatoxins exhibit markedly increased rates of liver diseases including liver cancer. Aflatoxin B1, the most potent toxin produced by *A. flavus* presents a particularly serious health threat in the United States.
Cytochrome P450 converts it into an epoxide derivative that damages DNA.



Under ideal conditions, a small tripeptide glutathione will attack the epoxide ring, making the aflatoxin derivative soluble so that it can be excreted in the urine. If the reactive epoxide derivatives escape the attack of glutathione, they are free to attack guanine rings in DNA.



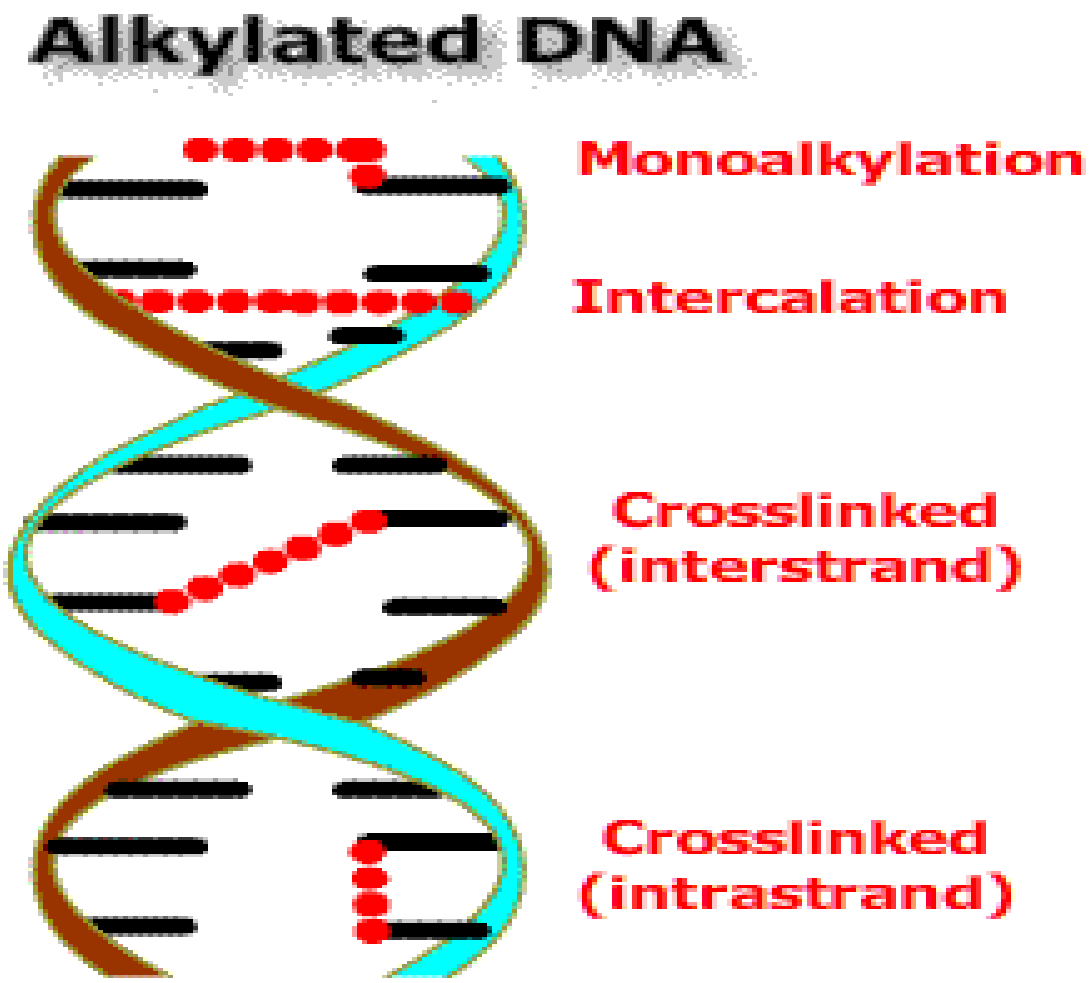
Guanine Base in DNA (dG)

AFB_1-N7-dG-Adduct

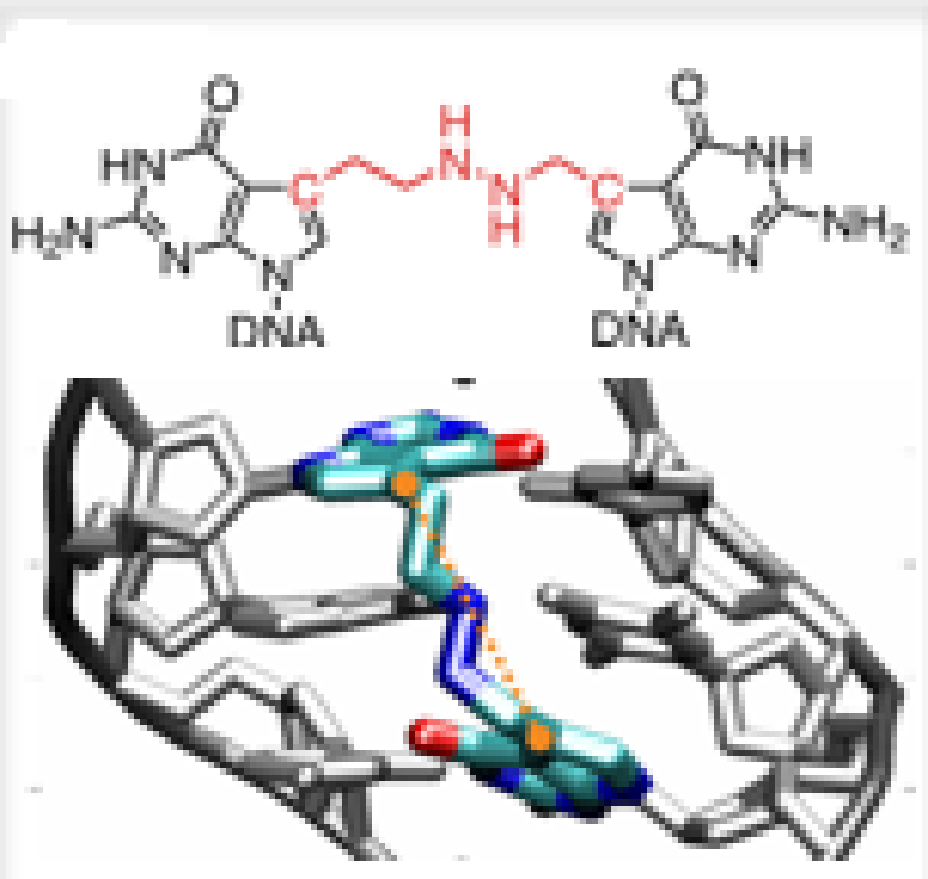
The flat aflatoxin ring system inserts between DNA bases causing helical distortion that in turn leads to replication errors. The flat aflatoxin ring system inserts between DNA bases causing helical distortion that in turn leads to replication errors.

68. Chemical Cross-Linking Agents

Many alkylating agents, having two reactive sites can form intra-strand or inter-strand cross-links in addition to forming monoadducts. Inter-strand cross-links are of special interest because they prevent strand separation an, if not corrected, are lethal.



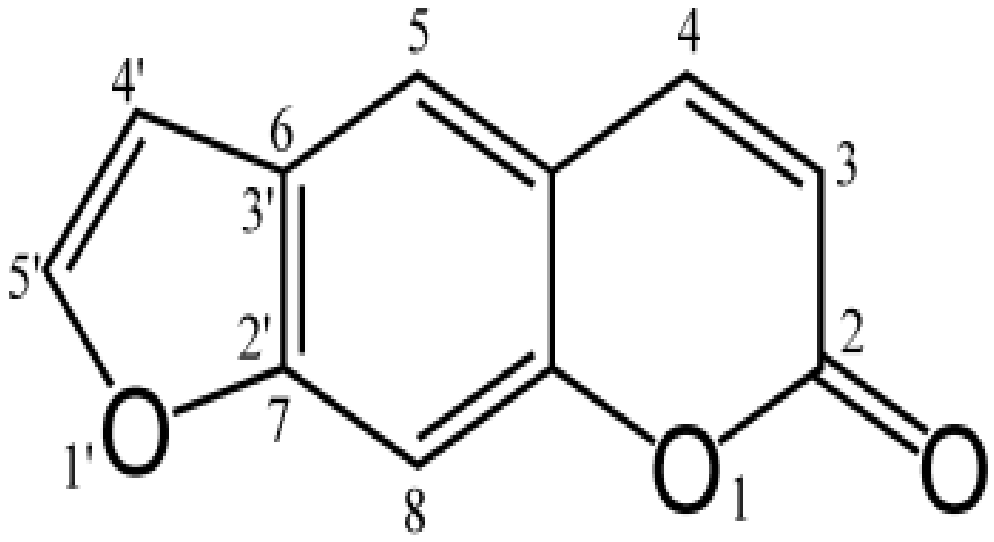
One of the simplest cross-linking agents, nitrogen mustard gas (bis[2-chloroethyl] methylamine damages DNA by forming inter-strand cross-links. It does so by attacking N-7 on two guanines, which are on opposite strands of DNA double helix.



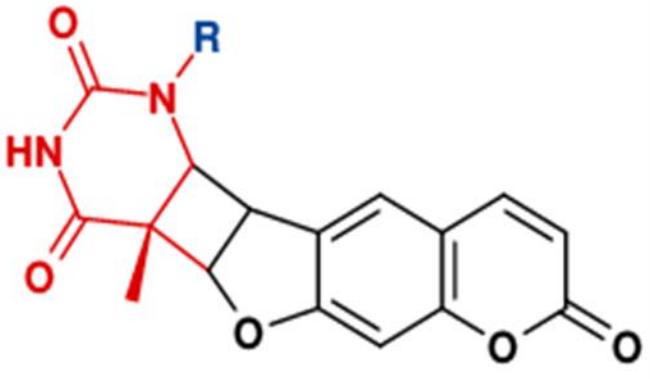
Although a very toxic substance, nitrogen mustard gas has found clinical application as a chemotherapeutic agent for treating certain forms of cancers.

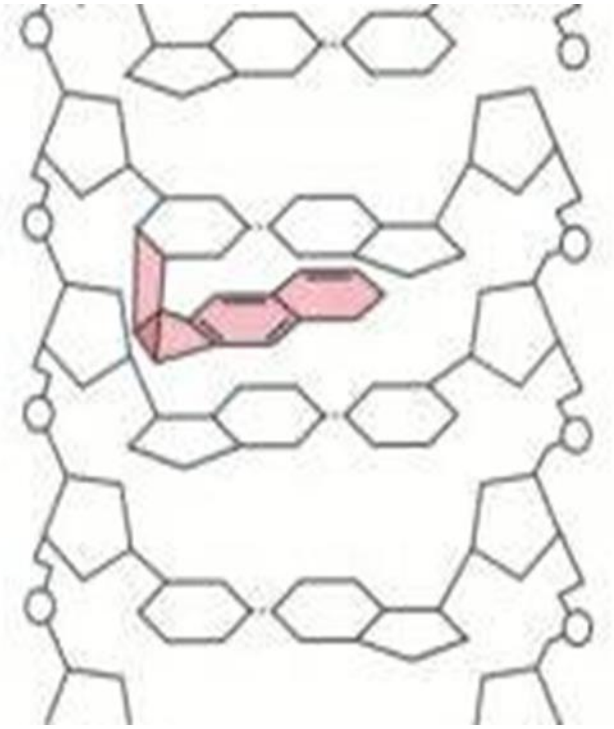
69. DNA damage by Psoralen

Psoralen can form mono-adducts or cross-links. It is a naturally occurring substance that can alkylate DNA if photoactivated. The planar psoralen molecule, which consist of a furan ring fused to a heterobicyclic ring system called coumarin, intercalates into the DNA molecule.

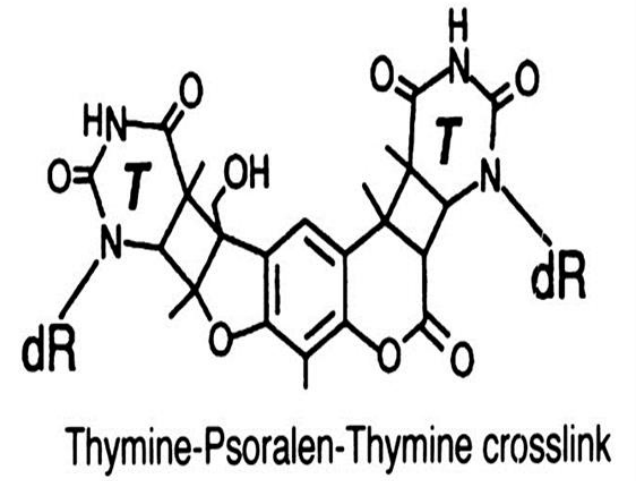


Upon exposure to light with a wavelength of 400 to 450nm, the furan ring in psoralen becomes activated and adds across the 5,6 double bond in a pyrimidine base to form the 4',5', monoadduct.

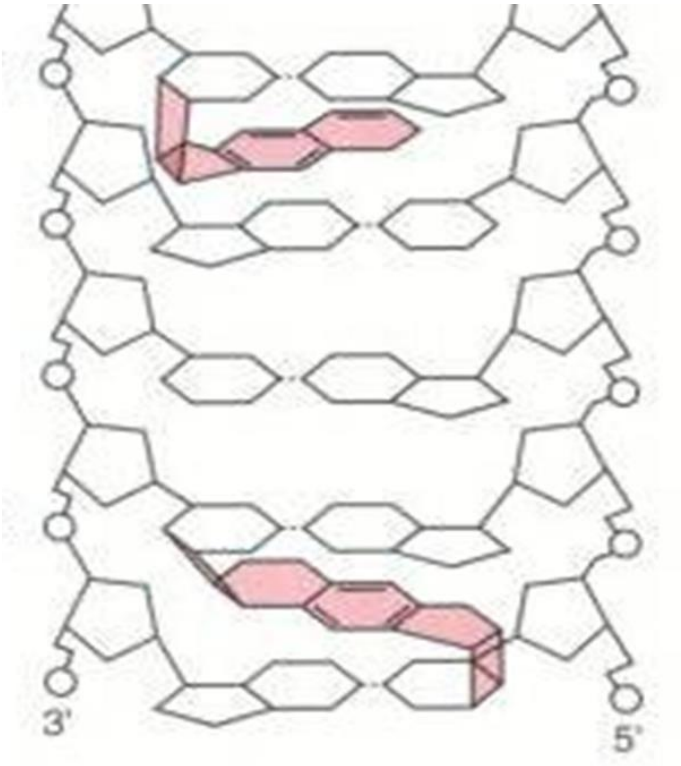




The planar tricyclic psoralen derivative in the monoadduct is in position to combine with a second pyrimidine base on the opposite DNA strand. But to do so, it must be activated by light with the wavelength of 320 to 400nm. The resulting photoproduct contain a cross-link between pyrimidine bases on opposite strands of the DNA duplex.



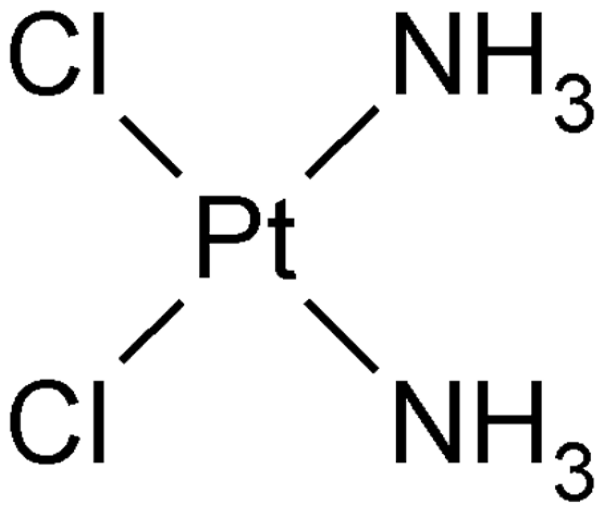
Thymine-Psoralen-Thymine crosslink



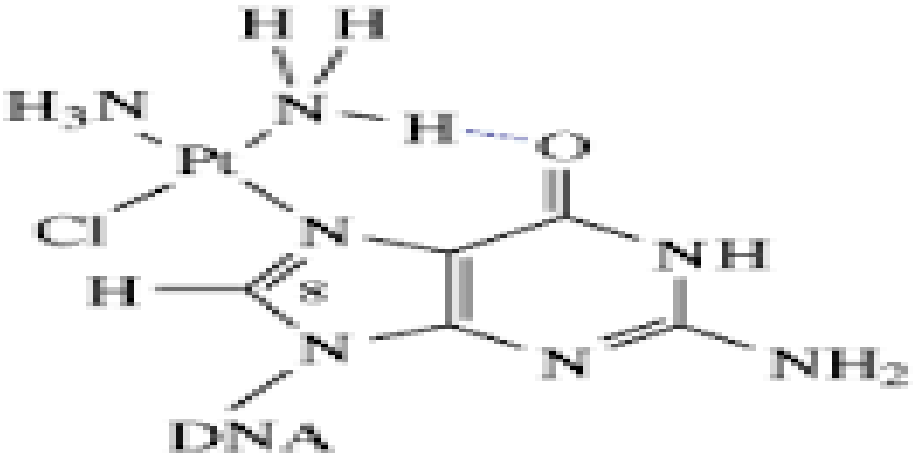
If not properly repaired, psoralen damage causes mutations and is lethal to cells.

70. DNA damage by Cisplatin

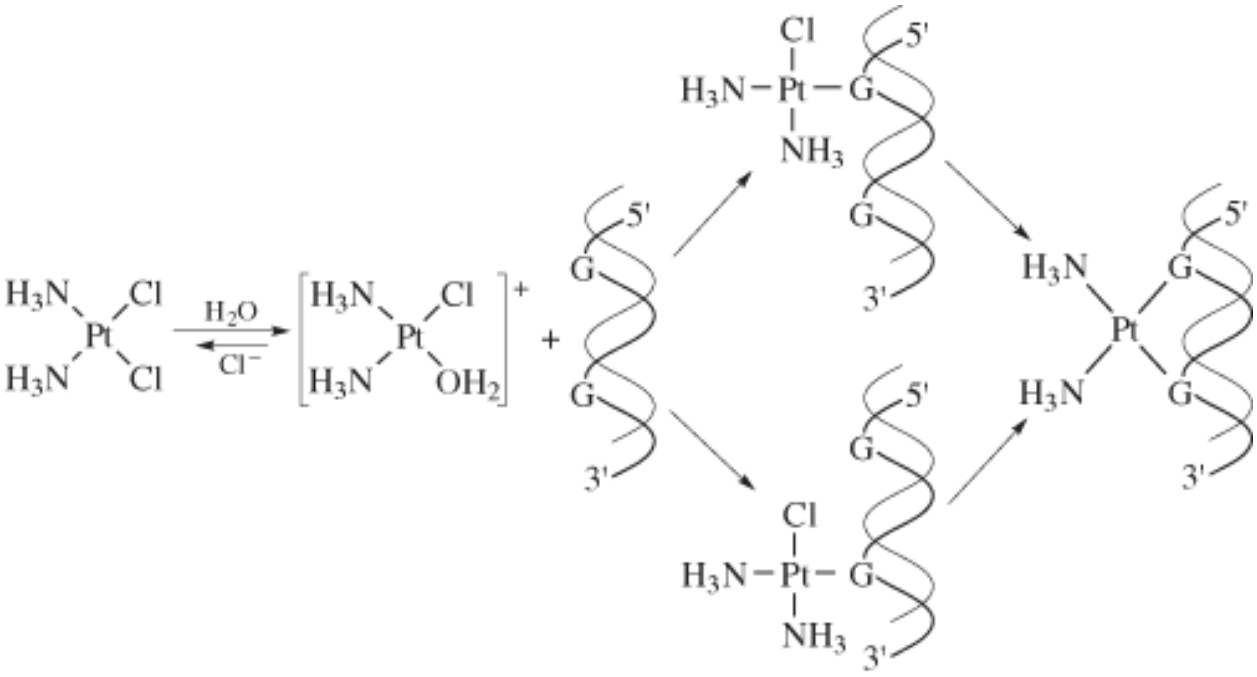
While examining the effects of electric currents on *E. coli*, a new compound viz; *cis*-diamminedichloroplatinum, better known as Cisplatin was found to block cell division in *E. coli.* Efforts were made to see if cisplatin would also inhibit cell division in other kinds of cells. It was revealed that cisplatin blocks the division of tumor cells.



After cisplatin enters the cell by passive diffusion or active transport, it undergoes hydrolysis to produce a highly reactive and charged complex ($Pt(NH_3)_2ClH_2O^+$). This complex coordinates to the N-7 atom of either a guanine or adenine base in DNA.



Then the remaining chloride ligand is displaced by hydrolysis , allowing the platinum to coordinate to a second purine base on the same or opposite strand of the double stranded DNA.



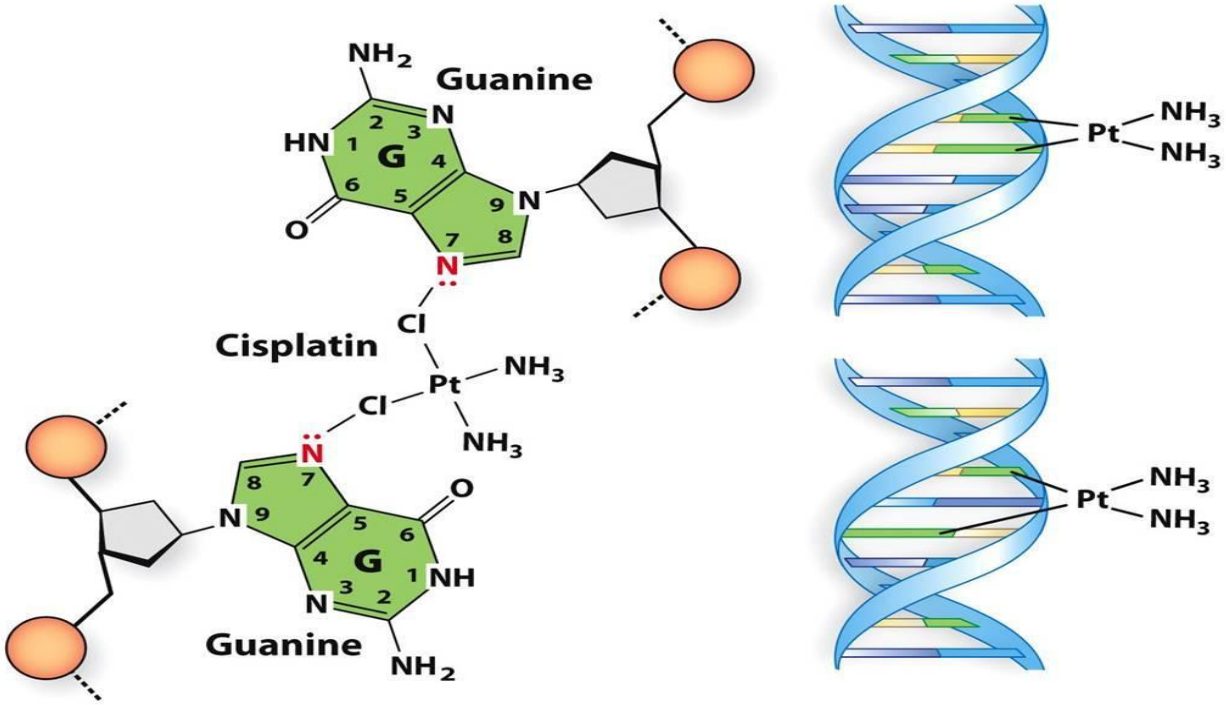
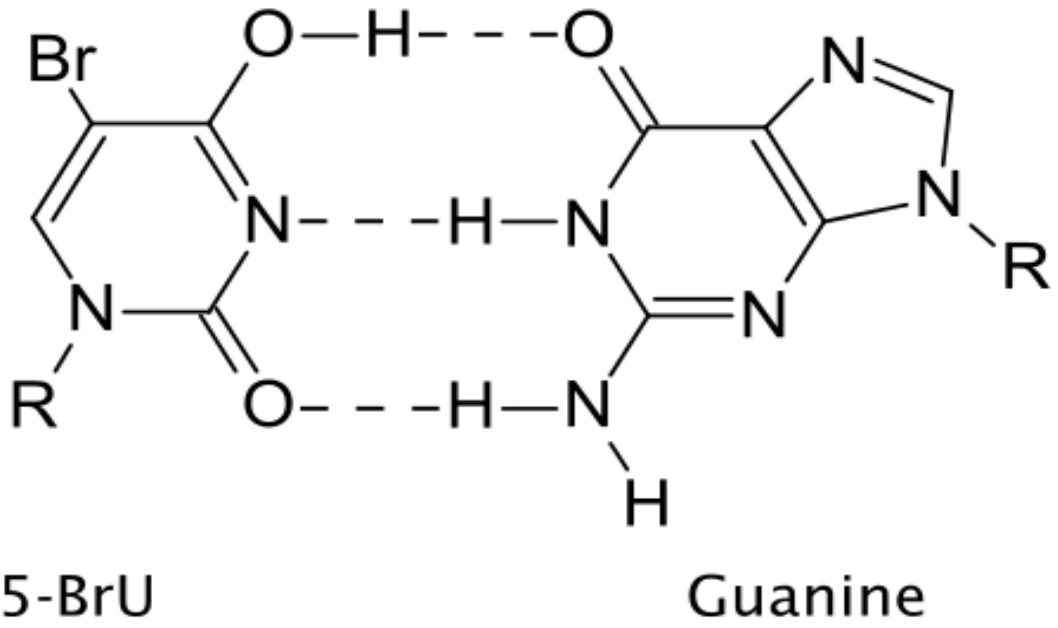


Figure 12-14a
Molecular Biology: Principles and Practice

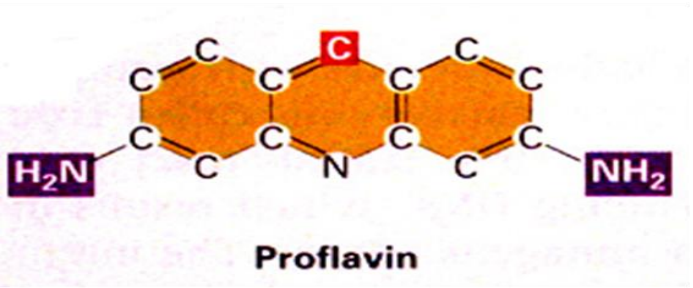
Cisplatin's cytotoxic effects appear to be due to inter-strand cross-linking which blocks the replication and transcription machinery. Cisplatin is a very effective chemotherapeutic agent for treating cancers of the bladder, ovaries and testicles, but has a number of side effects.

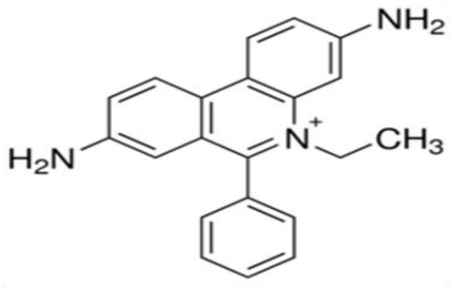
71. Base Analogs and Intercalating Agents

Mutations are also caused by compounds that substitute for normal bases called Base analogs or or the compounds that slip between the bases called Intercalating agents and cause errors in replication. Base analogs are structurally similar to proper bases but differ in ways that make them treacherous to the cell. Thus, base analogs are similar enough to the proper bases to get taken up by cells, converted into nucleoside triphosphates, and incorporated into DNA during replication. But, because of the structural differences from the proper bases, the analogs base-pair inaccurately, leading to frequent mistakes during the replication process. One of the most mutagenic base analogs is 5-bromouracil, an analog of thymine. The presence of the bromo substituent allows the base to mispair with guanine.

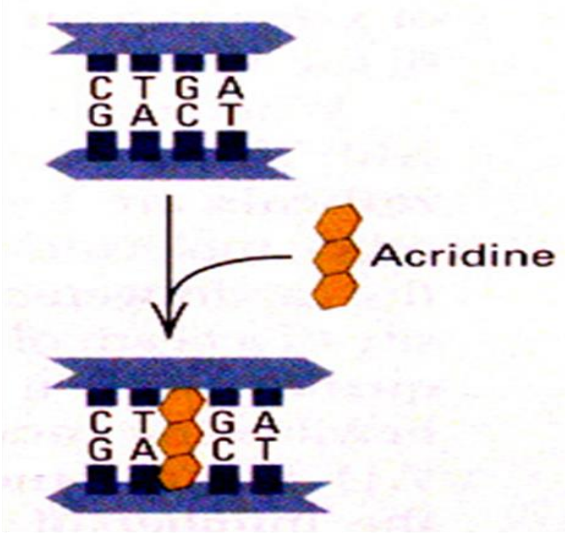


Intercalating agents are flat molecules containing several polycyclic rings that bind to the purine or pyrimidine bases of DNA, just as the bases bind or stack with each other in the double helix.





Ethidium



Intercalating agents, such as proflavin, acridine, and ethidium, cause the deletion or addition of a base pair or even a few base pairs. One possibility in the case of insertions is that, by slipping between the bases in the template strand, these mutagens cause the DNA polymerase to insert an extra nucleotide opposite the intercalated molecule. Conversely, in the case of deletions, the distortion to the template caused by the presence of an intercalated molecule might cause the polymerase to skip a nucleotide.

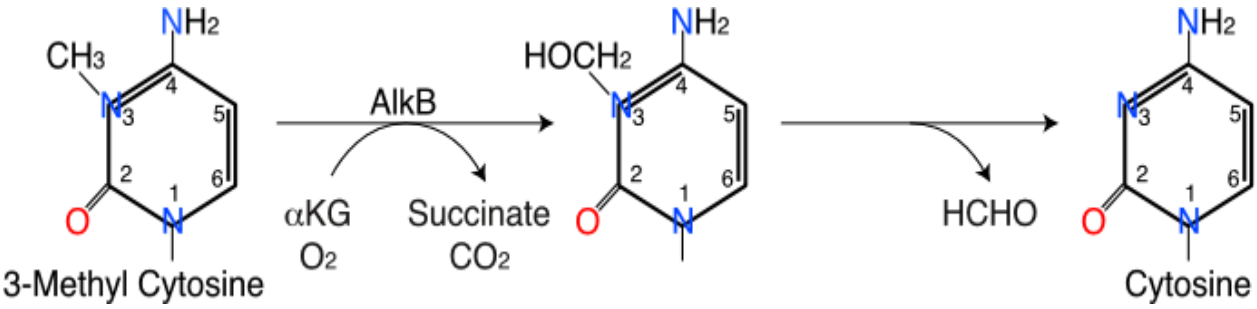
Chapter 9. DNA Repair

72. Direct Reversal of DNA Damage

The first clue to the existence of an enzyme that catalyzes the direct reversal of DNA damage was reported by Albert Kelner in 1949. In his experiments, Kelner first irradiated bacteria with UV light at doses that killed most of the bacteria. The he tested the survivors to isolate the desired mutants. Even though Kelner was very careful in his experimentation, he noticed a great deal of variation in the number of survivors from one experiment to the other. Finally, he found that cells placed in dark after UV treatment had a very low survival rate whereas those placed in light had a high survival rate. Exposure to light thus reversed the UV light's bactericidal effects. Similar phenomenon was also observed by Renato Dulbecco while studying UV-irradiated phage T2. Dulbecco prepared multiple plates each containing the same number of UV-irradiated phages and sensitive bacteria, and placed the plates in a stack under a florescent bulb in the lab. Each of the stacked plates should have about the same number of plaques.

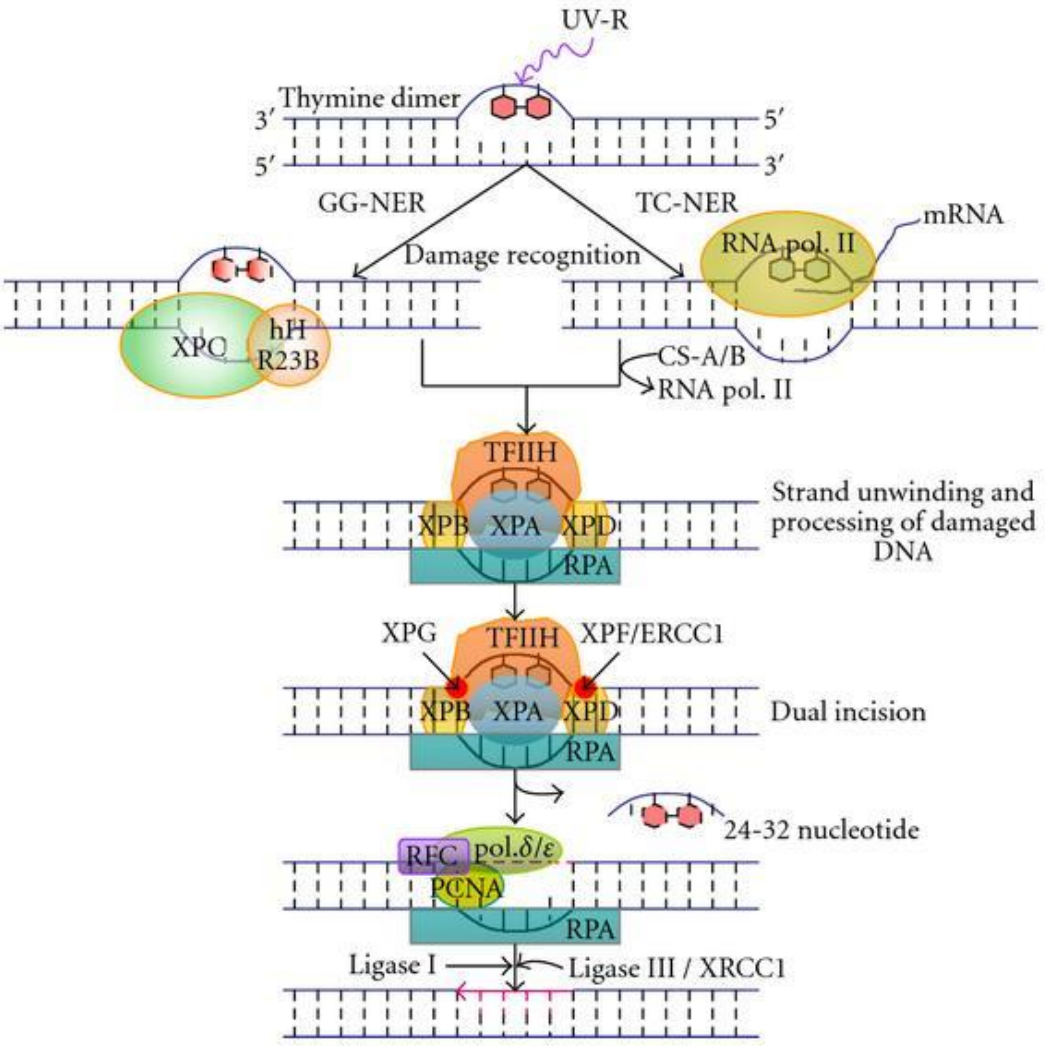
However, the plaque number decreased dramatically, going from top to bottom of the stack. Dulbecco explained this by proposing that the plates on the top of the stack were exposed to more light from the bulb as compared to the plates on the bottom of the stack. He tested this hypothesis by exposing some of the plates to more to fluorescent light while keeping others in dark. As expected, the number of plaques on plates exposed to light was much higher than those left in the dark. The bacteria were somehow using the visible light to repair the UV damaged DNA in the phage. The chemical basis for this light-dependant phenomenon, which Dulbecco called photoreactivation, remained to be elucidated.

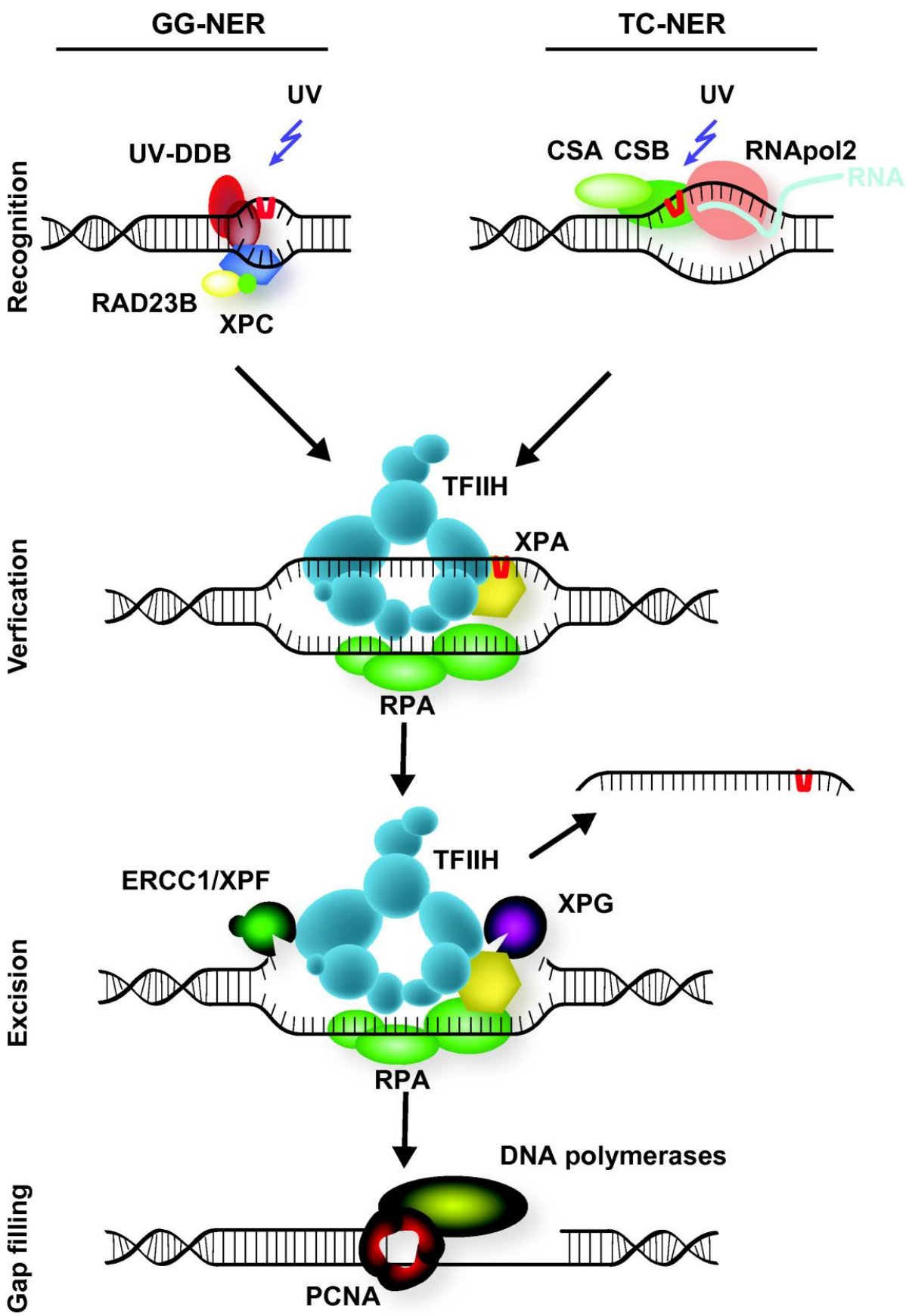
Direct DNA damage reversal is also provided by AlkB, a protein that is found in most, perhaps all, living organisms (including humans). This protein is a member of the class of enzymes called alpha-keto-glutarate-dependent and iron-dependent oxygenases (aKG-Fe(II)-oxygenases), which use iron-oxo intermediates to oxidize chemically inert compounds. In the process, alpha-keto-glutarate is converted to succinate and CO_2, which is similar to one of the steps in the tricarboxylic acid cycle. AlkB is capable of reversing methylation at the 1 position of adenine and the structurally similar 3 position of cytosine, as shown in this diagram:



73. Photoreactivation

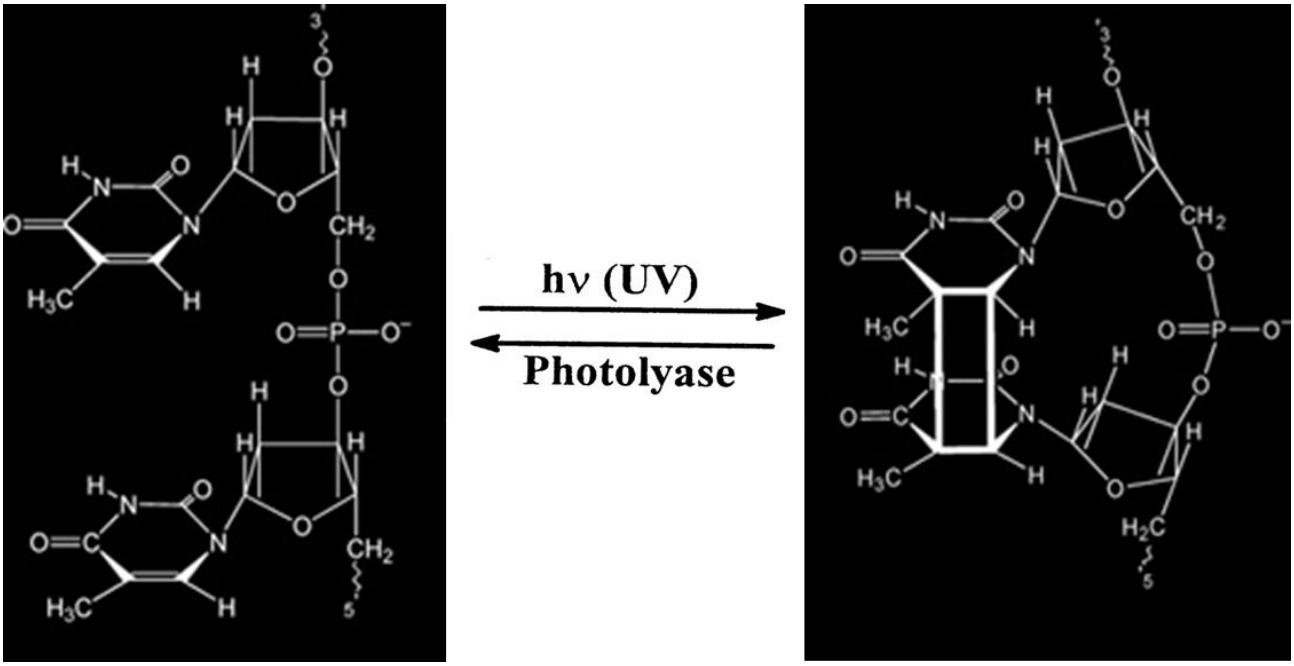
Claud S. Rupert and co-workers devised an in vitro photoreactivation system in 1957, taking a major step toward determining the chemical mechanism of photoreactivation. They used a straightforward approach in which they isolated DNA from the gram-negative bacteria *Haemophilus influenzae*. They irradiated this DNA with UV light to inactivate its transforming ability. Then they demonstrated that a cell-free *E. coli* extract restored the transforming activity in the presence of light. Although, there study had the potential to open the way for the purification and characterization of photoreactivation enzyme, investigators still needed to establish the chemical nature of this repair.


UV-R
Thymine dimer
3′
5′
5′
3′
GG-NER
TC-NER
Damage recognition
RNA pol. II
mRNA
XPC
hH
R23B
CS-A/B
RNA pol. II
TFIIH
XPB
XPA
XPD
RPA
Strand unwinding and processing of damaged DNA
XPG
TFIIH
XPF/ERCC1
XPB
XPA
XPD
RPA
Dual incision
24-32 nucleotide
RFC
pol.δ/ε
PCNA
RPA
Ligase I
Ligase III / XRCC1


GG-NER
TC-NER
UV
UV
UV-DDB
CSA CSB
RNApol2
RNA
Recognition
RAD23B
XPC
TFIIH
XPA
Verfication
RPA
TFIIH
ERCC1/XPF
XPG
Excision
RPA
DNA polymerases
Gap filling
PCNA

74. CPD Photolyase

The problem of the chemical nature of photoreactivation was resolved over next few years when investigators came to know that UV irradiation induces the formation of cyclobutan pyrimidine dimer in DNA. Further studies showed that the photoreactivation enzyme reverses the UV induced damage by using the energy provided by blue light (350-450nm) to drive cyclobutane ring disruption. When it was recognized that the photoreactivation enzyme catalyzes the disruption of carbon – carbon bonds, it was given a more descriptive name of CPD photolyase.



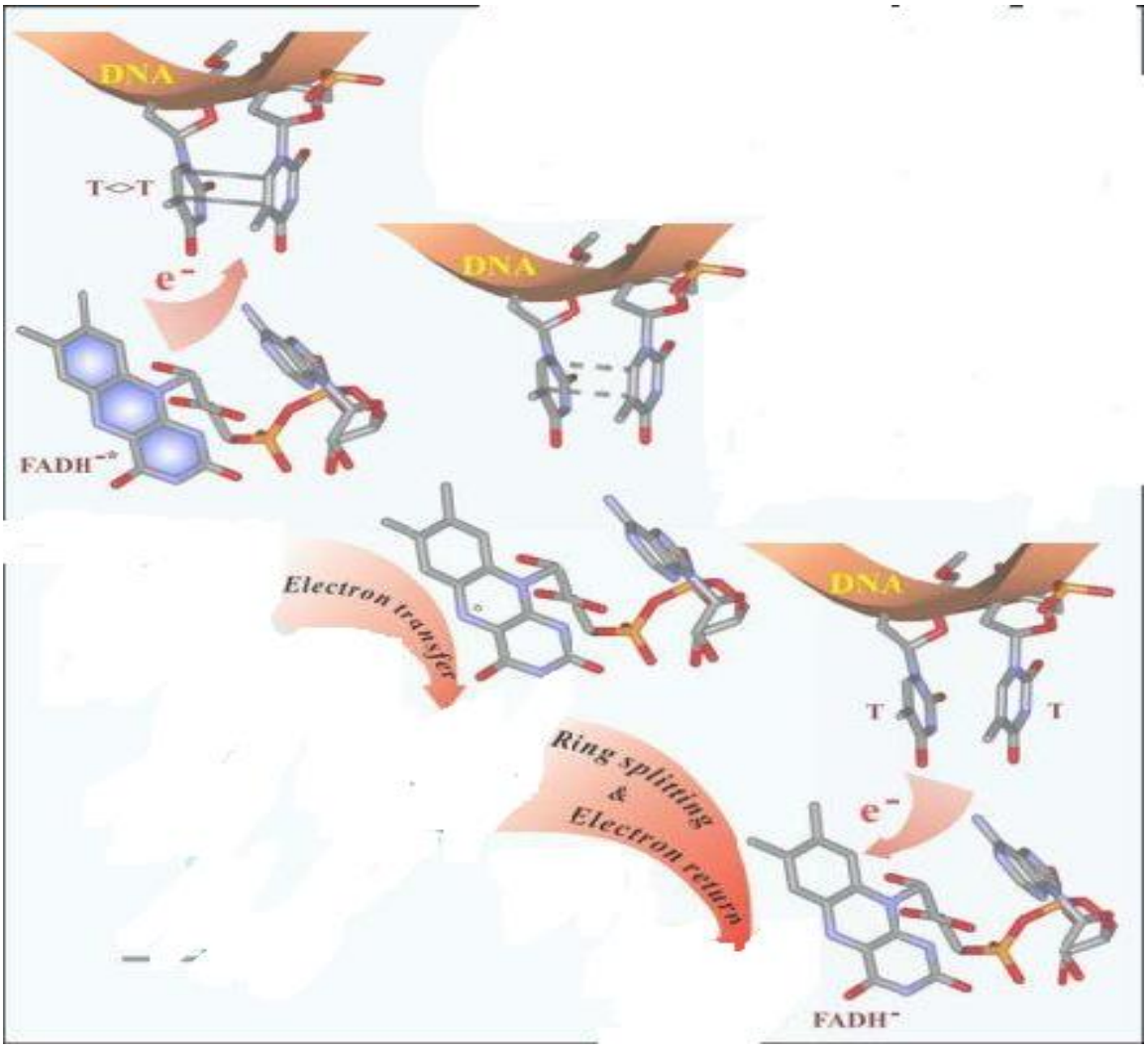
The bacteria that lack CPD photolyase can't repaire cyclobutane pyrimidine lesions through photoreaction thus can't reverse the UV induced damage in DNA. CPD photolyases are present in a wide variety of organisms including bacteria, archaea, plants, and animals but not in humans and other placental mammals. These are monomeric proteins ranging in size from about 450-550 amino acid residues.

All CPD photolyases have two domains, designated as N – and C – terminal domains. A light absorbing pigment, or chromophore pigment binds to each domain through non-covalent bonding. This chromophore factor acts as a photoantenna to capture light with wavelengths that would not otherwise be available.

75. Mechanism of CPD Photolyase

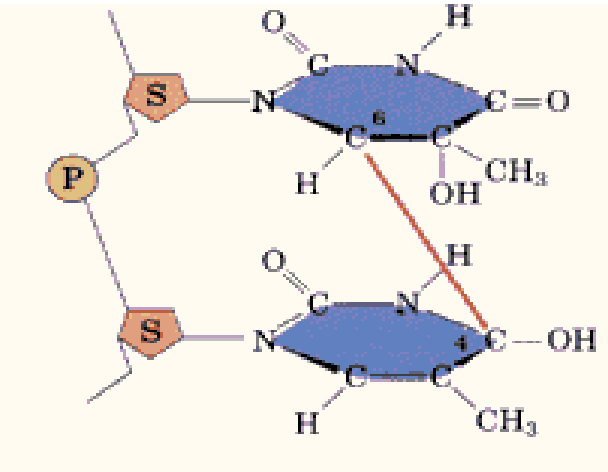
CPD photolyase can bind to DNA in the dark by recognizing the altered DNA structure caused by a CPD formation rather than a specific nucleotide sequence. This binding is about 10^5 tighter

when a DNA segment contains a CPD than when it does not. Half of the binding energy appears to come from interactions between enzyme and the DNA back bone. The other half of energy comes from interactions between the FADH at the active site and the CPD. However, the light harvesting antenna pigment does not influence binding. Once the enzyme – DNA complex is formed, the CPD is flipped out of the DNA double helix and into the enzyme's active site. After the CPD flips into the enzyme's active site, the energy of an absorbed photon is transferred from the light harvesting antenna pigment to the FADH . FADH then transfers an electron to the CPD to induce cyclobutane ring cleavage. The catalytic cycle is completed when the electron is transferred back from the repaired thymine to the FADH cofactor.

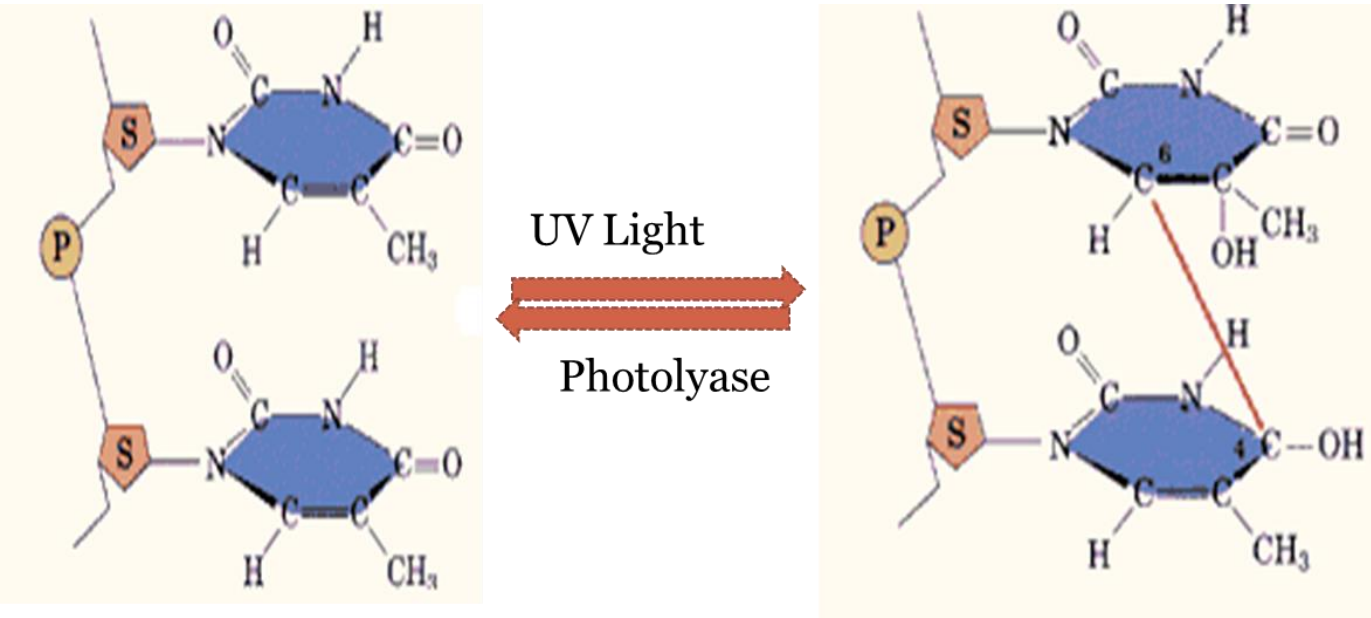


76. (6-4) Photolyase

UV irradiation also induces the formation of another type of pyrimidine dimer, the (6-4) photoproduct.



Takeshi Todo, Taisei Nomura and coworkrs reported in 1993 that *Drosophila melanogaster* has a photolyase that reverses (6-4) photoproduct lesions in the DNA. This photolyase was designated as (6 – 4) photolyase. It is widely distributed in plants and animals, but has not been detected in bacteria and mammals.



Considerably less information is known about the mechanism of action of (6-4) photolyase as compared to CPD photolyase; however, the two enzymes seem to work in a similar way. The (6-4) photolyase binds to damaged DNA, causing the (6-4) photoproduct to flip out of the DNA and into the active site of the enzyme. In the active site of the enzyme, it undergoes a rearrangement to produce a product that receives an electron from an excited FADH molecule. The final outcome is that the organisms containing (6-4) photolyase can use light energy to convert (6-4) photoproducts back to normal pyrimidine rings. Organisms can also repair dimer lesions introduced by UV light by excising damaged nucleotides and replacing them with normal nucleotides. This type of excision repair is the major pathay for repairing UV-induced damage to DNA in organisms such as humans that lack both the types of photolyases.

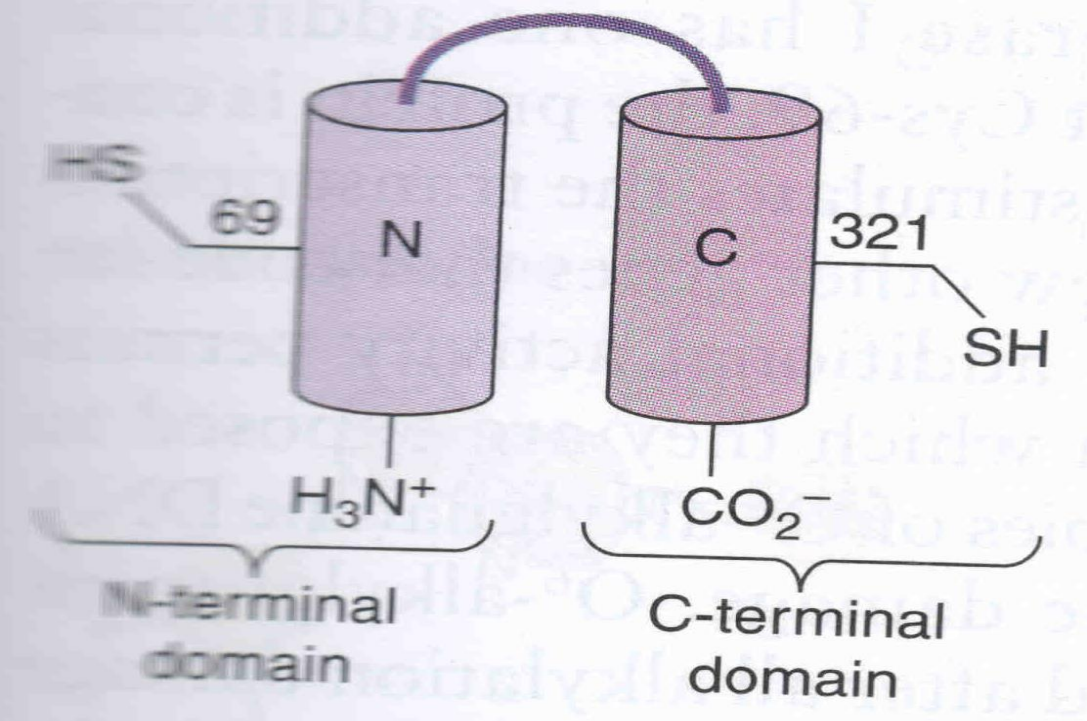
77. Damage Reversal by Dealkylation

Another means of direct reversal of DNA damage is by dealkylation. Direct dealkylation reactions have probably been most extensively studied in *E. coli*. Three different proteins can catalyze the direct removal of alkyl groups attached to oxygen atoms in DNA.
These include:-

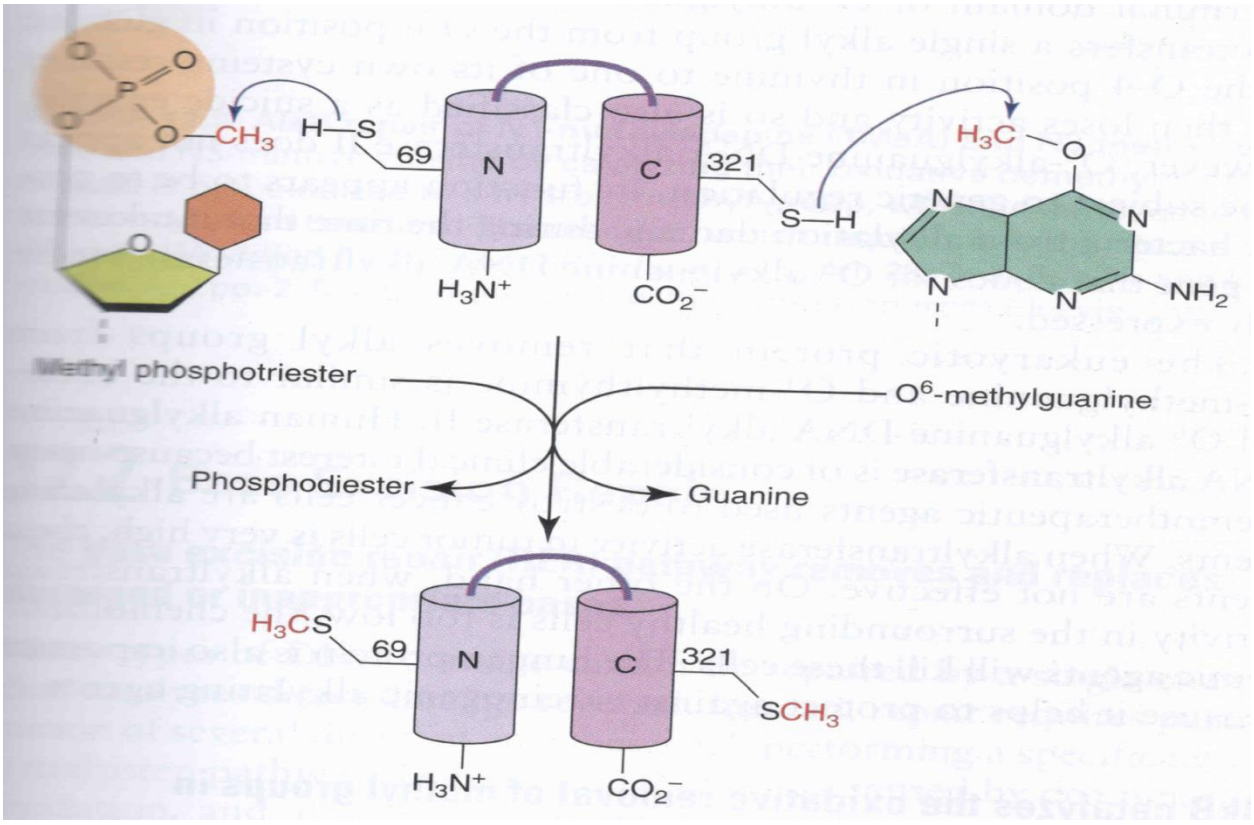
1) O^6-alkylguanine DNA alkyltransferase I
2) O^6-alkylguanine DNA alkyltransferase II
3) Alkylguanine DNA alkyltransferase

O^6-alkylguanine DNA alkyltransferase I can remove methyl & other alkyl groups attached to O-6 in guanine. It can also remove alkyl groups attached to O-4 of thymine and to phosphotriesters.

This enzyme is a monomer that has a flexible linker connecting its N- and C- terminal domains. Each domain has an active site that performs a specific function.



The N- terminal domain transfers an alkyl group from an Sp phosphotriester to one of its own cysteine residues, Cys-69.



The C- terminal domain transfers an alkyl group from either O^6-alkylguanine or O^4-alkylthymine to one of its own cysteine residues, Cys-321. Once alkylated, the protein can't be regenerated and therefore behaves more like an alkyl transferring agent than an enzyme. Such proteins as O^6-alkylguanine DNA alkyltransferase I, which lose their activity after acting one time, are called suicide enzymes. O^6-alkylguanine DNA alkyltransferase I, has one additional remarkable function. After methylation at Cys-69, this protein is converted to a transcriptional activator. This activator stimulates the transcription of the gene that codes for it as well as some other genes that code for the proteins that repair DNA damage. This additional activity permits the bacteria to adapt to environments in which they are exposed to alkylating agents by synthesizing more copies of these enzymes which can repair damaged DNA. O^6-alkylguanine DNA alkyltransferase I that is synthesized after all the alkylation damage has been repaired will remain unmethylated. This unmethylated form of the enzyme blocks transcription of the same genes that were activated by the methylated proteins.

78. Dealkylation Enzymes

O^6-alkylguanine DNA alkyltransferase II

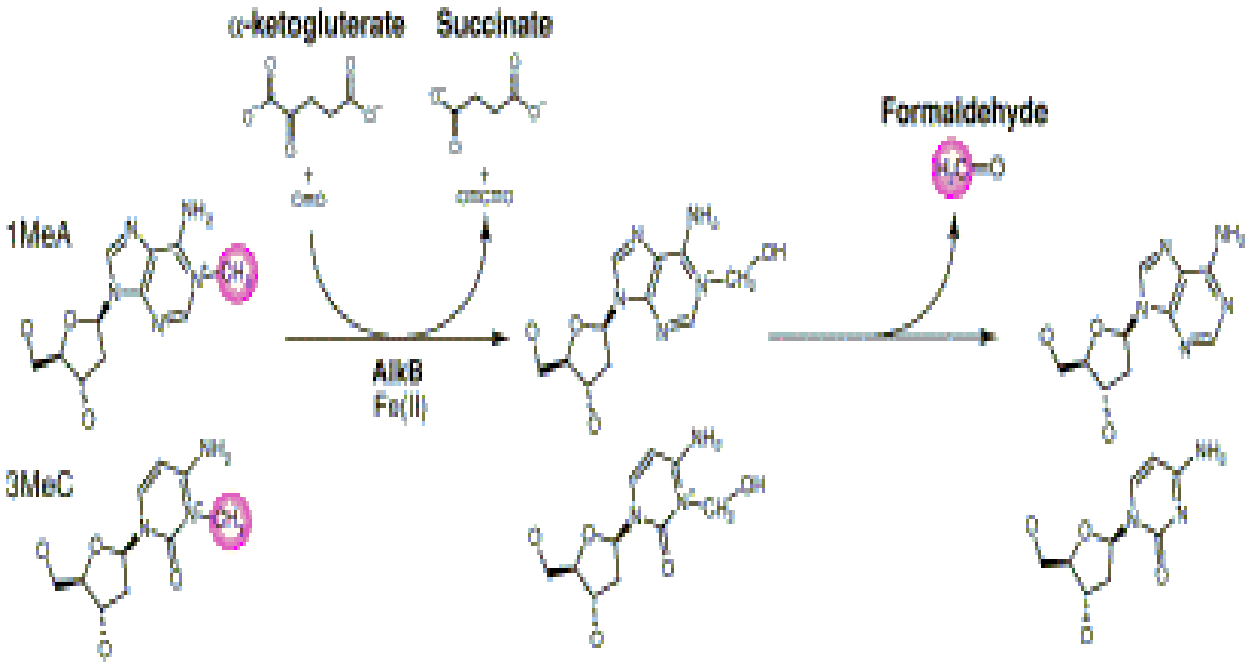
O^6-alkylguanine DNA alkyltransferase II is another E. coli's alkyl transferring enzyme. It has properties very similar to that of C-terminal domain of O^6-alkylguanine DNA alkyltransferase I. It also transfers a single alkyl group from the O – 6 position in guanine or the O-4 position in thymine to one of its own cysteine residues. After this transfer, the enzyme loses its activity and so is also classified as a suicide enzyme. However, this enzyme doesn't appear to be subjected to genetic regulation. Its function appears to be to protect bacteria from alkylation damage during the time that it takes for the gene of O^6-alkylguanine DNA alkyltransferase I to be fully expresed.

Alkylguanine DNA alkyltransferase

The eukaryotic protein that removes alkyl group from O^6-methylguanine or O^4-methylthymine is similar to the bacterial O^6-alkylguanine DNA alkyltransferase II. Human alkylguanine DNA alkyltransferase is of considerable clinical interest because many chemotherapeutic agents used to destroy cancer cells are alkylating agents. When alkyltransferase activity is very high in the tumor cells, these agents are not effective. On the other hand, when alkyltransferase activity in the surrounding healthy cells is too low, the chemotherapeutic agents will kill these cells. The human protein/enzyme is also important because it helps to protect the cells against carcinogenic alkylating agents.

AlkB

An entirely different type of alkylation damage repair activity was found in *E. Coli* in 2002 which is carried out by a protein AlkB. AlkB catalyzes the direct conversion of 1-methyladenine, 1-methylguanine, 3-methylcytosine and 3-methylthymine to adenine, guanine, cytosine and thymine, respectively. The AlkB catalyzed reaction requires, $Fe2^+$, molecular oxygen, and α-ketoglutarate.
A similar type of enzyme has been found in other organisms including human & mammals.

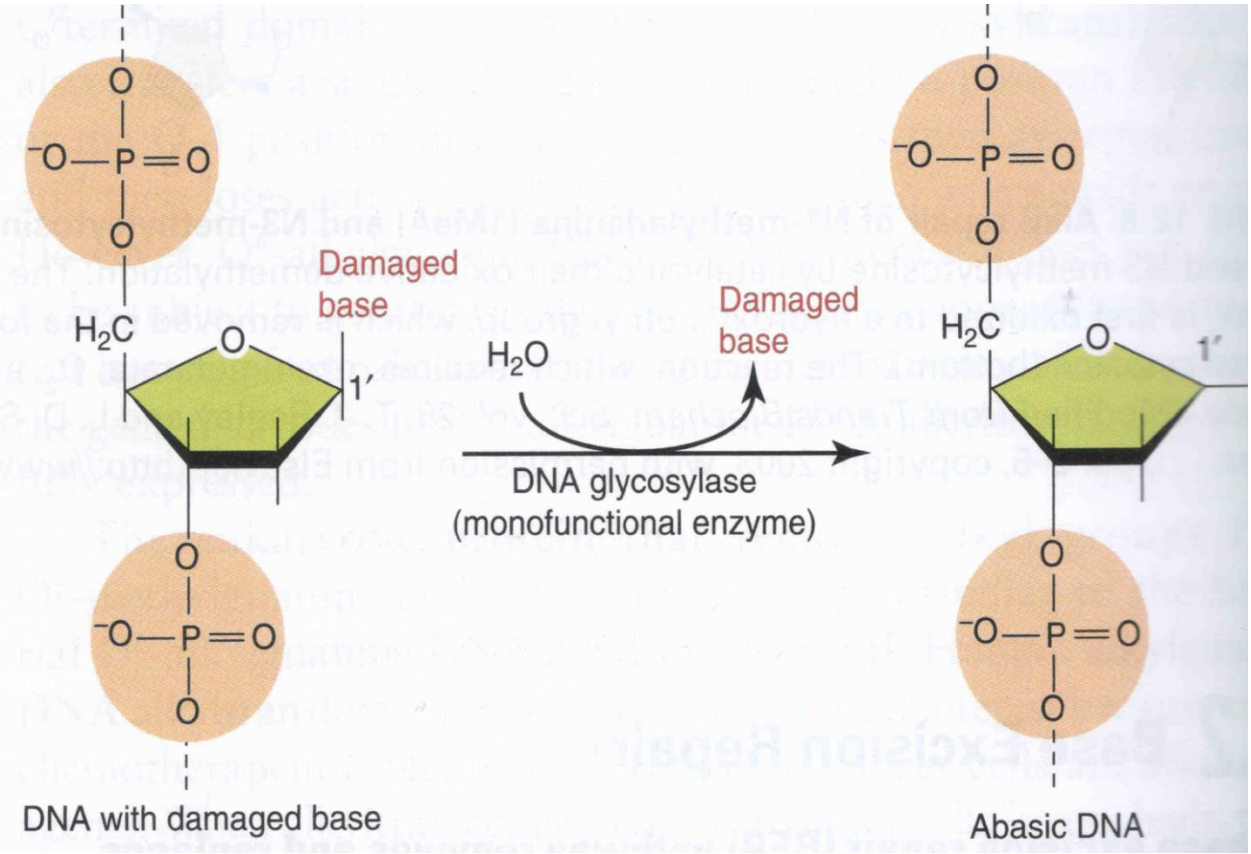


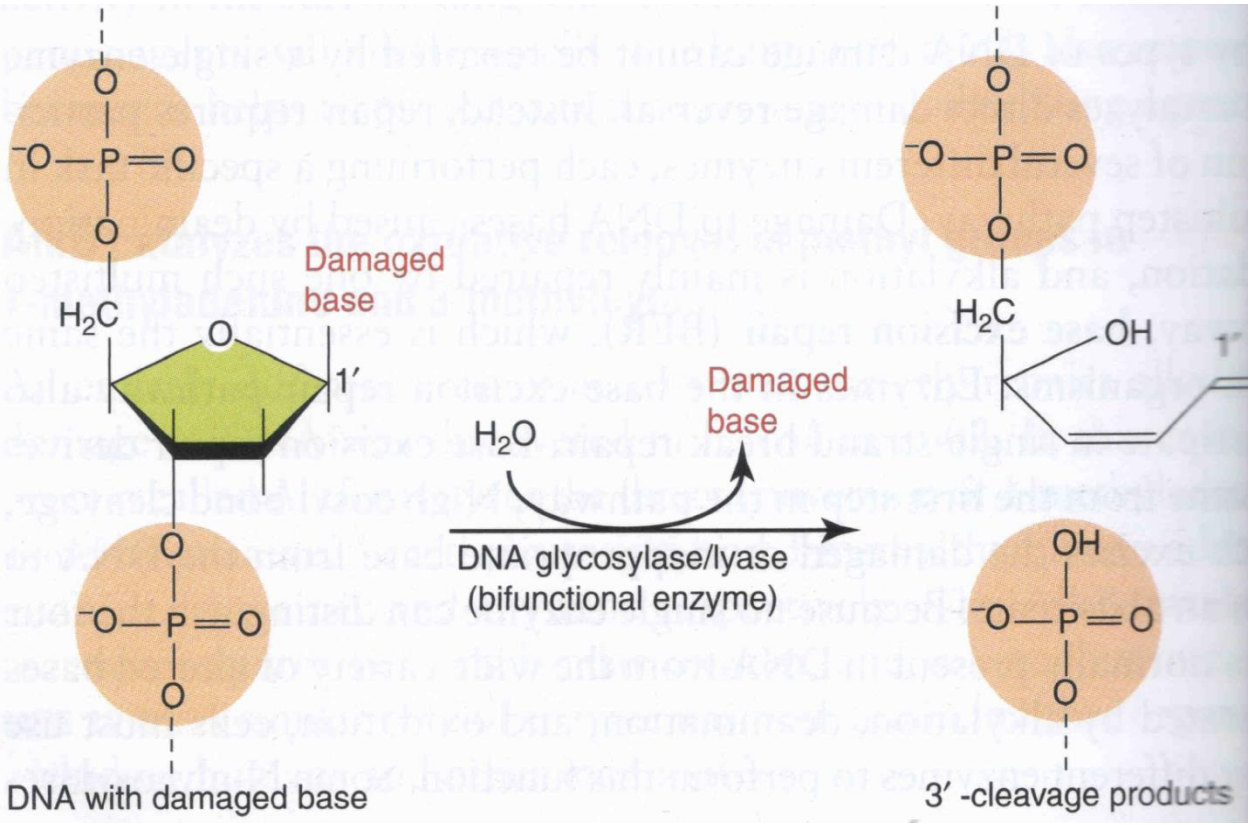
79. Base Excision Repair

Many types of DNA damages can't be repaired by a single enzyme that catalyzes direct damage reversal. Instead, repair requires the participation of several different enzymes, each performing a specific task in multistep pathway. Damage to DNA bases caused by deamination, oxidation, and alkylation is mainly repaired by one such multistep pathway which is called, Base Excision Repair (BER). BER pathway is same in all organisms. Enzymes involved in the base excision repair pathway also participate in the repair of single-strand break in DNA. Base excision repair derives its name from the first step in the pathway, N-glycosyl bond cleavage.

This cleavage excises the damaged or inappropriate base from the DNA to form an abasic site. Because no single enzyme can distinguish the four bases present in DNA from a wide variety of altered bases, cells must use many different enzymes to perform this function. Some N-glycosylases are monofunctional enzymes with only the ability to excise a damaged base.

Such enzymes are called DNA glycosylases. Other N-glycosylases have an AP lyase activity that cleaves the bond between sugar and the phosphate 3' to the damaged site. This enzymes is designated as DNA glycosylase/lyase.

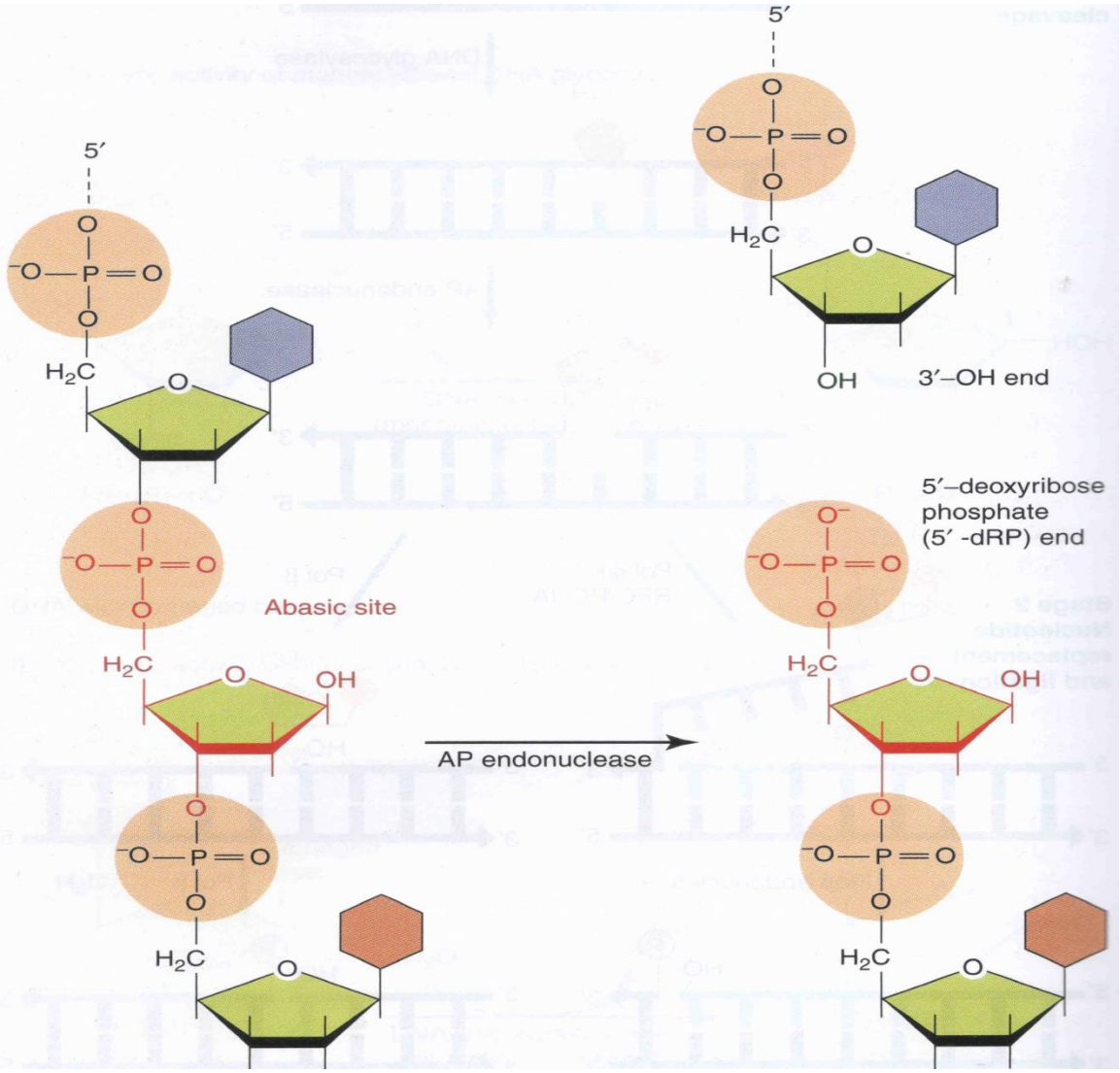

Damaged base
H_2O
Damaged base
DNA glycosylase
(monofunctional enzyme)
DNA with damaged base
Abasic DNA



Both types of enzymes detect the damaged base, flip it out of the DNA helix into an active site pocket, and then cleave the N-glycosyl bond. Although some glycosylases excise a specific base, most have a somewhat broader specificity. The *E.coli* enzyme Uracil N-glycosylase (Ung) is specific for uracil. Whereas another *E. coli* N-glycosylase, 3-methyladenine DNA glycosylase (AlkA), acts on 3- or 7-methyl purines, 3- or 7-ethylpurine, ethenoadenine, and O^2-methyl pyrimidine. A null mutation in a gene that codes for a DNA glycosylase or DNA glycosylase/lyase is not lethal which probably reflects overlapping functional abilities among the glycosylases.

80. Base Excision Repair Pathway

The DNA glycosylase catalyzes base excision to produce an AP (apurinic and apyrimidininc) site. The next enzyme in the pathway, AP endonuclease, hydrolyzes the phosphodiester bond 5' to the AP site to generate a nick.



E.coli has two well-characterized AP endonucleases:-

- Exonuclease III (Xth)
- Endonuclease IV (Nfo)

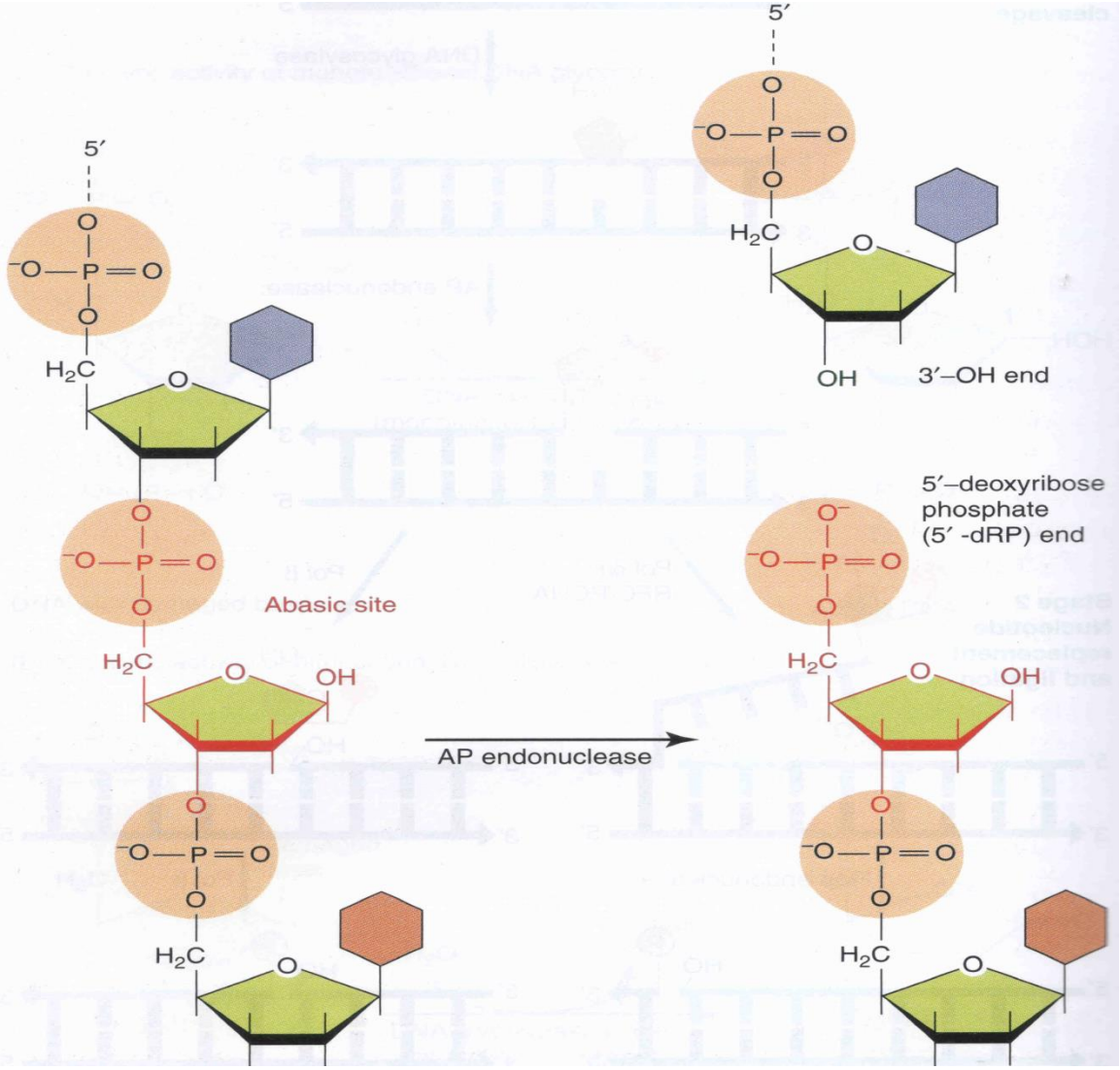
Exonuclease III (Xth) despite its name accounts for most of the bacterial AP endonuclease activity. Both are multifunctional enzymes that have 3'-phosphate and 3'-repair phosphodiesterase activities.

The former activity removes phosphate groups from the 3' end of a DNA strand. The latter activity i.e., 3'-repair phosphodiesterase activity removes the 3'-unsaturated aldehydic group produced by DNA glycosylase/lyase action. These activities are important because DNA polymerase can't attach new nucleotides to a blocked 3'-end.

The mammalian AP endonuclease, APE1 is homologous to *E. coli* exonuclease III.

81. Short Patch Repair

Base excision by DNA glycosylase and strand cleavage by AP endonuclease introduces a gap with a 5'-deoxyribose phosphate (5'-dRP) on one side and a 3'-OH on the other.



Additional enzymes are required to fill in the gap and remove the 5'-deoxyribose phosphate. Cells can repair the damage by two different pathways viz. short patch repair & long patch repair.

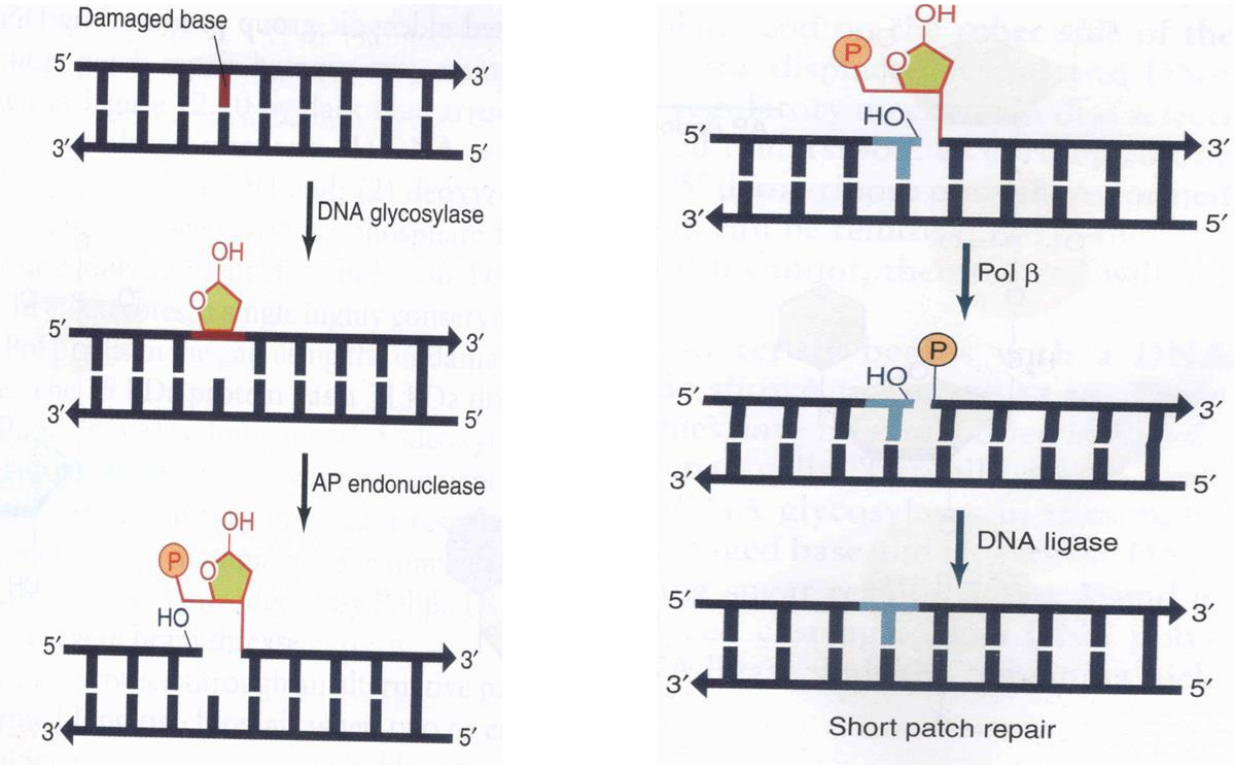
In the short patch repair, only a single nucleotide is replaced. This repair pathway involves the following enzyme catalyzed reactions:-

(1) DNA polymerase adds a deoxyribonucleotide to the 3'-OH end.

(2) Deoxyribose phosphate lyase (dRPase) removes 5'-deoxyribose phosphate from the 5'-end; and

(3) DNA ligase joins the adjacent ends.

In *E. coli*, DNA polymerase I fills in the gap. While in eukaryotes, a single highly conserved enzyme, DNA polymerase β (Pol β) fills in the gap using the undamaged DNA strand as the template.

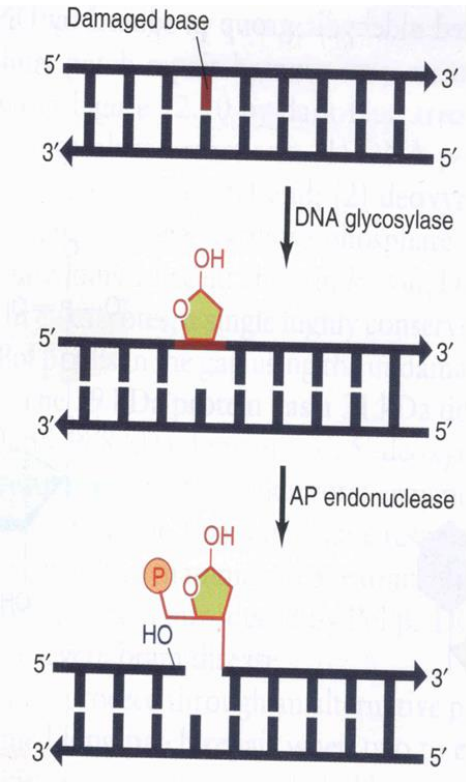


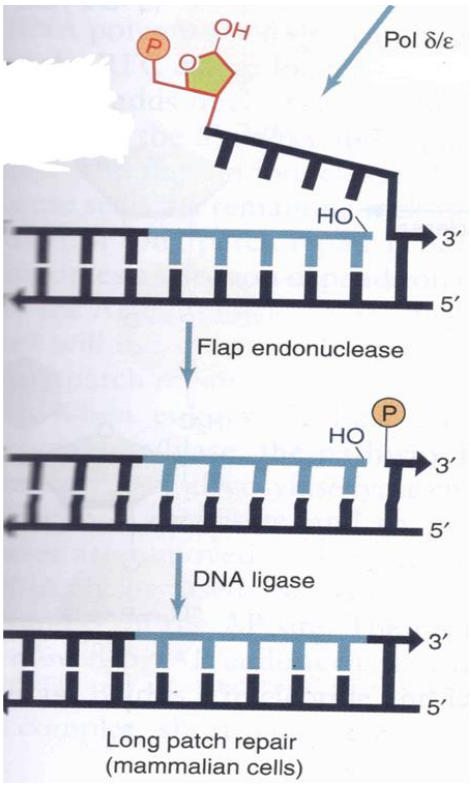
The eukaryotic enzyme differs from its bacterial counterpart in one very important respect, i.e., it lacks 3' → 5' proofreading activity. Two mammalian 3' exonucleases, TREX1 and TREX2 may correct the errors introduced by Pol β.

82. Long Patch Repair

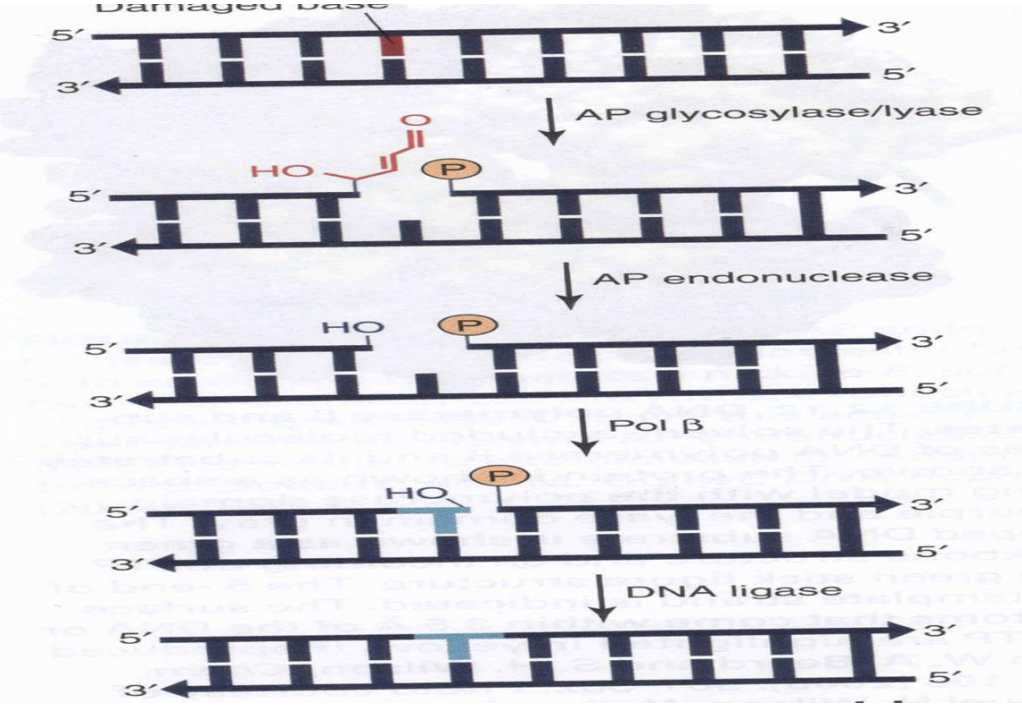
The alternative pathway to repair the damage in mammalian cells is termed as long patch repair. In this pathway, 2 – 8 nucleotides are replaced. In this case, DNA polymerase δ or Ɛ catalyzes chain extension with the assistance of the RFC (replication factor C) clamp loader and the PCNA sliding clamp. As the polymerase adds nucleotides to the 3'-OH end on one side of the gap, it displaces the 5'- deoxyribose phosphate end on the other side of the gap. The flap endonuclease cleaves the displaced strand and DNA ligase seals the remaining nick.

The regulatory mechanism that selects long or short patch repair is not well understood.

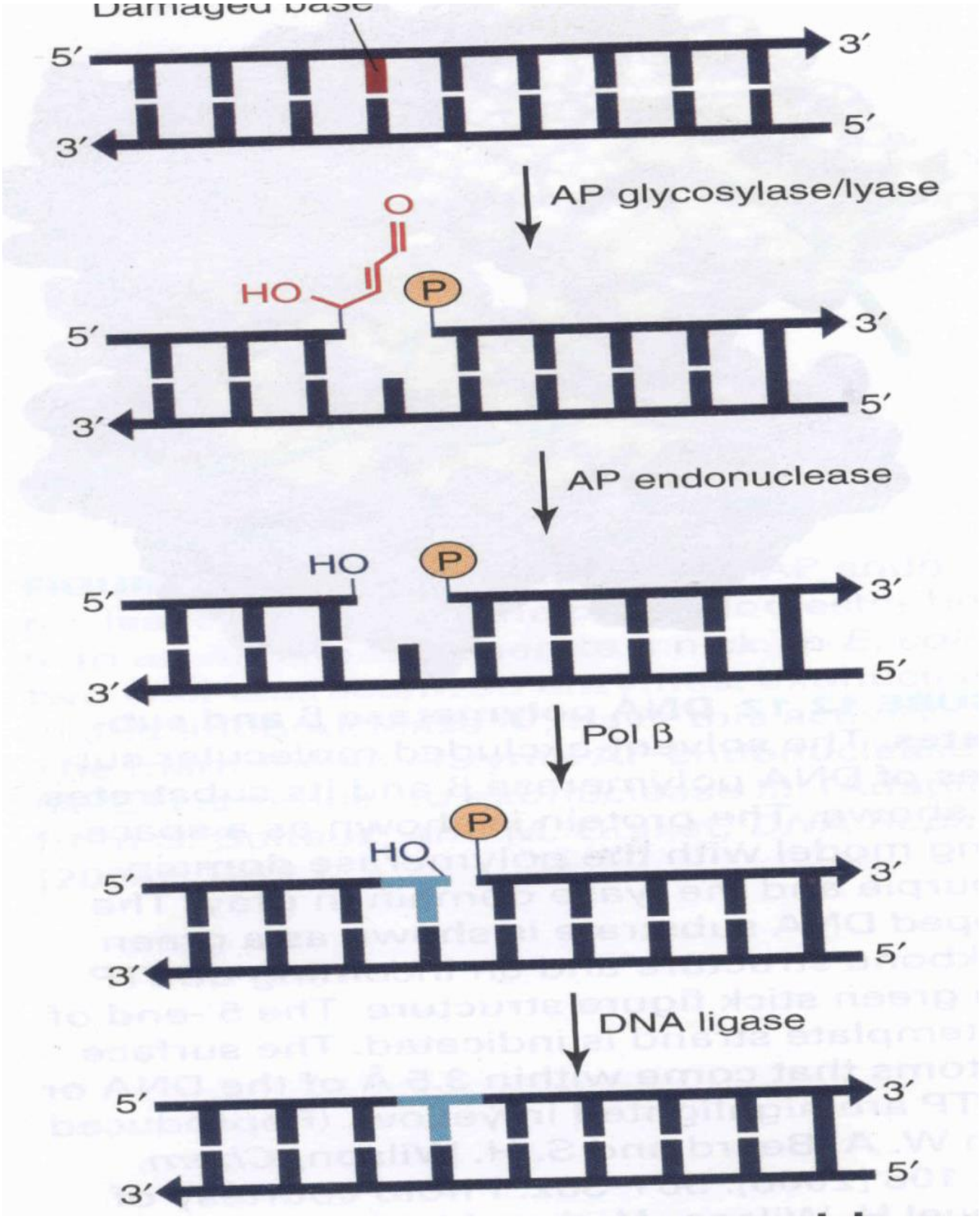




When eukaryotic base excision repair begins with a DNA glycosylase/lyase, the pathway is shown below:-



Three distinct DNA glycosylase/lyase enzymes have been identified in *E. coli*, three in *S. cerevisiae*, and six in human cells. Virtually all oxidized bases are removed by bifunctional DNA glycosylases in mammals. DNA glycosylase/ lyase excises the damaged base and cleaves the DNA strand 3' of the AP site. The resulting sugar residue at the 3'-end is removed by the AP endonuclease catalyzed cleavage. Then DNA polymerase β adds a nucleotide and DNA ligase seals the remaining nick to complete short patch repair.


Damaged base
5′
3′
3′
5′
AP glycosylase/lyase
O
HO
P
5′
3′
3′
5′
AP endonuclease
HO
P
5′
3′
3′
5′
Pol β
P
HO
5′
3′
3′
5′
DNA ligase
5′
3′
3′
5′

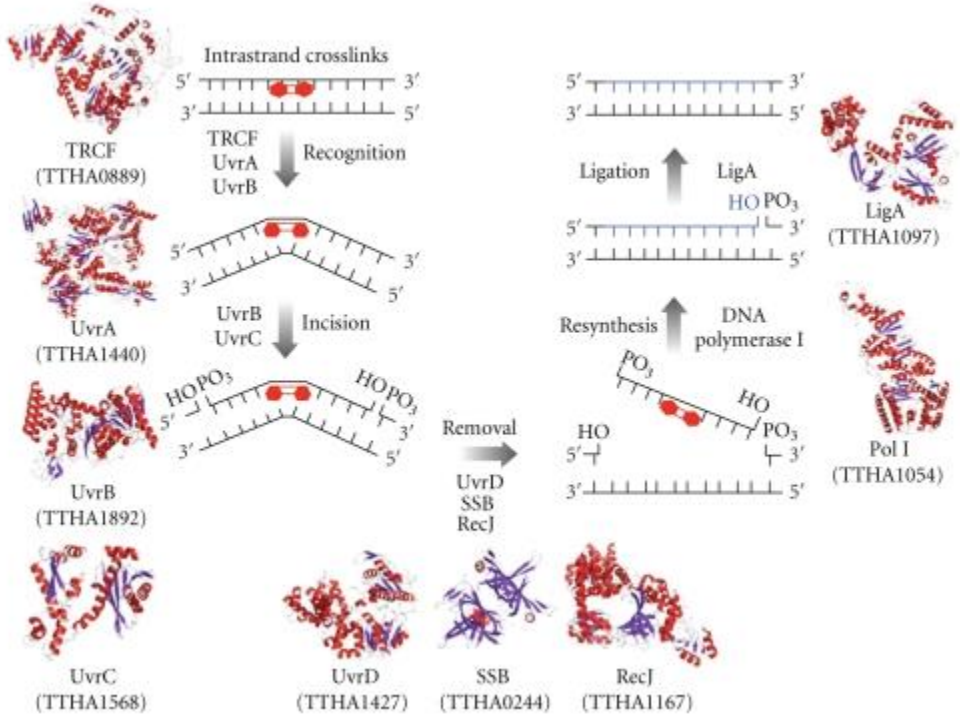
83. Nucleotide Excision Repair

Nucleotide excision repair (NER) pathway removes bulky adducts from DNA by excising an oligonucleotide bearing the lesion and replacing it with new DNA. This repair mechanism excises UV-induced cyclobutane pyrimidine dimers, (6-4) photoproducts, damaged bases formed by alkylating agents and certain types of cross-links. The efficiency of repair for different kinds of lesions can vary a lot.

In general, there is a direct relationship between the amount of helical distortions produced by lesion and the efficiency of this repair. The basic nucleotide excision repair pathway is same in all the organisms.It involves the following steps:-

1. damage recognition
2. an incision in the damaged DNA strand on each side of the lesion,
3. excision of the oligonucleotide created by the incision,
4. synthesis of new DNA to replace the excised DNA segment using the undamaged DNA strand as a template, and
5. ligation of the remaining nick.

Although the basic nucleotide excision repair pathway is similar in all the organisms, there are considerable differences in the proteins that carry out the various steps.



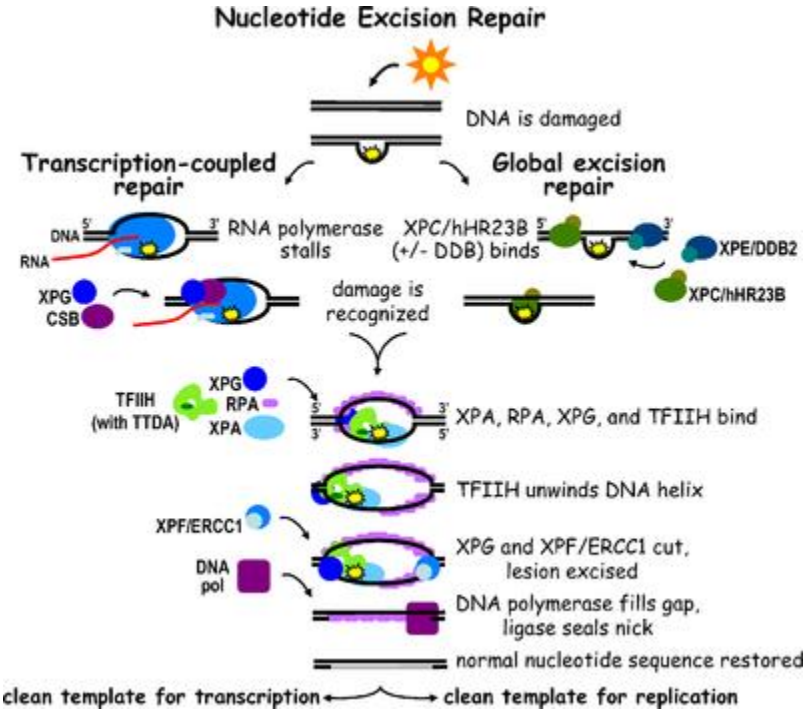
84. Nucleotide Excision Repair of UV-induced Damage

UV-irradiated *E. coli* can regain their ability to survive after incubation in dark. However, they recover more slowly than when incubated in the light. This observation suggests that the bacteria use some process other than photoreactivation to repair the UV-induced DNA damage. Richard Setlow & William Carrier and, working independantly, Richard Boyce & Paul Howard-Flanders, used a similar approach to investigate this alternative process in 1964. Both groups cultured *E. coli* in the presence of [^{3}H]thymine to label the DNA and then irradiated the cells with UV light to induce the formation of cyclobutane thymine dimer.

Then they:-

1) incubated the UV-irradiated cells in the dark so that the photoreactivation could not take place;
2) removed samples after various incubation times and added TCA to them;
3) separated acid-insoluble DNA from acid soluble nucleotides;
4) digested the DNA and oligonucleotides to release intact thymine cyclobutane dimers; and
5) detected the released dimers by chromatography.

The experiments revealed that as the incubation time in the dark increases, Cyclobutane thymine dimers disappear from the acid-insoluble DNA and appear in the acid soluble oligonucleotide fraction.These results were correctly interpreted to mean that bacteria can excise an oligonucleotide containing a lesion and replace the excised oligonucleotides with newly synthesized DNA. Subsequent studies showed that eukaryotes and the archaea also have nucleotide excision repair pathways.



85. UvrA, UvrB, and UvrC Proteins

UvrABC endonuclease is a multienzyme complex in *Escherichia coli* involved in DNA repair by nucleotide excision repair, and it is, therefore, sometimes called an excinuclease. This UvrABC repair process, sometimes called the short-patch process, involves the removal of twelve nucleotides where a genetic mutation has occurred followed by a DNA polymerase, replacing these aberrant nucleotides with the correct nucleotides and completing the DNA repair. The subunits for this enzyme are encoded in the *uvrA*, *uvrB*, and *uvrC* genes. This enzyme complex is able to repair many different types of damage, including cyclobutyl dimer formation.

Genetic studies revealed that three *E. coli* genes viz., *uvrA*, *uvrB*, and *uvrC*, code for proteins that are essential for damage recognition, incision, and excision. All three genes have been cloned and the proteins that they encode (UvrA, UvrB, and UvrC) have been purified and characterized. Under normal physiological conditions, *E. coli* has about 25 molecules of UvrA, 250 molecules of UvrB, and 10 molecules of UvrC. After UV damage of DNA, UvrA and UvrB levels increase ten- and four folds, but the UvrC level remains the same. Although UvrA, UvrB, and UvrC do not combine to form a stable ternary complex, the polypeptides nevertheless are said to be part of a UvrABC damage-specific endonuclease. The UvrABC damage-specific endonuclease is called as UvrABC endonuclease in short but some investigators preferably term them as UvrABC excinuclease. UvrABC excinuclease is termed so because the proteins participate in incision and excision reactions. The three polypeptides work in the order suggested by their names, that is their order of action is UvrA, UvrB, and the UvrC.

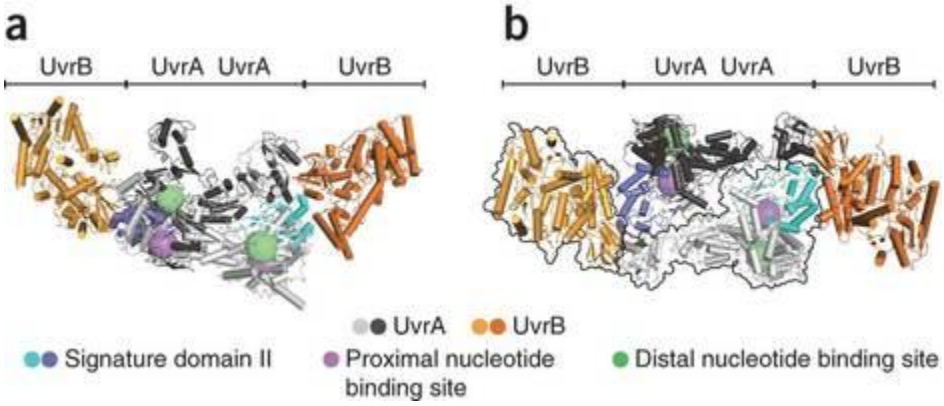


Fig. Architecture of the UvrA–UvrB DNA damage sensor.close

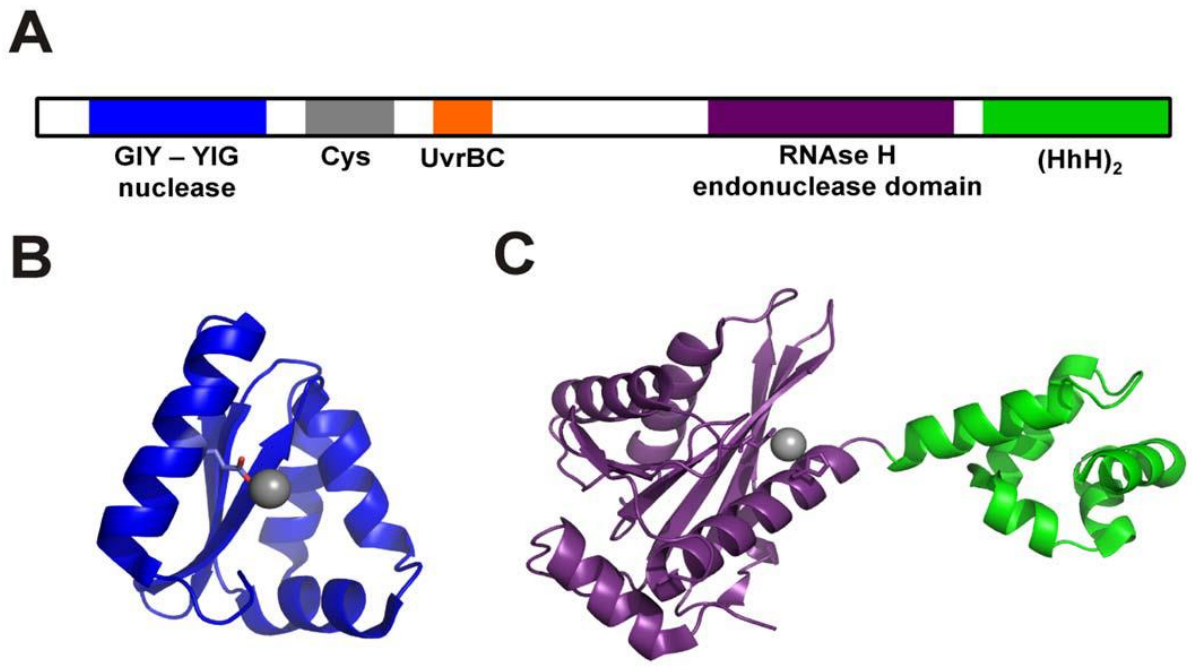


Fig. (A) UvrC consists of two endonuclease domains, particularly an N-terminal GIY-YIG nuclease domain (blue) and a C-terminal RNAse H-like endonuclease domain (purple). UvrC also comprises a Cys-rich region (gray), an UvrB-interacting domain (orange) and a C-terminal tandem Helix-hairpin-helix motif (green). (B) Crystal structure of the N-terminal GIY-YIG endonuclease domain of T. maritima UvrC (PDB: 1YD1) (Truglio et al, 2005). The residue E76 coordinating the Mg2+-ion (gray sphere) is shown as stick model. (C) Crystal structure of the C-terminal RNAse H-like endonuclease domain (purple) and the (HhH)2-domain (green) of T. maritima UvrC (PDB entry: 2NRZ) (Karakas et al, 2007). The catalytic triade DDH consisting of residues D367, D429 and H488 which coordinates one Mn2+-ion (gray sphere) is represented as stick model.

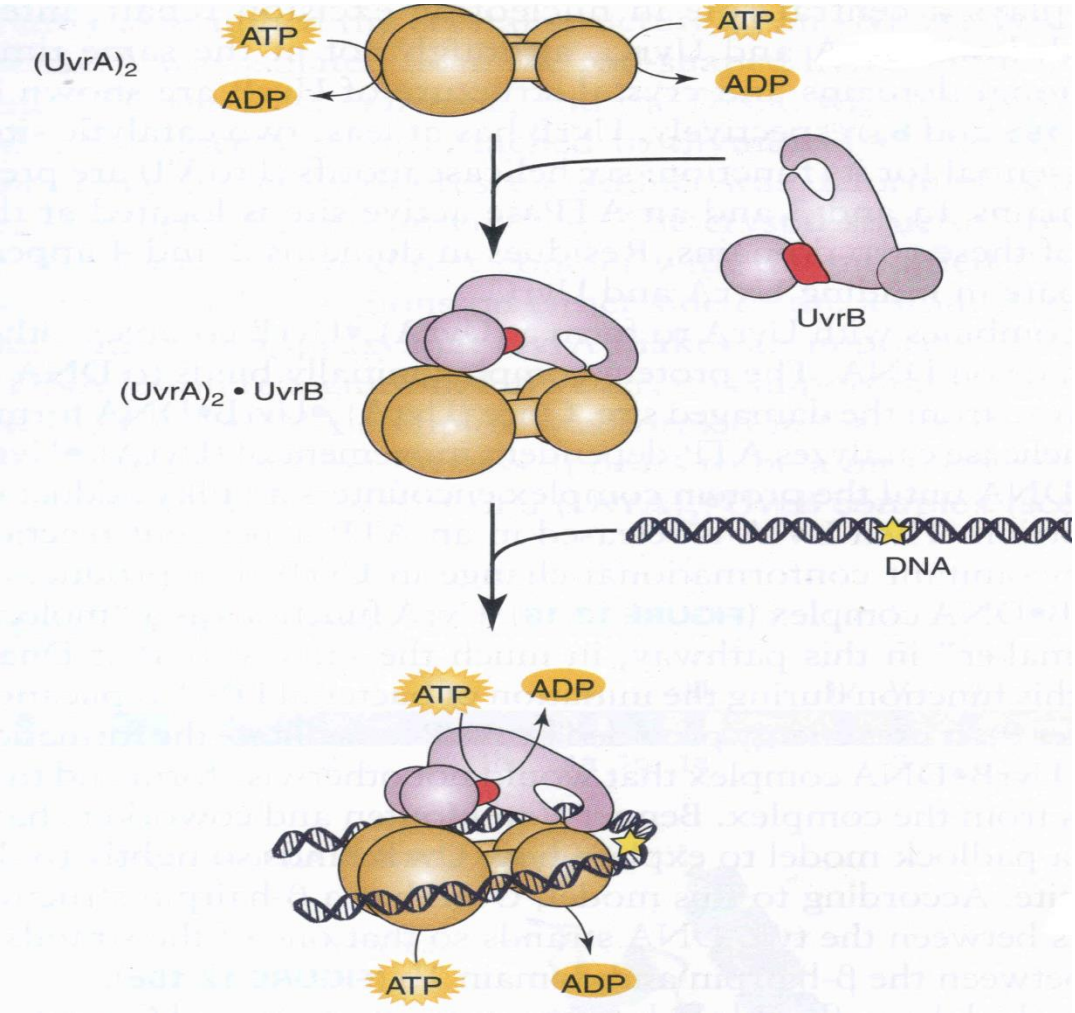
86. The NER Pathway

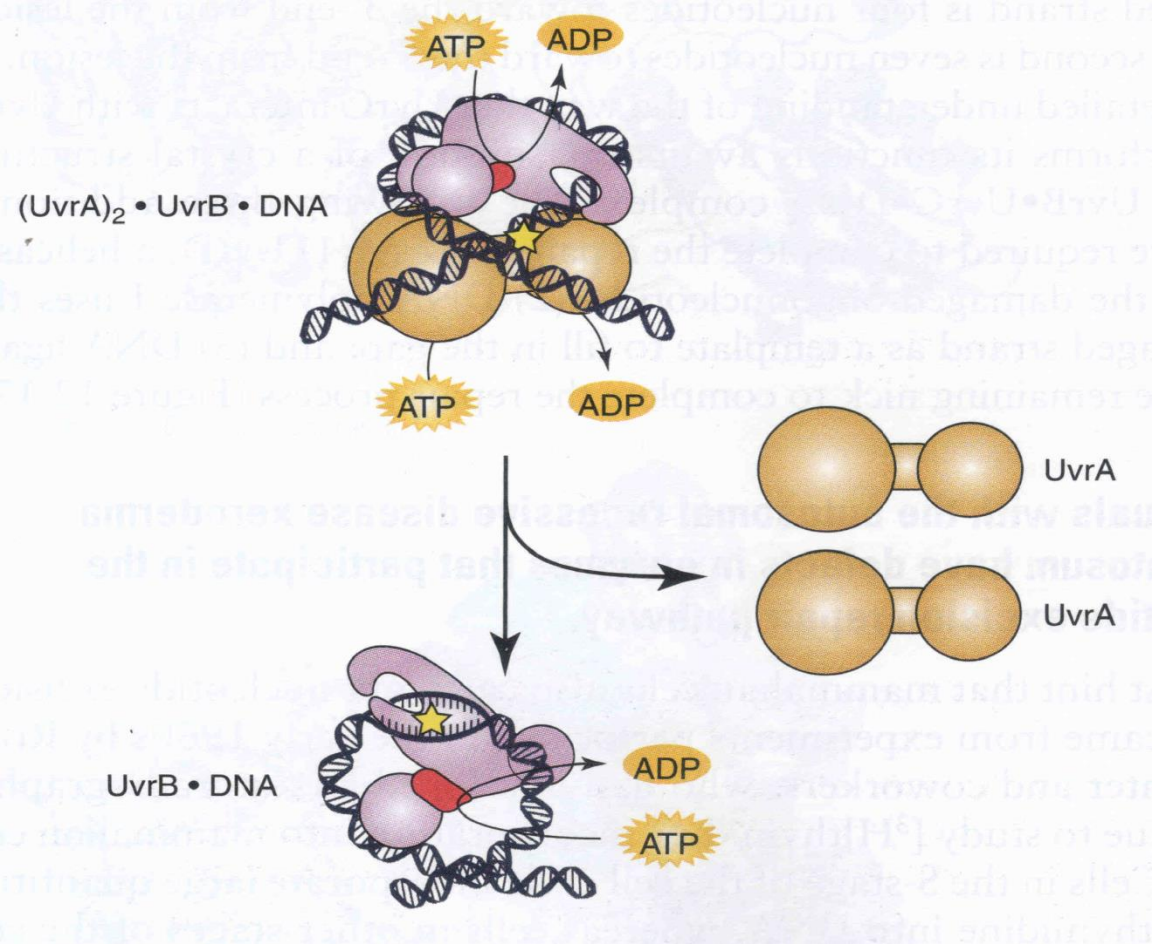
The crystal structure of UvrA•DNA complex reveals that UvrA does not make direct contact with the modified thymine but does bind to DNA regions on either side of the lesion. Based on this information, it appears that UvrA makes an important contribution towards recognition of DNA damaged lesion.

This UvrA•DNA complex is formed in vitro in the absence of UvrB. The recognition process appears to be more complicated in vivo where UvrA is a part of a (UvrA)2•UvrB complex. UvrB plays a central role in nucleotide excision repair by interacting with both the UvrA and UvrB, although not at the same time.

UvrB has at least two catalytic sites that are essential for its function. UvrB combines with UvrA to form a (UvrA)2•UvrB complex either in solution or on DNA.

Initially, this protein complex binds to DNA at some distance from the damaged site. Once (UvrA)2•UvrB • DNA complex is formed, the UvrB helicase catalyzes ATP-dependant movement of (UvrA)2•UvrB along the DNA until the protein complex encounters a bulky adduct or helix distortion. The UvrA is released in an ATP-dependant reaction with a concomitant conformational change in the UvrB that produces a stable UvrB•DNA complex.

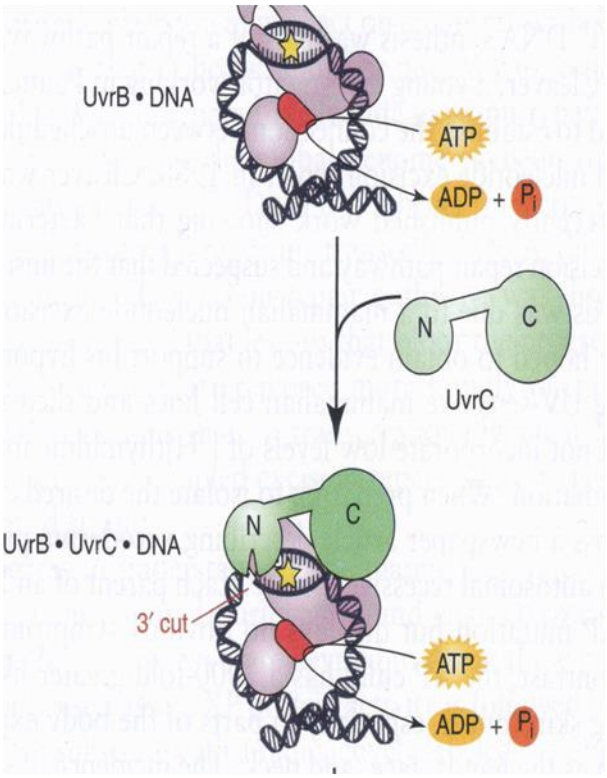

ATP
ADP
$(UvrA)_2$
ATP
ADP
UvrB
$(UvrA)_2$ • UvrB
DNA
ATP
ADP
ATP
ADP

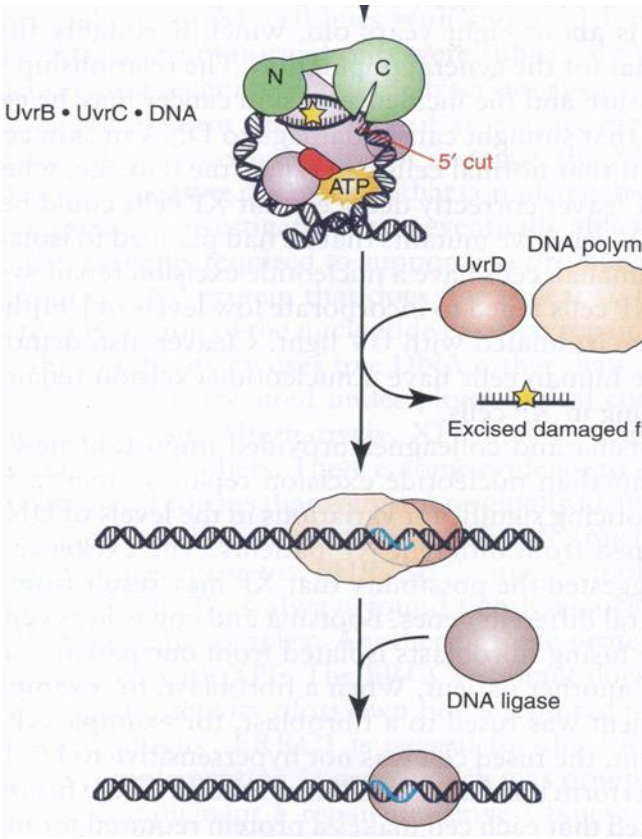


UvrA functions as a "molecular match maker" in this pathway. It uses energy of ATP to facilitate the formation of UvrB•DNA complex and then dissociates from the complex. UvrC, which has a flexible linker that connects its N- and C-terminal domains, binds to the UvrB•DNA complex and makes two incisions, one on each side of the lesion. The first incision of the damaged strand is four nucleotides towards the 3'-end from the lesion and the second is seven nucleotides toward the 5'-end from the lesion. How UvrC interacts with UvrB and performs its function is still to be described.

Following three more steps are required to complete the repair process:-

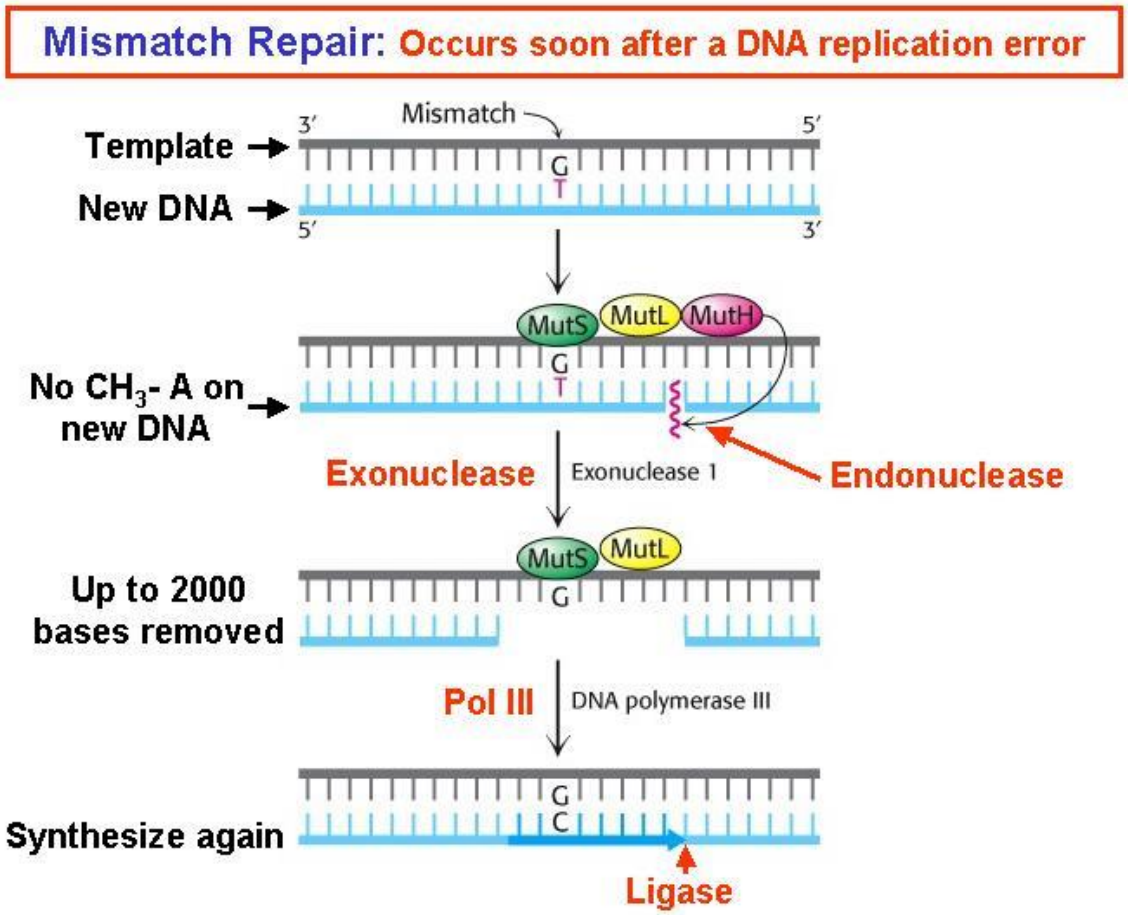
- 1). UvrD, a helicase, excises the damaged oligonucleotide;
- 2). DNA polymerase I uses the undamaged strand as a template to fill the gap;
- 3). DNA ligase seals the remaining nick to complete the repair process.





87. Mismatch Repair

The mismatch repair system corrects rare base pair mismatches and short insertions or deletions that appear in DNA following replication. DNA polymerases introduce about one mispaired nucleotide per 10^5 nucleotides.
However, the 3' → 5' proofreading exonuclease increases replication fidelity by 100-fold by removing mispaired nucleotides. Although an error frequency of one nucleotide in 10^7 may seem extremely low, it would result in a high mutation rate. The second type of error that occur during replication is the short insertions and deletions, which result from the fact that repeated-sequence motifs sometimes dissociate and then re-anneal incorrectly. As a result, the newly synthesized strand will have a different number of repeats than the template strand.
Introduction of an insertion or deletion into the newly synthesized DNA is likely to produce a mutation. Cells with a non-functional mismatch repair system have a high rate of mutation due to their inability to efficiently repair base pair mismatches, short insertions or deletions that arise during replication.

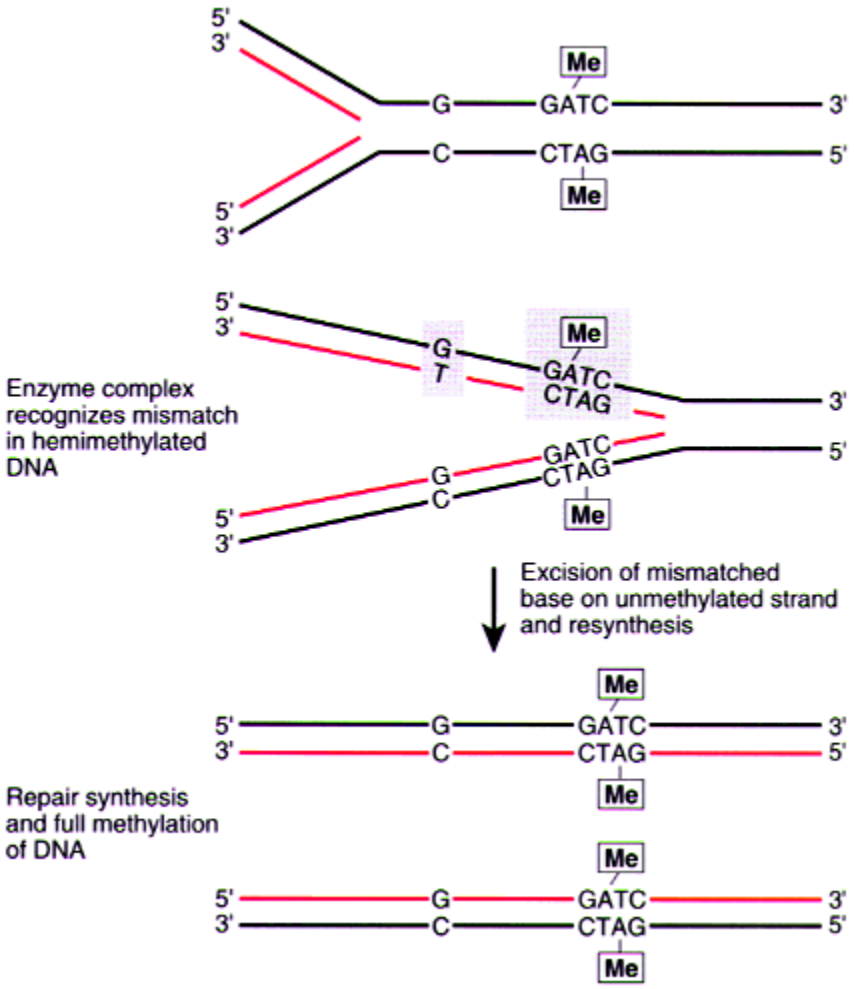


88. Mismatch Repair System in *E. coli*

Let us begin the examination of mismatch repair by considering the *E. coli* mismatch repair system because this system has been the most extensively studied. Although this system provides valueable informations for studying mismatch repair in other organisms, it differs from the mismatch repair systems of gram-positive bacteria & eukaryotes in one important respect. The *E. coli* mismatch repair system can distinguish a newly synthesized strand from a parental strand because only the latter has methyl groups attached to sites with the sequence GATC. *E. coli* has a deoxy- adenosine methylase that transfers methyl groups from S-adenosylmethionine molecules to deoxyadenosines in GATC sequences. The time of methylation by deoxyadenosine methylase, however, lags behind that of nucleotide addition at the replication fork about two minutes, so the newly synthesized strand is transiently unmethylated. The *E. coli* mismatch repair system exploits this period of transient unmethylation to identify and cut GATC sites in a newly synthesized strand with a mismatch.

GATC sequences in DNA are normally methylated (Me) at the 6 position of adenine. During semiconservative DNA synthesis, a G-T mismatch arises in one of the sister DNA duplexes. The enzymatic mechanism for repairing this lesion depends on discrimination between the newly

synthesized (red) and parental (black) strands. This is achieved by recognition of the temporary lack of methylation of the newly synthesized strand before postreplicative DNA methylation takes place. The nonmethylated daughter strand containing the incorrect base is enzymatically attacked by mismatch correction enzymes, and the misincorporated base is removed. Repair synthesis and daughter-strand methylation at GATC sequences restore the sister DNA duplexes to their native state.



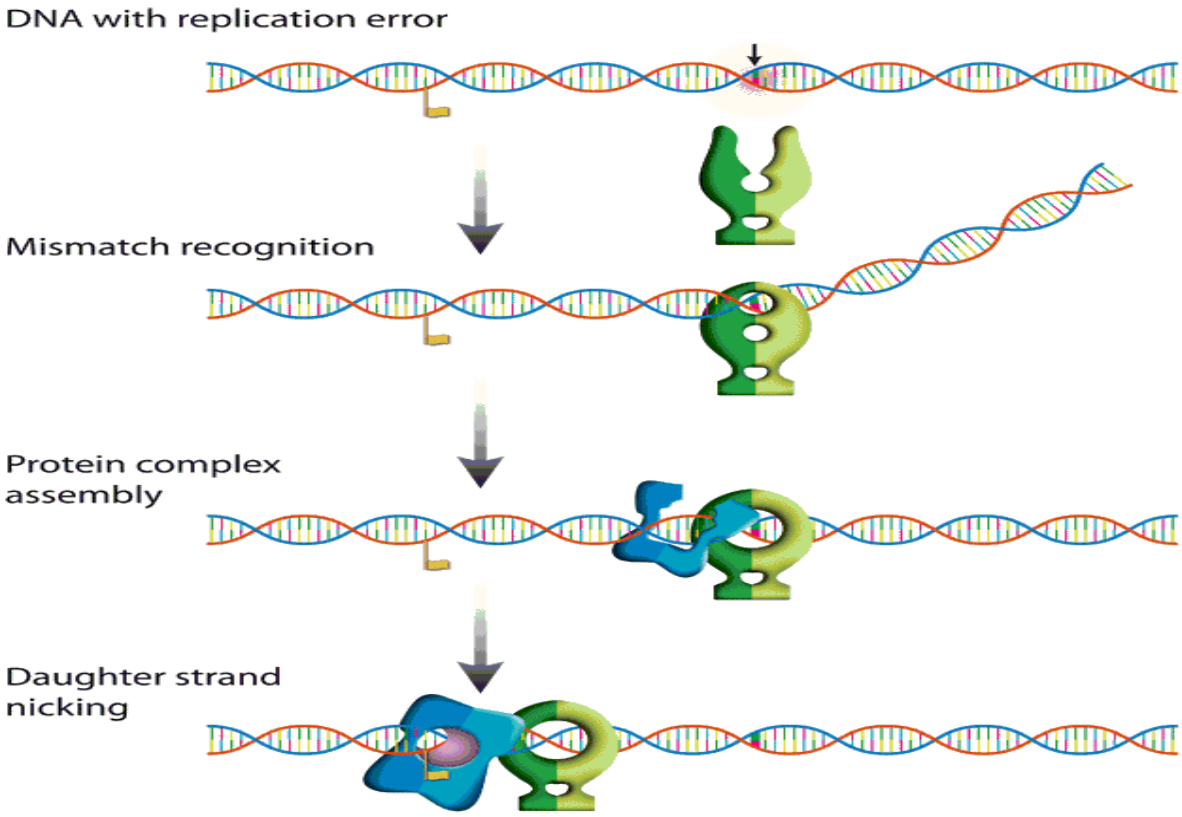
89. MutS, MutL, & MutH Proteins

Genetic and biochemical studies have demonstrated that three *E. coli* proteins viz., MutS, MutL, and MutH are dedicated to mismatch repair. Although these proteins are essential for mismatch repair, they are not sufficient. Several additional enzymes and protein factors also make important contributions. Among these enzymes and protein factors are:-

- i. DNA helicase II (UvrD),
- ii. Single-stranded DNA binding protein (SSB)

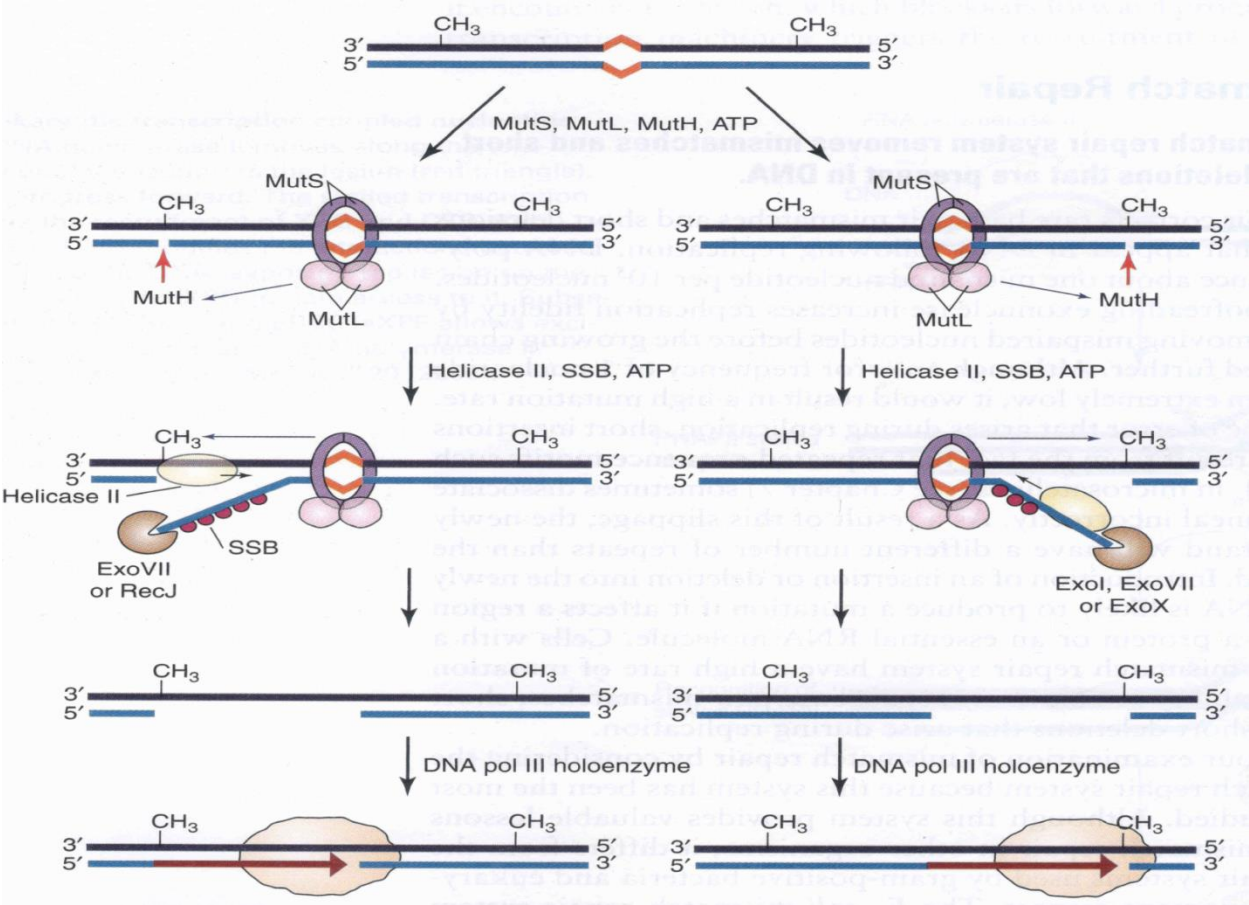
- iii. 5’⟶3’ exonucleases (ExoVII or RecJ)
- iv. 3’⟶5’ exonucleases (ExoI, ExoVII, or ExoX),
- v. DNA polymerase III holoenzyme,
- vi. DNA ligase, and
- vii. Deoxyadenosine methylase.

The process of repair begins when MutS (a homodimer or homotatramer) binds to the mismatch. MutS recruits MutL (a homodimer) in an ATP-dependant fashion. Then the MutS•MutL complex activates MutH, which makes an incision at the nearest unmethylated GATC site, either 5’ or 3’ to the mismatch, in the newly synthesized strand.



MutH shares sequence homology with the type II restriction endonuclease, Sau3AI. Both enzymes recognize and cleave GATC sequences. MutH does not bind or cleave fully methylated GATC sites, whereas Sau3AI cleaves fully, hemi and unmethylated GATC sites. Next in the process is that helicaseII (UvrD) unwinds the DNA and SSB binds to the resulting single strands. When the incision is 5’ to the mismatch, ExoVII or RecJ exonucleases hydrolyze the nicked strand in a 5’ 3’ direction. When the incision is 3’ to the mismatch, ExoI, ExoVII or ExoX exonucleases hydrolyze the nicked strand in a 3’ 5’ direction. DNA polymerase III

holoenzyme fills the gap with new DNA. DNA ligase seals the remaining nick and deoxyadenosine methylase adds a methyl group to the GATC site.



All organisms that have a mismatch repair system have MutS and MutL homologs. However, MutH is present only in *E. coli* and some other gram-negative bacteria.

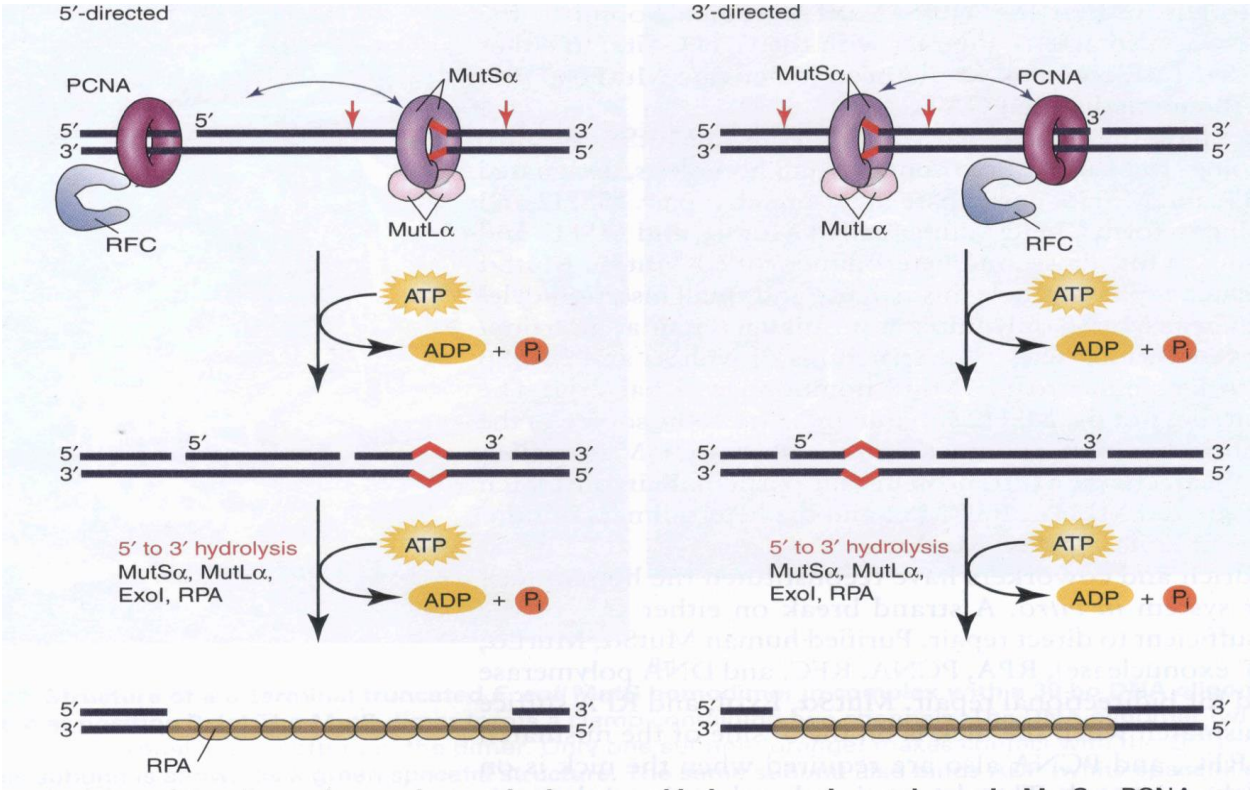
90. Mismatch Repair in Eukaryotes

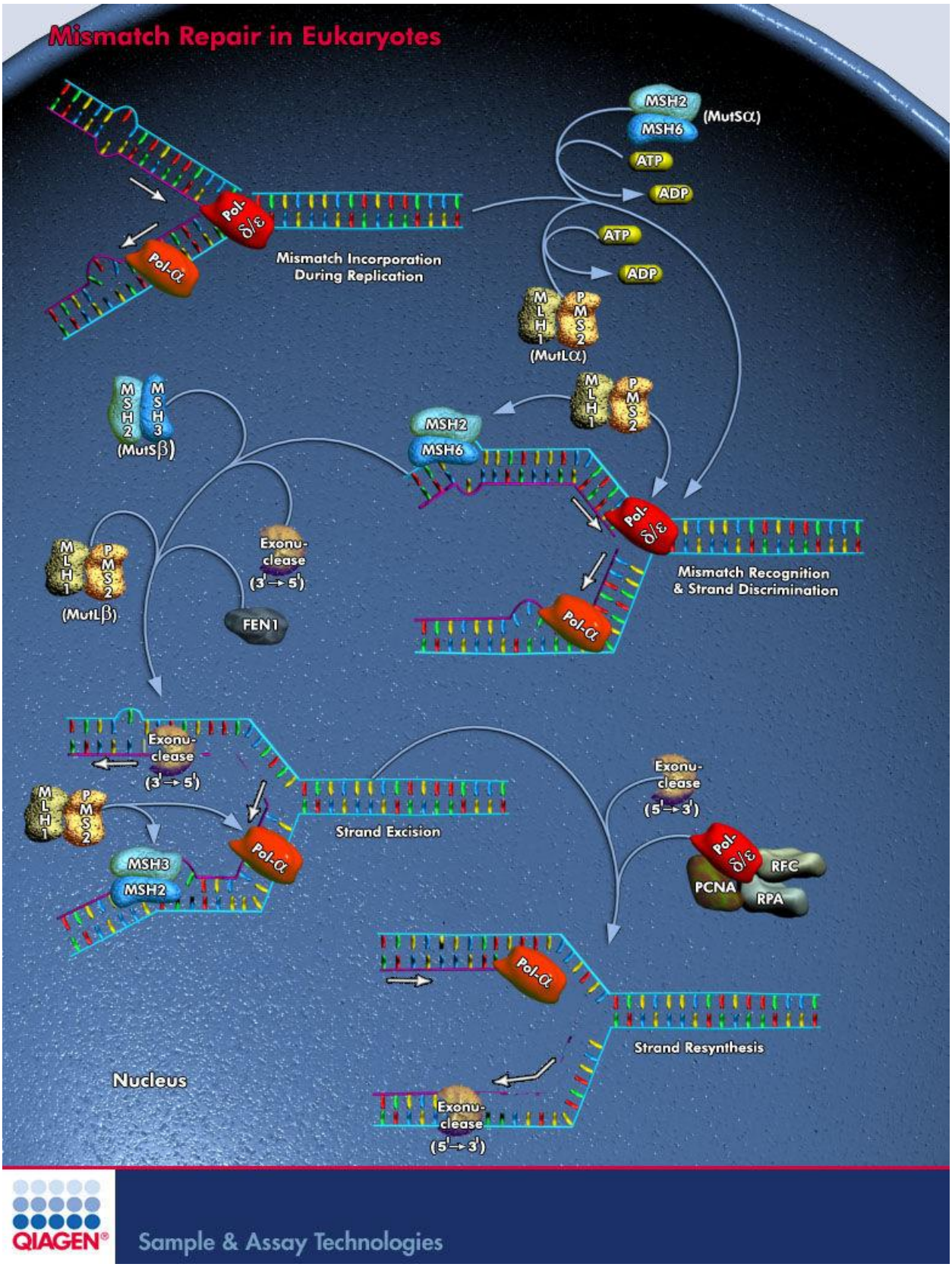
Eukaryotes have proteins that are homologous to MutS and MutL but lack homologs to MutH. There are three human MutS homologs designated as MSH2, MSH3, and MSH6 which participate in mismatch repair. MSH2 and MSH6 combine to form a hetrodimer called MutSα, and MSH2 and MSH3 combine to form a second heterodimer called MutSβ. The structures of MutSα and MutSβ are thought to be similar to the MutS homodimer in bacteria.
MutSα initiates mismatch repair at single mismatches and small insertions/deletion loops, whereas MutSβ only initiates mismatch repair at insertion/deletion loops of various sizes. Mammalian homologs of the bacterial MutL protein that participate in mismatch repair are designated as MLH1 and PMS2 and the heterodimer containing these two subunits is called MutLα. Paul Modrich and coworkers have reconstituted the human mismatch repair system *in vitro*.

A strand break on either side of the mismatch is sufficient to direct repair.

91. **Human Mismatch Repair System**

Paul Modrich and coworkers have reconstituted the human mismatch repair system *in vitro*. A strand break on either side of the mismatch is sufficient to direct repair. Purified human MutSα, MutLα, ExoI (a 5’ to 3’ exonuclease), RPA, PCNA, RFC, and DNA ploymerase□ are required for bidirectional repair. MutSα, ExoI, and RPA are adequate to excise a mismatch when the nick is on the 5’ side of the mismatch but MutLα, PCNA, and RFC are also required when the nick is on the 3’ side of the mismatch. The observation that the mismatch repair system can degrade newly synthesized strands with a nick on the 3’-side of the mismatch was very puzzling because ExoI degrades DNA in a 5’ to 3’ direction. Modrich and co-workers solved the puzzle by demonstrating that MutLα is a latent endonuclease that is activated in a mismatch. Once activated, MutLα preferentially makes incisions in the strand that already has a nick, that is, the discontinuous strand during replication. The endonuclease activity appears to require an amino acid motif present in the PMS2 but not the MLH1 subunit. In the human mismatch repair pathway, MutSα, PCNA, and RFC cooperate to activate the latent MutLα endonuclease. This MutLα endonuclease then nicks the discontinuous strand of a DNA duplex on both the 5’ and 3’ sides of the mismatch. When the original nick is on the 3’ side of the mismatch, MutLα incisions produce a new 5’ terminus on the far side of the mismatch that can serve as an entry site for MutSα-activated ExoI. This activated ExoI then removes the mismatch using its 5’ to 3’ exonuclease activity. RPA stimulates ExoI activity in the presence of MutSα as long as the mismatch is present. Once the mismatch has been removed, RPA inhibits the exonuclease, probably by displacing ExoI from the DNA.


5′-directed
PCNA
MutSα
5′
5′
3′
3′
5′
RFC
MutLα
ATP
ADP + Pi
5′
3′
5′
3′
3′
5′
5′ to 3′ hydrolysis
MutSα, MutLα,
ExoI, RPA
ATP
ADP + Pi
5′
3′
3′
5′
RPA
3′-directed
MutSα
PCNA
5′
3′
3′
5′
MutLα
RFC
ATP
ADP + Pi
5′
3′
5′
3′
3′
5′
5′ to 3′ hydrolysis
MutSα, MutLα,
ExoI, RPA
ATP
ADP + Pi
5′
3′
3′
5′


Mismatch Repair in Eukaryotes
MSH2
MSH6
(MutSα)
ATP
ADP
Pol-δ/ε
Pol-α
Mismatch Incorporation During Replication
ATP
ADP
MLH1
PMS2
(MutLα)
MSH2
MSH3
(MutSβ)
MSH2
MSH6
MLH1
PMS2
(MutLβ)
Exonu-clease (3'→5')
FEN1
Pol-δ/ε
Pol-α
Mismatch Recognition & Strand Discrimination
Exonu-clease (3'→5')
MLH1
PMS2
MSH3
MSH2
Pol-α
Strand Excision
Exonu-clease (5'→3')
Pol-δ/ε
RFC
PCNA
RPA
Pol-α
Strand Resynthesis
Nucleus
Exonu-clease (5'→3')
QIAGEN®
Sample & Assay Technologies

91. Human Mismatch Repair System

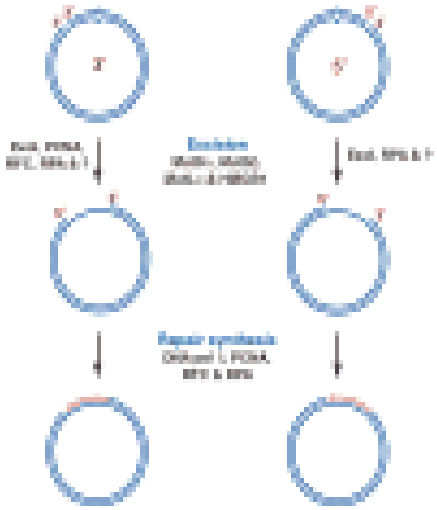
Inactivation of the human mismatch repair system confers a large increase in spontaneous mutability and a strong predisposition to tumor development. Mismatch repair provides several genetic stabilization functions: it corrects DNA biosynthetic errors, ensures the fidelity of genetic recombination, and participates in the earliest steps of checkpoint and apoptotic responses to several classes of DNA damage. Defects in this pathway are the cause of typical and atypical hereditary nonpolyposis colon cancer, but may also play a role in the development of 15 to 25% of sporadic tumors that occur in a number of tissues. The system is also of biomedical interest because mismatch repair-deficient tumor cells are resistant to certain cytotoxic chemotherapeutic drugs, a manifestation of its involvement in the DNA damage response. Of the several mutation avoidance functions of mismatch repair, the reaction responsible for replication error correction has been the most thoroughly studied, and the discussion that follows is restricted to this pathway.

Mismatch Repair in Eukaryotic Cell Extracts

Correction of DNA biosynthetic errors requires targeting of mismatch repair to the newly synthesized strand at the replication fork. In contrast to *E. coli*, where mismatch repair is directed by the transient absence of adenine methylation at d(GATC) sites within newly synthesized DNA, the strand signals that direct replication error correction in eukaryotes have not been identified. However, the function of the hemimethylated d(GATC) strand signal in *E. coli* mismatch repair is provision of a nick on the unmethylated strand, which serves as the actual signal that directs the reaction . Similarly, a strand-specific nick or gap is sufficient to direct mismatch repair in extracts of mammalian and *Drosophila* cells, as well as *Xenopus* egg extracts. These findings, coupled with the observation that mismatch repair is more efficient on the lagging strand at the replication fork, suggest that DNA termini that occur as natural intermediates during replication (3'-terminus on the leading strand; 3' and 5' termini on the lagging strand) may suffice as strand signals to direct the correction of DNA biosynthetic errors in eukaryotic cells.

Available information on the mechanism of eukaryotic mismatch repair is largely derived from analysis of the nick-directed repair of circular heteroduplexes in mammalian cell extracts. The strand break that directs repair may reside either 3' or 5' to the mispair as viewed along the shorter path linking the two sites in the circular substrate, and processing of such molecules in extracts is largely restricted to this region. Examination of intermediates produced in HeLa nuclear extracts when repair DNA synthesis is blocked has demonstrated that mismatch-provoked excision removes that portion of the incised strand spanning the shorter path between the nick and the mismatch, with excision tracts extending from the strand break to a number of sites within a region ≈ 90 to 170 nucleotides beyond the mispair. Radiolabeling of repair DNA synthesis tracts is also consistent with this view. The mammalian repair system thus displays a bidirectional capability in the sense that it responds to both 3'- and 5'-heteroduplex orientations,

and functionality is retained at nick-mismatch separation distances as large as 1,000 base pairs (bp).



Substrates and requirements for *in vitro* mismatch repair

Mammalian MutS and MutL Activities

The activities responsible for initiation of *E. coli* mismatch repair are MutS and MutL, which function as homo-oligomers. MutS is responsible for mismatch recognition and MutL serves to interface mismatch recognition by MutS to activation of downstream activities. Mammalian cells possess two MutS activities that function as heterodimers and share MSH2 as a common subunit : MutSα (MSH2•MSH6 heterodimer) and MutSβ (MSH2•MSH3 heterodimer). MutSα, which represents 80 - 90% of the cellular MSH2, preferentially recognizes base-base mismatches and insertion/deletion (ID) mispairs in which one strand contains 1 or 2 unpaired nucleotides, but is also capable of recognition of larger ID heterologies with reduced affinity. MutSβ recognizes ID mismatches of 2 to about 10 nucleotides, weakly recognizes single-nucleotide ID mispairs, and is essentially inert on base-base mismatches. *MSH2*and *MSH6* defects have been implicated in tumor development, but the cancer predisposition conferred by*MSH6* inactivation is less severe. The association of *MSH3* defects with tumor development appears to be limited.

Three eukaryotic MutL activities have been identified, and like eukaryotic MutS activities function as heterodimeric complexes, with MLH1 serving as a common subunit. MutLα, a heterodimer of MLH1 and PMS2, is the primary MutL activity in human mitotic cells and supports repair initiated by either MutSα or MutSβ. MutLα accounts for ≈ 90% of the MLH1 in human cells, but two low abundance complexes involving MLH1 have also been identified. A human MLH1•PMS1 heterodimer (MutLβ) has been isolated, but involvement in mismatch

repair has not been demonstrated. However, the MutLγ MLH1•MLH3 complex has been reported to support modest levels of base-base and single nucleotide ID mismatch repair *in vitro*, events that are presumably initiated by MutSα. Genetic inactivation of *MLH1*or *PMS2* confers cancer predisposition, but mutations in *PMS1* do not. Involvement of *MLH3* defects in tumor development is uncertain.