Mu'tah University
Faculty of Medicine
Department of Pharmacology & Biochemistry

Course title:	Molecular biology and genetics
Course code:	1503102
Credit hours:	2 credit hours
Calendar description:	17 weeks / Second semester / 1st year Medicine

Course description: This two-credit hours course is mandatory for first-year medical students. It consists of one lecture (1.5 hours) and a laboratory session a week. The course is designed to introduce students to the basics of molecular biology and genetics via starting with the chemical structures of the genetic material, cellular regulation of gene expression, the concept and types of inheritance, mutations, and their relationship to health and disease.

The laboratory is intended to facilitate students' understanding of theoretical concepts of both molecular biology and biochemistry.

Objectives (intended learning outcomes):
The overall objective is to enhance student understanding of advanced molecular biology and genetics-based medical topics to be covered in later courses.

1. **Knowledge:**
   - **A. Lectures**
   - Understand the involvement of molecular biology in medicine
   - Understand the chemical structures of DNA and chromosomes
   - Understand the mechanism and regulation of DNA replication
   - Understand the mechanism and regulation of gene expression (transcription and translation)
   - Understand the types and causes of DNA mutations
   - Understand the mechanisms of DNA repair
   - Understand the mechanism of inheritance
   - Understand the mechanism of disease inheritance
   - Understand the mechanisms of signal transduction pathways and their regulation of gene expression
   - Keep up with advancement in the field of molecular biology

   - **B. Laboratory**
   - Be acquainted of laboratory safety, data collection and reporting
   - Be proficient in liquid handling and pipetting
   - Determination of DNA measurement
   - Extraction and measurement of plasmid DNA
   - Cleavage of DNA by restriction endonucleases
   - Performance of DNA fingerprinting
   - Performance of DNA electrophoresis
   - Performance of Polymerase chain reaction
   - Understand and apply bioinformatics
2. **Skills:**
   A. Understand the different functions of genes, their regulation and inheritance, and genetic association to health and disease
   B. Be able to read and comprehend molecular biology and genetics textbooks

List of lectures

- **Introduction**
  Significance of molecular biology and genetics in biology and medicine
  Living systems and experimental models
  Chromosomes, genes, and DNA
- **DNA and RNA structures**
  General characteristics of DNA and RNA structures
  Chemical structure of DNA
  DNA-protein interaction and binding
- **Structure of chromosomes**
  Chromosomal structure and chromatin in prokaryotes and eukaryotes
  Chromosomal karyotyping and Chromosomal Disorders
- **DNA replication in prokaryotes and eukaryotes**
- **RNA and gene transcription**
  Types and structure of RNA in prokaryote and eukaryotes
  Transcription in prokaryote and eukaryotes
  Post-transcriptional regulation
- **Translation in prokaryote and eukaryotes**
  - Post-translational regulation, Eukaryotic Pre-mRNA Processing, RNA Splicing
- **Genetic Testing**
  Newborn screening, Presymptomatic testing, Prenatal testing
- **Cancer**
  Causes, cancer genetics, p53, treatment
- **Stem cell: Types and properties**

- Gene therapy: Introduction, uses and risk

- Signal transduction: General principles and examples
Examples of signaling pathways

- **Gene Regulation (Operon system)**
- **DNA mutations and chromosomal anomalies**
  Types of chromosomal anomalies
  Examples of genetic diseases
  Types of DNA mutation
- **Mechanisms of DNA repair**

List of laboratories

1- Introduction.
2- Extraction of genomic DNA from bacteria
3- Concept of restriction endonucleases and DNA cloning introduction
4- Restriction endonucleases Lab
5- Gel electrophoresis
6- PCR introduction.
7- PCR Lab
8- Extraction and measurement of plasmid DNA

Please note that some labs need more than one class to complete.

Introduction to Molecular Biology

Molecular biology is the study of biological molecules and the molecular basis of structure and function in living organisms. The goal of molecular biology is to understand the five basic cell behaviour patterns (growth, division, specialization, movement, and interaction) in terms of various molecules that are responsible for them. That is molecular biology wants to generate a complete description of the structure, function, and interrelationship of the cells macromolecules and thereby to understand why living cells behave the way they do.

Following the rapid advances in biological science brought about by the development and advancement of the Watson-Crick model of DNA (deoxyribonucleic acid) during the 1950s and 1960s, molecular biologists studied gene structure and function in increasing detail. In addition to advances in understanding genetic machinery and its regulation, molecular biologists continue to make fundamental and powerful discoveries regarding the structure and function of cells and of the mechanisms of genetic transmission. The continued study of these processes by molecular biologists and the advancement of molecular biological techniques require integration of knowledge derived from physics, microbiology, mathematics, genetics, biochemistry, cell biology and other scientific fields.

The complete set of genes containing the genetic instructions for making an organism (cell and creature) is called its genome. It contains the master blueprint (DNA) for all cellular structures and activities for the lifetime of the cell or organism. The human genome consists of tightly coiled threads of deoxyribonucleic acid (DNA) and associated protein molecules organized into structures called chromosomes. In humans, as in other higher organisms, a DNA molecule consists of two strands that wrap around each other to resemble a twisted ladder whose sides (backbone) made of sugar and phosphate molecules are connected to nitrogen-containing chemicals called nitrogenous bases. Each strand is a linear arrangement of repeating units called nucleotides. The particular order of the bases arranged along the sugar-phosphate backbone is called the DNA sequence; the sequence specifies the exact genetic instructions required to create a particular organism with its own unique traits.

Each time a cell divides into two daughter cells, its full genome is duplicated; for humans and other complex organisms, this duplication occurs in the nucleus. Each daughter cell receives one old and one new DNA strand. The cell's adherence to these base-pairing rules ensures that the new strand is an exact copy of the old one. This minimizes the incidence of errors (mutations) that may greatly affect the resulting organism or its offspring.

Each DNA molecule contains many genes, the basic physical and functional units of heredity. A gene is a specific sequence of DNA nucleotide bases, whose sequences carry the information required for constructing a specific polypeptide, which provide the structural components of cells and as well as enzymes for essential biochemical reactions.

The chromosomes of prokaryotic microorganisms differ from eukaryotic microorganisms, in terms of shape and organization of genes. Prokaryotic genes are more closely packed and are usually is arranged along one circular chromosome.

The central belief of molecular biology states that DNA is copied to make mRNA (messenger RNA), and mRNA is used as the template to make proteins. Formation of mRNA is called transcription and formation of protein is called translation. Transcription and translation processes are regulated at various stages and the regulation steps are unique to prokaryotes and eukaryotes. DNA regulation determines what type and amount of mRNA should be transcribed, and this subsequently determines the type and amount of protein. This process is the fundamental control mechanism for growth and development.

All living organisms are composed largely of proteins, the end product of genes. Proteins are large, complex molecules made up of long chains of subunits called amino acids. The protein-coding instructions from the genes are transmitted indirectly through messenger ribonucleic acid (mRNA), a transient intermediary molecule similar to a single strand of DNA. For the information within a

gene to be expressed, a complementary RNA strand is produced (a process called transcription) from the DNA template. In eukaryotes, messenger RNA (mRNA) moves from the nucleus to the cellular cytoplasm, but in both eukaryotes and prokaryotes mRNA serves as the template for protein synthesis.

Twenty different kinds of amino acids are usually found in proteins. Within the gene, sequences of three DNA bases serve as the template for the construction of mRNA with sequence complimentary codons that serve as the language to direct the cell's protein-synthesizing machinery. Codons specify the insertion of specific amino acids during the synthesis of protein. For example, the base sequence ATG codes for the amino acid methionine. Because more than one codon sequence can specify the same amino acid, the genetic code is termed a degenerate code (i.e., there is not a unique codon sequence for every amino acid).

Areas of intense study by molecular biology include the processes of DNA replication, repair, and mutation. Other areas of study include the identification of agents that cause mutations (e.g., ultra-violet rays and some chemicals) and the mechanisms of rearrangement and exchange of genetic materials.

Recombinant DNA technologies and genetic engineering are an increasingly important part of molecular biology. Advances in biotechnology and molecular medicine also carry profound clinical and social significance. Advances in molecular biology have led to significant discoveries concerning the mechanisms of the embryonic development, disease and immunologic response.

Model of biological system

Viruses: are the simplest organisms. It is made up of DNA (in some cases RNA) surrounded by protein coat. The key to the virus simplicity is its parasitic nature. It borrows functions from its host cells. The host include bacteria cells (bacteriophage is bacteria viruses), plant cells, and animal cells. Viruses helped in proving that DNA and not proteins contain the genetic information.

Bacteria: a unicellular cells that have a single chromosome, and they are simple in their organization. Bacteria lack a membrane-bounded nucleus (they are prokaryotes) and mitochondria, are surrounded by a cell wall, and divide by binary fission. Bacteria are suitable object to study some process because they can grow easily and rapidly, they are simple in their need, best example is Escherichia coli (E. coli) which divide every 20 minutes at optimal conditions. They can be grown in liquid (broth) and solid surface (agar). If the liquid is complex extract of biological material it is called broth, if the growth medium is a simple mixture containing no organic compounds other than a carbon source such as sugar it is called a minimal medium. A typical minimal medium contain the ions and a source of carbon such as glucose, glycerol, or lactate. If a bacterium can grow in a minimal medium it is called prototroph and if other substance has to be added for growth the bacterium is called auxotroph.

Bacteria is normally grown on agar, a jellying agent obtained from seaweed. Agar is resistant to the action of bacterial enzyme.

Metabolism in bacteria is precisely regulated. They rarely synthesis substances that are not needed. For example if tryptophan is present in the growth medium bacteria will not make it but once used up the tryptophan synthesizing enzyme will be activated.

Yeast: they are eukaryotes; mutant strains of yeast were often used to discover the genes that control growth, division, and cell behaviour patterns. They also are used to produce large number of human genes.

Animal cells (and embryos) many animal cells can be grown and cultured in lab. Primary cell culture represent normal animal tissue. Animal cells grow well initially but eventually die off. Tumour cells grow indefinitely and are easier to propagate in culture. Embryonal stem cells holds a great promise as starting point for the production of tissues and organs as human replacement parts. In addition molecular biologist has succeeded in injecting foreign genes into animal eggs and thus

generating transgenic animals. In some animal cases improved agriculture productivity such as enhanced milk production from dairy cows. Attempts will be made to inject foreign genes into humans cells to correct genetic defect (gene therapy).

Plant cells the discovery that cells of many plants can be raised and cultivated into whole complete plants expanded the horizons of molecular biologist to learn about gene function in plants.

Genetics and Medicine

Most disease processes can be viewed as resulting from environmental influences interacting with the individual genetic makeup of the affected individual. A disease is genetically determined if it is mainly or exclusively caused by disorders in the genetic program of cells and tissues. More than 3000 defined human genetic diseases are known to be due to a mutation at a single gene locus (monogenic disease) and to follow a Mendelian mode of inheritance.

An important category of disease results from genetic predisposition interacting with environmental factors (multigenic or multifactorial diseases). This includes many relatively common chronic diseases (e.g., high blood pressure, hyperlipidemia, diabetes mellitus, gout, psychiatric disorders).

The identification of disease-related genes has led to an increase in the number of available genetic tests that detect disease or an individual's risk of disease. Genetic testing may be undergone prior to marriage, during pregnancy, or after birth.

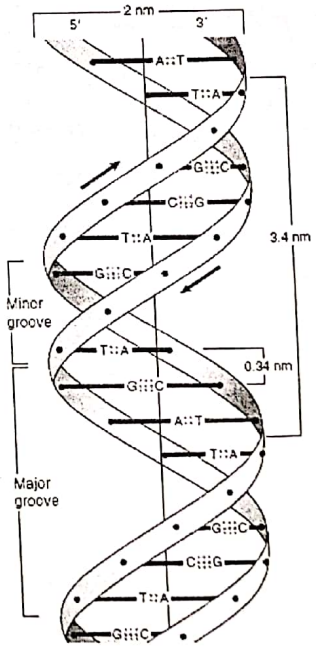
Genetic counselling is generally offered prior to marriage or conception, in order to predict the likelihood of conceiving an affected child, during pregnancy, in order to determine the condition of the fetus, or to an adult, in order to determine susceptibility to a certain disease.

Gene therapy is used to correct defective genes that cause disease. Gene therapy is not yet an active current therapy but is still in a stage of clinical investigation. However cloning has a high potential for curing many diseases but it is still under investigations.

Genetically determined diseases are not a marginal group, but make up a substantial proportion of diseases. More than one-third of all paediatric hospital admissions are for diseases and developmental disorders that, at least in part, are caused by genetic factors. The total estimated frequency of genetically determined diseases of different categories in the general population is about 3.5–5.0%.

Nucleic Acids

Nucleic acids occur in two forms, **Deoxyribonucleic acid** (DNA) and **ribonucleic acid** (RNA). Both are polymers of subunits termed **nucleotides**. DNA is found in the nucleus of eukaryotes and the cytoplasm of prokaryotes and functions as the molecule of heredity in all cells except some viruses where RNA is the molecule of heredity. RNA molecules are synthesized on DNA templates and participate in protein synthesis in the cytoplasm.



DNA & RNA

The first correct three-dimensional structure of the **deoxyribonucleic acid** DNA molecule was proposed in 1953 by James Watson and Francis Crick. The structure suggests how DNA duplicates itself, controls hereditary traits, and undergoes mutation. In the Watson-Crick structure, DNA consists of two long chains of nucleotides twisted around one another to form a double-stranded helix.

DNA contains all the information required to build the cells and tissues of an organism. The exact replication of this information in any species assures its genetic continuity from generation to

generation and is critical to the normal development of an individual. The information stored in DNA is arranged in hereditary units, known as **genes** that control identifiable traits of an organism. In the process of **transcription,** the information stored in DNA is copied into **ribonucleic acid (RNA),** which has distinct roles in protein synthesis. Chemically RNA differs from DNA in two respects: 1- it contains ribose sugar instead of deoxyribose, 2- it has uracil (U) instead of thymine (T). There are three classes of RNA based on their functions: (1) **transfer** RNAs (tRNAs); (2) **messenger** RNAs (mRNAs); and (3) **ribosomal RNAs** (rRNAs).

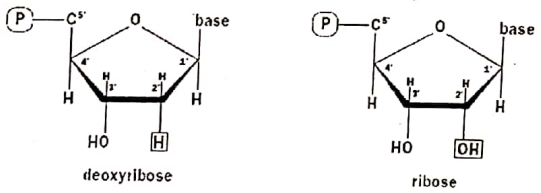
DNA exists predominantly as right-handed antiparallel double-stranded helix. The two strands are helically coiled, which maximizes the exposure of the negatively charged phosphate backbone to water and shields the **hydrophobic** bases in the middle from water.
DNA has a regular helix making a complete turn every $34A^0$ (3.4 nm) with a diameter of about 20 A^0 (2 nm).
The bases of DNA are flat and stacked above one another. Each base pair is rotated about 36^0 in B DNA around the axis of helix relative to the next base pair, so about 10 bases pairs make a complete 360^0. The chemical nature of the bases in double-stranded DNA creates a slight twisting force that gives DNA its characteristic gently coiled structure, known as the double helix with a minor groove about 12 A^0 across and a major groove about 22 A^0.

Nucleotides

DNA and RNA are long, unbranched polymer of nucleotides. A nucleotide is made up of one molecule of sugar (deoxyribose in DNA and ribose in RNA) bound to one molecule of phosphate and to one nitrogen-containing base. In DNA nitrogen bases are adenine (A), guanine (G), cytosine (C), and thymine (T), while in RNA adenine (A), guanine (G), cytosine (C), and uracil (U)).
Nucleotides are named after the nitrogen bases present, therefore, four different nucleotide bases are found in DNA: **A, G, C,** and **T** and four in RNA: **A, G, C,** and **U**.
The genetic information is stored in the sequence of bases, or base pairs that determine the sequence of amino acids in proteins.

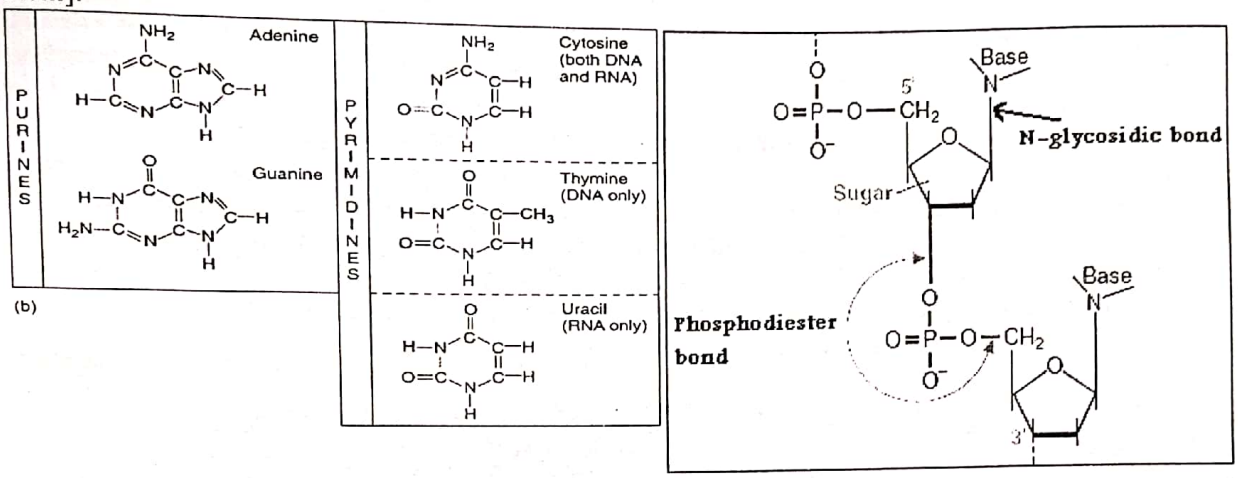
Components of nucleotides:

1- Deoxyribose Sugar: The carbon atoms on the sugar ring are numbered 1' (one prime) to 5' to distinguish them from atoms in the bases. The ribose in DNA differs from that of RNA by the absence of oxygen at the carbon atom number 2 and is thus 2-deoxy-β-D-ribose in DNA while in RNA β-D-ribose.



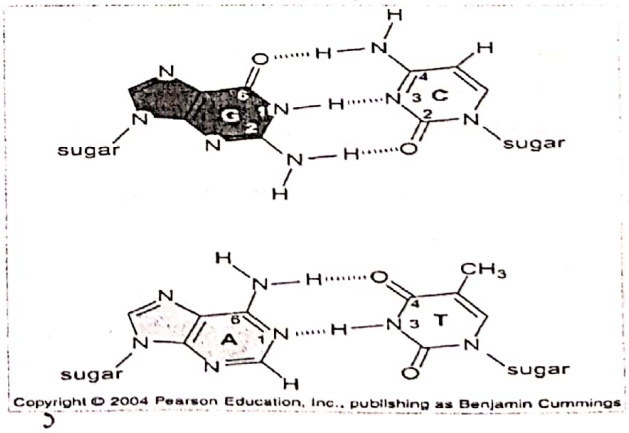
2- Nitrogen-containing bases: Because of their nitrogen content and basic qualities they are known as **nitrogenous bases.** The organic bases are of two general types: **purines** and **pyrimidines**. Each purine consists of a six-sided ring attached to a five-sided ring, whereas each pyrimidine consists of a six-sided ring only. The purines are **adenine** (A) and **guanine** (G). The pyrimidines are **cytosine** (C), **thymine** (T), and **uracil** (U). Thymine is found primarily in DNA and uracil is found only in RNA. Uracil differs from thymine by lacking a methyl group on its C 5 ring. All of these bases are hydrophobic. Purines and pyrimidines have basic amine groups and therefore can be described as organic bases.

In a nucleotide, the nitrogen atom in position 9 of a purine or in position 1 of a pyrimidine is bound to the carbon number 1 of the sugar by N-glycosidic bond [a covalent bond that joins a carbohydrate molecule to another group which may or may not be a carbohydrate through nitrogen atom].



3- Phosphate group: A phosphoric acid (H_3PO_4) is bounded to one oxygen by a double bond and three hydroxyl groups (–OH). Two of the hydroxyl groups can form covalent bonds, phosphodiester bonds, with the sugar hydroxyl groups by splitting out water [phosphodiester bond is a covalent bond between a phosphate group and two ribose or deoxyribose sugar]. The third -OH group on the phosphate is free and dissociates a hydrogen ion (H^+ ions) and leaving negatively charged oxygens at physiologic pH. In this form, the structure is referred to as phosphate. Therefore, phosphate is a negatively charged which gives the polymer its acidic property and promoting their attraction to positively charged histone proteins that partially neutralized this negative charges and because of the phosphate charges present in their component nucleotides, both DNA and RNA are negatively charged.

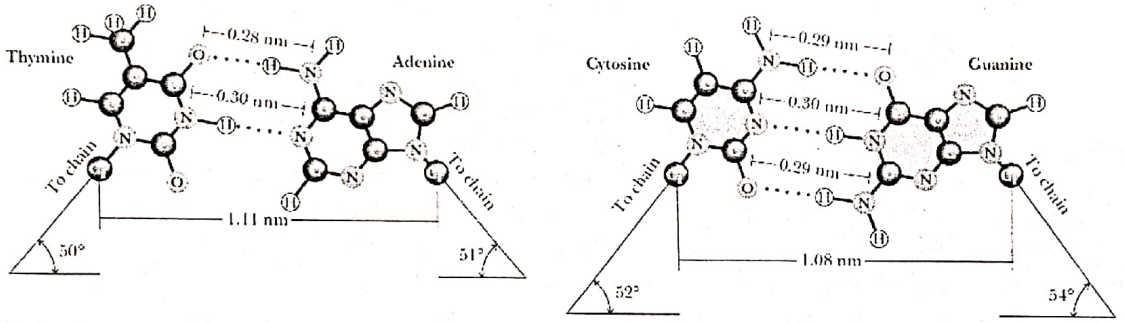
Electrostatic repulsion by negatively charged phosphates along the DNA backbone destabilize the double helix. For example, if the phosphates are left unshielded, as when DNA is dissolved in distilled water, the DNA strands will separate at room temperature. Neutralizing these negative charges by the addition of NaCl (which contributes positively charged sodium ions) to the DNA solution will prevent strand separation. In the cell, the phosphates also interact with positively charged (magnesium, potassium, or sodium) ions and with positively charged histone proteins.
As stated above, in a nucleotide, the nitrogen atom in position 9 of a purine or in position 1 of a pyrimidine is bound by N-glycosidic bond to the carbon number 1 of the sugar. On the other hand, the phosphate group links the 3' end of one nucleotide to the 5' end of the next nucleotide through phosphodiester bonds. This gives the sugar-phosphate backbone directionality a 5' end and a 3' end. In DNA the direction of the nucleotides in one strand is opposite to their direction in the other strand and thus called antiparallel. The 5' end has a terminal phosphate group and the 3' end has a terminal hydroxyl group.
This directionality, plus the fact that synthesis proceeds 5' to 3', has given rise to the convention that polynucleotide sequences are written and read in the 5' to 3' direction (from left to right); for example, the sequence AUG is assumed to be (5')AUG(3'). As we will see, the 5'-3' directionality of a nucleic acid strand is an important property of the molecule.



Base Pairs and bonds in DNA

Hydrogen bonding occurs between laterally opposed bases of the two strands of the DNA duplex according to Watson-Crick rules: adenine (A) specifically binds to thymine (T) by two hydrogen bonds (A=T) and cytosine (C) specifically binds to guanine (G) by three hydrogen bonds (C≡G). As a result, the base composition of DNA from different cellular sources is not random: the amount of adenine equals that of thymine, and the amount of cytosine equals that of guanine and this is known as Chargaff's rule.

The A-T and G-C bonds are the most stable of the various base-pairs on energetic grounds. In addition, these two base pairs have essentially identical dimensions within the double helix structure of DNA. That is the distance between the C1' carbons of the deoxyriboses of each base pair is 1.1 nm in each case (1.11 nm for the AT pair and 1.08 nm for GC) and thus avoiding partnership of purine-purine which will be too wide or pyrimidine-pyrimidine which is too narrow. This means that these are also the only two base pairs which will properly fit into the double helix that is the outside diameter of the double helix will be uniform over its length.



The backbone-to-backbone distance of an A:T pair is 1.11 nm, virtually identical to the 1.08 nm chain separation in G:C base pairs.

A hydrogen bond is a weak attraction between a hydrogen atom covalently bonded to one electronegative atom (either nitrogen or oxygen in DNA) and another electronegative atom. Hydrogen atoms of amino groups serve as the hydrogen bond donor while the carbonyl oxygens and ring nitrogens serve as hydrogen bond acceptors. The specific location of hydrogen bond donor and acceptor groups gives the bases their specificity for hydrogen bonding in unique pairs. As hydrogen bonds are weak they can be broken and rejoined relatively easily. Unlike covalent bonds which require considerable energy input to break them, noncovalent bonds like hydrogen bond are constantly being made and broken at physiological temperatures. As a result, they readily permit reversible and so transient molecular interactions, which are essential for biological function example ensuring faithful replication of DNA, transcription of RNA, codon-anticodon recognition, etc. Although individually weak, the combined action of numerous hydrogen bond bonds can make large contributions to the stability of the structure (conformation) of these molecules and so can be crucially important for specifying the shape of a macromolecule for example the intramolecular hydrogen bonding provides much of the shape of a transfer RNA molecule.

The important functional groups participating in H-bond formation are the amino groups of cytosine, adenine, and guanine; the ring nitrogens at position 3 of pyrimidines and position 1 of purines; and the strongly electronegative oxygen atoms attached at position 4 of uracil and thymine, position 2 of cytosine, and position 6 of guanine.

The paired nitrogenous bases tend to stack on top of one another (stacking interaction) in such a way as to exclude the maximum amount of water from the interior of the double helix. Hence a double stranded DNA molecule has a hydrophobic core composed of stacked bases, and it is the base stacking that provides double-stranded DNA with much of its chemical stability.

DNA can adopt different types of helical structure including, A DNA, B DNA, and Z DNA.

Both A-DNA and B-DNA are right-handed helices (ones in which the helix spirals in a clockwise direction as it moves away from the observer). They have respectively 11 and 10 base pairs per turn.

On the other hand, Z-DNA is a left-handed helix which has 12 base pairs per turn. Under physiological conditions, most of the DNA in a bacterial or eukaryotic genome is of the B-DNA.

The DNA Double Helix Is a Stable Structure

Several factors account for the stability of the double-helical structure of DNA. First, both internal and external hydrogen bonds stabilize the double helix. The two strands of DNA are held together by H-bonds that form between the complementary purines and pyrimidines, while polar atoms in the sugar–phosphate backbone form external H bonds with surrounding water molecules. Second, the negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and are free to interact electrostatically with cations in solution such as Mg^{2+}. Third, the core of the helix consists of the base pairs, which they stack together through hydrophobic interactions and van der Waals forces that contribute significantly to the overall stabilizing energy.

Significance of chemical differences between DNA and RNA

Two fundamental chemical differences distinguish DNA from RNA:
1. DNA contains 2-deoxyribose sugar instead of ribose sugar.
2. DNA contains thymine instead of uracil.

What are the consequences of these differences and do they hold any significance in common? An argument can be made that, because of these differences, DNA is a more stable polymeric form than RNA. The greater stability of DNA over RNA is consistent with the respective roles these macromolecules have assumed in heredity and information transfer.

Consider first why DNA contains thymine instead of uracil. The key observation is that cytosine deaminates (removal of amino group from biomolecule) to form uracil at a fixed rate in vivo (see figure below).

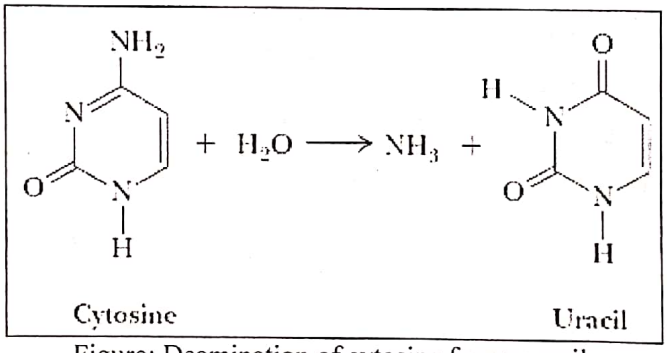


Figure: Deamination of cytosine forms uracil.

Because C in one DNA strand pairs with G in the other strand, whereas U would pair with A, conversion of a C to a U could potentially result in a heritable change of a CG pair to a UA pair. Such a change in nucleotide sequence would constitute a mutation in the DNA. To prevent this reaction from leading to changes in nucleotide sequence, a cellular repair mechanism "proofreads" DNA, and when a U arising from C deamination is encountered, it is treated as inappropriate and is replaced by a C. If DNA normally contained U rather than T, this repair system could not readily distinguish U formed by C deamination from U correctly paired with A.

The absence of the 2'-hydroxyl group in DNA further increases its resistance to hydrolysis. The ribose 2'-OH group of RNA is absent in DNA. This difference leads to a greater resistance of DNA to alkaline hydrolysis. To view it another way, RNA is less stable than DNA because its 2'-OH group makes the 3'-phosphodiester bond susceptible to nucleophilic cleavage.

On the other hand, Z-DNA is a left-handed helix which has 12 base pairs per turn. Under physiological conditions, most of the DNA in a bacterial or eukaryotic genome is of the B-DNA.

The DNA Double Helix Is a Stable Structure

Several factors account for the stability of the double-helical structure of DNA. First, both internal and external hydrogen bonds stabilize the double helix. The two strands of DNA are held together by H-bonds that form between the complementary purines and pyrimidines, while polar atoms in the sugar–phosphate backbone form external H bonds with surrounding water molecules. Second, the negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and are free to interact electrostatically with cations in solution such as Mg^{2+}. Third, the core of the helix consists of the base pairs, which they stack together through hydrophobic interactions and van der Waals forces that contribute significantly to the overall stabilizing energy.

Significance of chemical differences between DNA and RNA

Two fundamental chemical differences distinguish DNA from RNA:
1. DNA contains 2-deoxyribose sugar instead of ribose sugar.
2. DNA contains thymine instead of uracil.

What are the consequences of these differences and do they hold any significance in common? An argument can be made that, because of these differences, DNA is a more stable polymeric form than RNA. The greater stability of DNA over RNA is consistent with the respective roles these macromolecules have assumed in heredity and information transfer.

Consider first why DNA contains thymine instead of uracil. The key observation is that cytosine deaminates (removal of amino group from biomolecule) to form uracil at a fixed rate in vivo (see figure below).

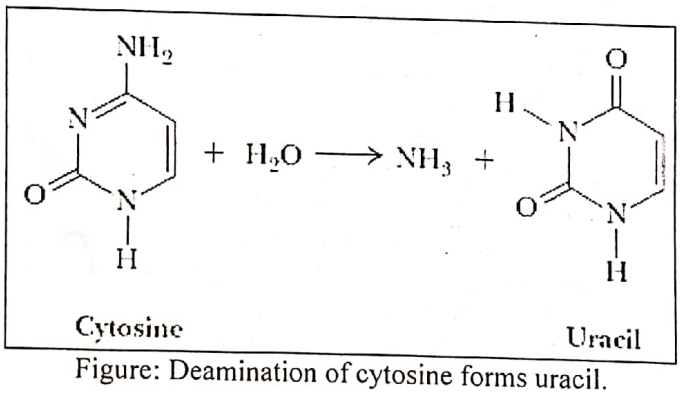


Figure: Deamination of cytosine forms uracil.

Because C in one DNA strand pairs with G in the other strand, whereas U would pair with A, conversion of a C to a U could potentially result in a heritable change of a CG pair to a UA pair. Such a change in nucleotide sequence would constitute a mutation in the DNA. To prevent this reaction from leading to changes in nucleotide sequence, a cellular repair mechanism "proofreads" DNA, and when a U arising from C deamination is encountered, it is treated as inappropriate and is replaced by a C. If DNA normally contained U rather than T, this repair system could not readily distinguish U formed by C deamination from U correctly paired with A.

The absence of the 2'-hydroxyl group in DNA further increases its resistance to hydrolysis. The ribose 2'-OH group of RNA is absent in DNA. This difference leads to a greater resistance of DNA to alkaline hydrolysis. To view it another way, RNA is less stable than DNA because its 2'-OH group makes the 3'-phosphodiester bond susceptible to nucleophilic cleavage.

Grooves (space between strands)

Why there is a grooves in DNA? Because the angle at which the two sugars protrude from the base pairs that is the angle between the glycosidic bonds is not equal, it is about 120° for the narrow angle (form the minor groove) and 240° for the wide angle (form the major groove). The major groove is 22 Å wide and the minor groove is 12 Å wide.

The edges of each base pair are exposed in the major and minor grooves, creating a pattern of hydrogen bond donors and acceptors and of van der Waals surfaces that identifies the base pair.

The edge of an A:T base pair displays the following chemical groups in the following order in the major groove: a hydrogen bond acceptor (the N7 of adenine), a hydrogen bond donor (the exocyclic amino group on C6 of adenine), a hydrogen bond acceptor (the carbonyl group on C4 of thymine) and a bulky hydrophobic surface (the methyl group on C5 of thymine).

Similarly, the edge of a G:C base pair displays the following groups in the major groove: a hydrogen bond acceptor (at N7 of guanine), a hydrogen bond acceptor (the carbonyl on C6 of guanine), a hydrogen bond donor (the exocyclic amino group on C4 of cytosine), a small non-polar hydrogen (the hydrogen at C5 of cytosine).

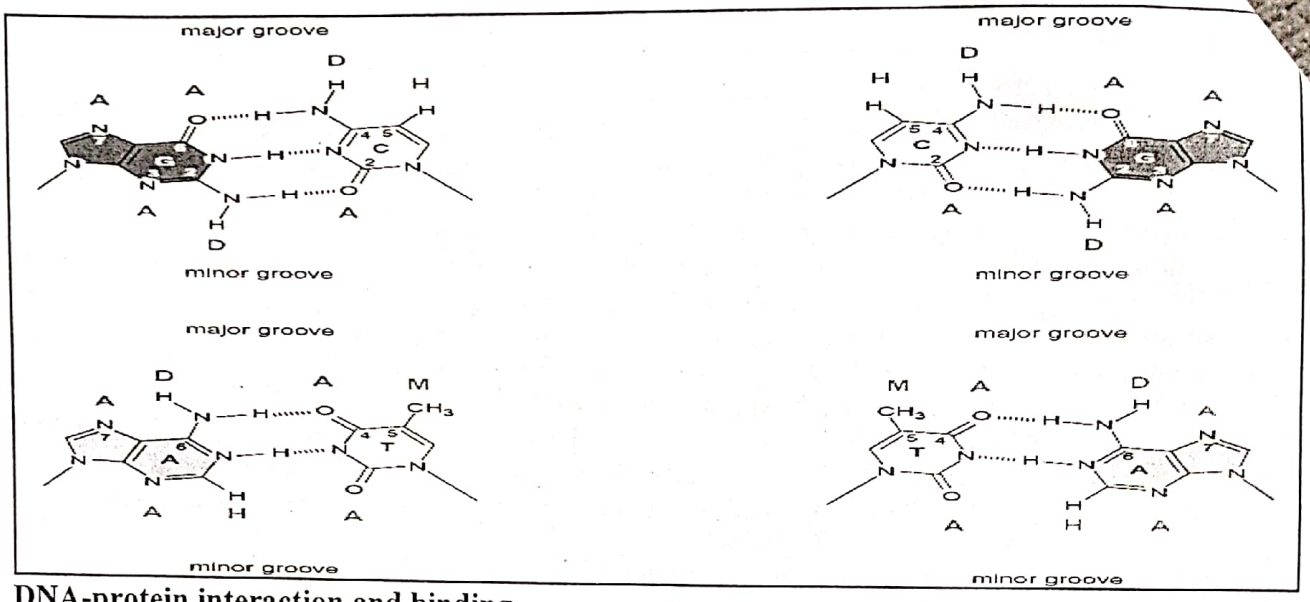
Thus, there are characteristic patterns of hydrogen bonding and of overall shape that are exposed in the major groove that distinguish an A:T base pair from a G:C base pair, and A:T from T:A, and G:C from C:G.

We can think of these features as a code in which **A** represents a **hydrogen bond acceptor, D** a **hydrogen bond donor, M** a **methyl group,** and **H** a **nonpolar hydrogen** (please see slides for clearer figure). In such a code, **A D A M** in the major groove signifies an A:T base pair, and **A A D H** stands for a G:C base pair. Likewise, **M A D A** stands for a T:A base pair and **H D A A** is characteristic of a C:G base pair.

In all cases, this code of chemical groups in the major groove specifies the identity of the base pair. These patterns are important because they allow proteins to easily recognize DNA sequences without having to open and thereby disrupt the double helix.

The minor groove is not as rich in chemical information and what information is available is less useful for distinguishing between base pairs. The small size of the minor groove is less able to accommodate amino acid side chains. Also, A:T and T:A base pairs and G:C and C:G pairs look similar to one another in the minor groove. An A:T base pair has a hydrogen bond acceptor (at N3 of adenine), a nonpolar hydrogen (at N2 of adenine) and a hydrogen bond acceptor (the carbonyl on C2 of thymine). Thus, its code is **A H A**. But this code is the same if read in the opposite direction, and hence an A:T base pair does not look very different from a T:A base pair from the point of view of the hydrogen bonding properties of a protein poking its side chains into the minor groove.

Likewise, a G:C base pair exhibits a hydrogen bond acceptor (at N3 of guanine), a hydrogen bond donor (the exocyclic amino group on C2 of guanine), and a hydrogen bond acceptor (the carbonyl on C2 of cytosine), representing the code **A D A**. Thus, from the point of view of hydrogen bonding, C:G and G:C base pairs do not look very different from each other either. The minor groove does look different when comparing an A:T base pair with a G:C base pair, but G:C and C:G, or A:T and T:A, cannot be easily distinguished (see Figure).



DNA-protein interaction and binding

Gene regulatory proteins that bind directly to DNA (regulatory DNA binding proteins) are usually called specific transcription factors or transactivators; they may be either activators or repressors of the transcription of specific genes. Several motifs mediate the binding of these regulatory proteins to DNA. The regulatory proteins bind with high affinity to the correct region of DNA. Four unique motifs the helix-turn-helix, the zinc finger, helix-loop-helix and the leucine zipper—account for many of these specific protein- DNA interactions. Comparison of the binding activities of the proteins that contain these motifs leads to several important generalizations:
(1) Binding must be of high affinity to the specific site and of low affinity to other DNA.
(2) Small regions of the protein make direct contact with DNA; the rest of the protein, in addition to providing the trans-activation domains, may be involved in the dimerization of monomers of the binding protein, may provide a contact surface for the formation of heterodimers, may provide one or more ligand-binding sites.
(3) The protein-DNA interactions are maintained by hydrogen bonds and van der Waals forces.
(4) Proteins with the helix-turn-helix or leucine zipper motifs form symmetric dimers, and their respective DNA binding sites are symmetric palindromes. In proteins with the zinc finger motif, the binding site is repeated two to nine times. These features allow for cooperative interactions between binding sites and enhance the degree and affinity of binding.

The Helix-Turn-Helix Motif

Consists of three antiparallel β sheets (β1–β3) and three α helices (α1–α3). The α3 helices form the DNA recognition surface of many proteins, and the rest of the molecule appears to be involved in stabilizing these structures. The helix-turn-helix motif is formed because the α3 and α2 helices are held at about 90 degrees to each other by a turn of four amino acids. Two monomers associate through the antiparallel β3 sheets to form a dimmer that has a twofold axis of symmetry. The average diameter of an α helix is 1.2 nm, which fit in the major groove in the B form of DNA. The DNA recognition domain interacts with 5 bp.

The Zinc Finger Motif

Zinc finger motifs (commonly found in the DNA binding domain of some hormone receptors) contain a bound zinc at four positions with either four cysteine, or two cysteine and two histidine in a sequence of approximately 20 amino acids. The result is a relatively small, tight, autonomously-folded domain. Each zinc finger contacts about 5 bp of DNA. The zinc is required to maintain the

tertiary structure of this domain. Eukaryotic transcription factors generally have two to six zinc finger motifs that function independently. At least one of the zinc fingers forms an α-helix containing a nucleotide recognition signal, a sequence of amino acids that specifically fits into the major groove of DNA.
Zinc finger motifs consist of an α-helix and a β- sheet. The nucleotide recognition signal (contained within the α-helix) of at least one zinc finger binds to a specific sequence of bases in the major groove of DNA.

The Leucine Zipper Motif

Leucine zippers also function as dimers to regulate gene transcription. The leucine zipper motif is an α-helix of 30 to 40 amino acid residues that contains a leucine every seven amino acids, positioned so that they align on the same side of the helix. Two helices dimerize so that the leucines of one helix align with the other helix through hydrophobic interactions to form a coiled coil. The portions of the dimer adjacent to the zipper bind the DNA through basic amino acid residues (arginine and lysine) that bind to the negatively charged phosphate groups. This DNA binding portion of the molecule also contains a nucleotide recognition signal.

Helix-loop-helix

It consists of a short a helix connected by a flexible loop of amino acids to a second longer helix. A highly basic set of amino acids in one of the helices binds to the DNA.
They also function as dimers that is similar to leucine zipper proteins. The dimerization region consists of a portion of the DNA-gripping helix and a loop to another helix.

Mitochondria DNA (mtDNA)

Mitochondrial DNA is a double strand and circular that is not covered with histone and lacks introns. The mtDNA contains 37 genes, 13 of these genes provide instructions for making enzymes involved in oxidative phosphorylation. The remaining genes provide instructions for making transfer RNA (tRNA) and ribosomal RNA (rRNA).
mtDNA is only inherited from our mother. The mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization and are present at the base of the sperm's tail, which is usually lost during fertilization. mtDNA also does not recombine; there is no shuffling of genes from one generation to the other, as there is with nuclear genes.

Why Study mtDNA?

There are many diseases caused by mutations in **mtDNA**. Mutations in mtDNA increase the production of potentially harmful molecules called reactive oxygen species (ROS) which is the cause of many diseases including cancer. Tumour development in most of cases is associated with mutation to mitochondrial DNA.
Because the mitochondria produce energy in cells, symptoms of mitochondrial diseases often involve degeneration or functional failure of tissue. For example, mtDNA mutations have been identified in some forms of diabetes, deafness, and certain inherited heart diseases. In addition, mutations in mtDNA are able to accumulate throughout an individual's lifetime due to lack of repair mechanism. Evidence suggests that the mtDNA mutations contribute to the progression of Parkinson's and Alzheimer's disease.
In addition to the critical cellular energy-related functions, mitochondrial genes are useful to evolutionary biologists because of their maternal inheritance. By studying patterns of mutations, scientists are able to reconstruct patterns of migration and evolution within and between species. For example, mtDNA analysis has been used to trace the migration of people from Asia to North and South America.