

# Chapter 16

## Gene Structure

### 16.1. Search to Identify the Genetic Material

#### (a) Discovery of Nucleic Acids - Friedrich Miescher, 1869

Miescher isolated the nuclei of white blood cells obtained from pus cells. His experiments revealed that nuclei contained a chemical that contained nitrogen and phosphorus but no sulphur. He called the **chemical nuclein** because it came from nuclei. It later became known as nucleic acid.

#### (b) Proteins Produce Genetic Traits - Archibald Garrod, 1909

Garrod noticed that people with certain genetic abnormalities (inborn errors of metabolism) lacked certain enzymes. This observation linked proteins (enzymes) to genetic traits.

#### (c) Genetic Material can Transform Bacteria - Frederick Griffith, 1931

When *Streptococcus pneumoniae* (pneumococcus) bacteria are grown on a culture plate, some produce smooth shiny colonies (S) while others produce rough colonies (R). This is because the S strain bacteria have a mucous (polysaccharide) coat, while R strain does not. Mice infected with the S strain die from pneumonia infection but mice infected with the R strain do not develop pneumonia. Griffith was able to kill bacteria by heating them. He observed that heat-killed S strain bacteria injected into mice did not kill them. When he injected a mixture of heat-killed S and live R bacteria, the mice died. Moreover, he recovered living S bacteria from the carcasses. He concluded that some substance needed to produce the mucous coat was passed from the dead bacteria (S strain) to the live ones (R strain); they became *transformed*. This must be due to a change in the genotype associated with the transfer of the genetic material.

#### (d) The Transforming Material is DNA - Oswald Avery, Colin MacLeod, and Maclyn McCarty, 1944

Prior to the work of Avery, MacLeod, and McCarty, the genetic material was thought to be protein. Avery, MacLeod, and McCarty worked to determine what the transforming substance was in Griffith's experiment. They purified chemicals from the heat-killed S cells to see which ones could transform live R cells into S cells. They discovered that DNA alone from S bacteria caused R bacteria to become transformed.

They also discovered that protein-digesting enzymes (proteases) and RNA-digesting enzymes (RNase) did not affect transformation, so the transforming substance was not a protein or RNA. Digestion with *DNase* did inhibit transformation, so DNA caused transformation. They concluded that DNA is the hereditary material, but not all biologists were convinced.

#### (e) More Evidence: The Genetic Material is DNA - Alfred D. Hershey and Martha Chase, 1952

Hershey and Chase worked with viruses that infect bacteria called bacteriophages.

The bacteriophage becomes attached to the bacteria and its genetic material then enters the bacterial cell. The bacterial cell treats the viral genetic material as if it was its own and subsequently manufactures more virus particles. Hershey and Chase worked to discover whether it was protein or DNA from the viruses that entered the bacteria. They grew a virus population in medium that contained radioactive phosphorus and another in medium that contained radioactive sulphur. Viruses grown in the presence of radioactive phosphorus contained radioactive DNA but not radioactive protein because DNA contains phosphorus but protein does not. Similarly, viruses grown on radioactive sulphur contained radioactive protein but not radioactive DNA because DNA does not contain sulphur. Radioactive bacteriophages were allowed to attach to *E. coli* bacteria. Then as the infection proceeded, the viral coats were removed from the bacteria by agitating them in a blender. The viruses particles were separated from the bacteria by spinning them in a centrifuge. Bacteria that were infected with viruses that had radioactive DNA were radioactive, indicating that DNA was the material that passed from the virus to the bacteria. Bacteria that were infected with viruses that had radioactive proteins were not radioactive. This indicates that proteins did not enter the bacteria from the viruses. DNA is therefore the genetic material that is passed from virus to bacteria.

## 16.2. Discovery of the Structure of DNA

### (a) Erwin Chargaff, 1940's and early 50's

DNA was thought to contain equal amounts of A, T, T, and C. **Chargaff found** that the base composition of DNA differs among species. His data showed that in each species, the percent of A equals the percent of T, and the percent of G equals the percent of C. so that 50% of the bases were purines (A + G) and 50% were pyrimidines (T + C). **Chargaff's Rule:** *The amount of A = T and the amount of G = C.*

### (b) M.H.F. Wilkins and Rosalind Franklin, early 50's

Wilkins and Franklin studied the structure of DNA crystals using X-rays. They found that the crystals contain regularly repeating subunits. Structures that are close together cause the x-ray to bend more than structures that are further apart. The X pattern produced by DNA suggested that DNA contains structures with dimensions of 2 nm, 0.34 nm, and 3.4 n. The dark structures at the top and bottom indicate that some structure was repeated, suggesting a helix.

### (c) James Watson and Francis H.C. Crick, 1953

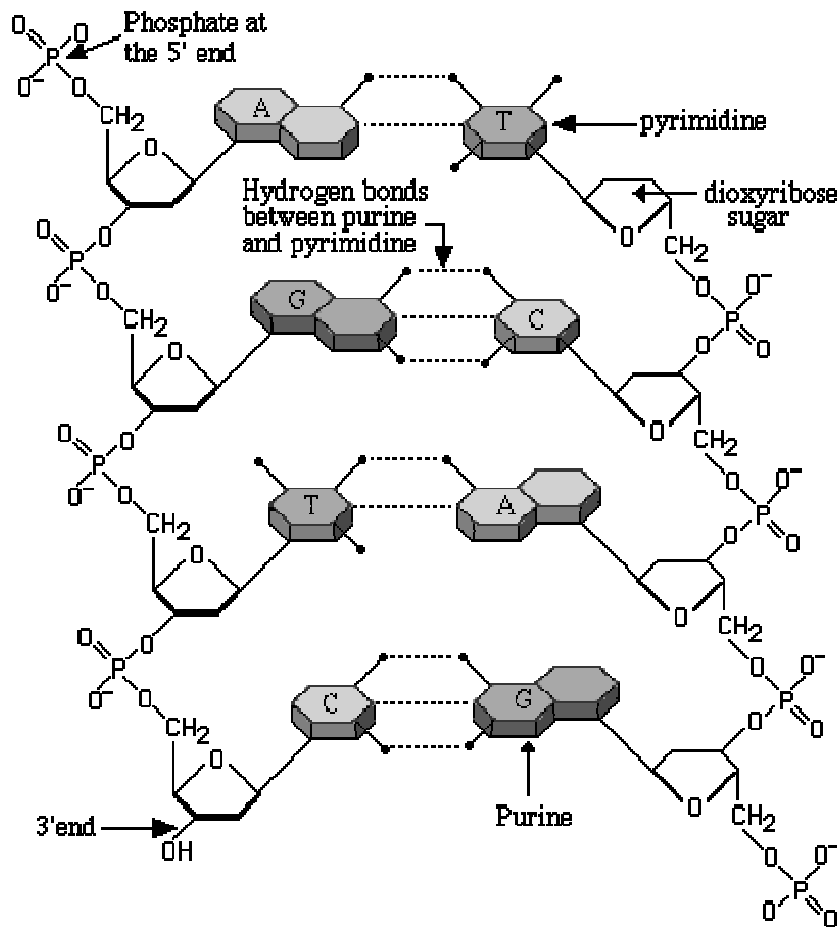
Watson and Crick used Chargaff's base data and Franklin's X-ray diffraction data to construct a model of DNA. The model showed that DNA is a double helix with sugar-phosphate backbones on the outside and the paired nucleotide bases on the inside, in a structure that fit the spacing estimates from the X-ray diffraction data. Chargaff's rules showed that A = T and G = C, so there was complementary base pairing of a purine with a pyrimidine, giving the correct width for the helix. The paired bases can occur in any order, giving an overwhelming diversity of sequences.

## 16.3. Properties of Genetic Material

DNA is an ideal genetic material because it can store information, is able to replicate, and is able to undergo changes (mutate).

## 16.3.1. Structure of DNA

DNA is composed of units called nucleotides. Each nucleotide contains a phosphate group, a deoxyribose sugar, and a nitrogenous base. The nucleotides joined together to form a chain. The phosphate end of the chain is referred to as the 5' end. The opposite end is the 3' end. DNA is composed of two chains of nucleotides linked together in a ladder-like arrangement with the sides composed of alternating deoxyribose sugar and phosphate groups and the rungs being the nitrogenous bases as indicated by the diagram below.



*Fig.16.1. Structure of DNA*

The "A" of one strand is always paired with a "T" on the other. Similarly, the "G" of one strand is paired with a "C" on the other. The two strands are held together by **hydrogen bonds** (electrostatic attraction). Two hydrogen bonds hold adenine to thymine. Three bonds attach cytosine to guanine as indicated in the diagram above. During the process of cell division, the DNA becomes tightly coiled, forming structures called **chromosomes**. The diagram below is a portion of a double-stranded chromosome showing the centromere and a portion of the base sequence. The diagram does not show the extensive looping and coiling and the proteins associated with coiling. Notice that the base sequence in the two chromatids is identical.

## 16.3.2. Storage of Genetic Information

The diagram below shows that one strand of the DNA double-helix serves as a template for the construction of mRNA. The sequence of nucleotides in this DNA strand is complementary (opposite) the sequence in mRNA. The diagram also shows that the sequence of nucleotides in mRNA determines the amino acids in the protein. For example GUG in mRNA (or CAC in DNA) codes for valine. The strand of DNA that contains the genetic code is called the anti-sense. It is often referred to as the coding strand or the template strand. The other strand (the sense strand) is not used. Notice that the sense strand has the same base sequence as mRNA except that mRNA has U instead of T.

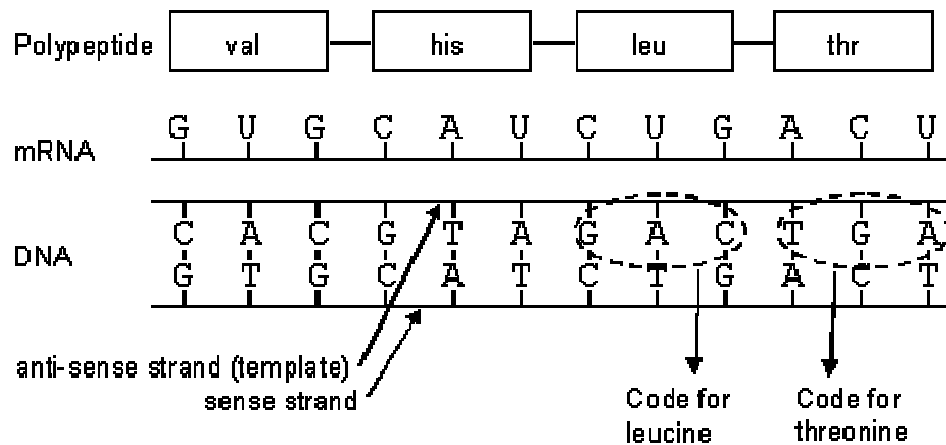


Fig.16.2. Storage of Genetic Information

The codes in DNA are copied to produce mRNA. Each three-letter code in mRNA (called a **codon**) codes for one amino acid. The sequence of amino acids in proteins is therefore most directly determined by the sequence of codons in mRNA, which in turn, are determined by the sequence of bases in DNA. There are four letters in the genetic alphabet (A, T, G, and C) and each codon contains three letters. It is therefore possible to have 64 different codons. Because there are only 20 different amino acids and 64 possible codons, some amino acids have several different codons. Terminators are codes that indicate the end of a genetic message (gene). An initiator codon (usually AUG) indicates where the genetic information begins.

### 16.3.3. DNA Replication

DNA replication involves: The DNA must be unwound and bonds between the bases broken so that the two strands become separated. Each strand serves as a template for the synthesis of a new strand. DNA polymerase adds nucleotides to match to the nucleotide present on the template strand. A is paired with T and G with C. Because each molecule of DNA contains one strand from the original strand, the replication process is **semiconservative replication**. The nucleotides used for synthesis are ATP, GTP, CTP and TTP. Each of these DNA nucleotides has three phosphate groups. Two of the phosphates will be removed when the nucleotide is attached to the growing chain of new DNA. The strand shown on the right side of the diagram must be synthesized in fragments because DNA polymerase always moves in the direction of 5' to 3'. The area in a DNA molecule where unwinding is occurring is called a **replication fork**. In the diagram, it looks like an upside-down Y. The resulting fragments are called **Okazaki fragments**. Covalent bonds must be formed between the newly-added nucleotides. DNA helicase unwinds the DNA

molecule by breaking hydrogen bonds. DNA polymerase lengthens the strand that is being synthesized by adding nucleotides that are complementary to those on the template strand (A paired with T and G paired with C). It proofreads the new strand as it synthesizes it. Incorrectly paired bases are removed and the correct one is inserted. DNA polymerase cannot initiate a new strand, it can only elongate a strand that is already present. Synthesis of new DNA therefore cannot begin until a short strand of nucleotides is added. This short strand is called a primer. Primase creates an RNA primer. DNA polymerase can extend this strand by adding DNA nucleotides. The RNA primer will be removed and replaced by DNA. DNA ligase catalyzes the formation of the covalent bonds between the Okazaki fragments. DNA synthesis occurs at numerous different locations on the same DNA molecule (hundreds in a human chromosome). These form bubbles of replication with a **replication fork** at the growing edge. The replication rate of eukaryotic DNA is 500 to 5000 base pairs per minute. A human cell typically requires a few hours to duplicate the 6 billion base pairs.

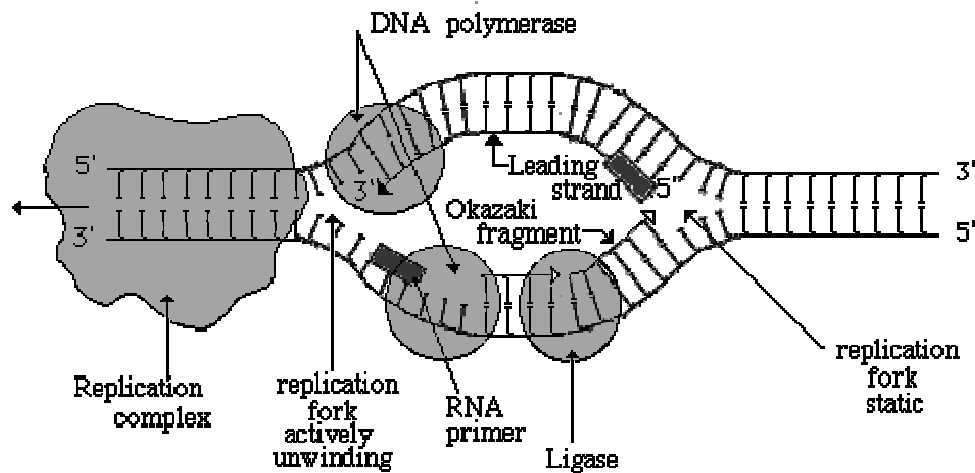


Fig.16.3. DNA Replication

### 16.3.4. Repair of Damaged DNA

Changes in the DNA code are called **mutations**. Repair enzymes repair most of the errors that occur in DNA. There are three different classes of repair mechanisms.

1. Proofreading corrects errors made during the DNA replication process.
2. Mismatch repair corrects base pair mismatching (A-T and G-C).
3. Excision repair removes and replaces small segments of damaged DNA.

#### (a) Errors Corrected as a Result of DNA Synthesis

##### (i) Proofreading

The overall error rate during DNA replication in E.coli is one base in one million ( $10^6$ ). DNA polymerase proofreads the new strand of DNA as it is synthesized and it removes mismatched bases and replaces them with the correct bases. After proofreading, the error rate is 1 in 1 billion ( $10^9$ ) base pairs.

##### (ii) Mismatch Repair

After DNA is replicated, some enzymes function to locate mismatched base pairs, remove a short segment of nucleotides containing the error, and replace the segment with

the correct nucleotides. The new segment is then sealed to the original strand by DNA ligase. Recall that this is the enzyme that seals the Okazaki fragments during DNA synthesis.

When repair enzymes detect a pairing error, how do they know which DNA strand contains the error? The repair enzymes are capable of distinguishing between the original strand of DNA and the new strand that contains the error because the new strand is not **methylated**. Methylation involves adding methyl groups (CH<sub>3</sub>) after DNA is synthesized. Shortly after DNA is synthesized, however, the new strand is not yet methylated.. Mismatch repair enzymes are able to detect which strand is not methylated.

### **(b) Other Mutations**

A number of environmental agents such as radiation (UV, X-rays, radioactive elements) and chemicals (pesticides, cigarette smoke) can cause mutations (changes) in DNA. A number of enzymes monitor the DNA and repair these changes. For example, excision repair occurs when a mutated segment of DNA is removed and replaced with a new segment. A common type of mutation caused by ultraviolet radiation occurs when two thymines become bonded to each other, forming a kink in the DNA molecule. This type of mutation, called a **thymine dimer**, can result in incorrect nucleotides being paired with it when the strand is replicated. To repair this mutation, an enzyme removes a segment of DNA that contains the dimer and replaces the removed nucleotides with nucleotides complimentary to the opposite strand. The new DNA is then bonded to the original strand with DNA ligase. *Xeroderma pigmentosum* is a genetic disease in which some repair enzymes do not function.

## **16.3.5. Organization of DNA**

Chromosomes are structures composed of condensed DNA and associated proteins. When DNA condenses, the molecule becomes wrapped around proteins called **histones**. The histones are then arranged in a coiled pattern to produce a larger fibre. This larger fibre is further compacted by looping to produce looped domains. The looped domains are coiled and compacted to produce chromosomes. **Chromatin** is DNA and its associated protein. **Heterochromatin** is DNA that is coiled and condensed. In this state, it is not transcribed. **Euchromatin** is less condensed and is actively transcribed. During interphase, looped domains may be attached to protein supporting structures on the inside of the nuclear membrane. Some of the DNA is coiled and compacted but other parts are not. Less than 5% of eukaryotic DNA functions to code for proteins. Approximately 1.5% of human DNA codes for protein. The function of the remaining DNA is not known but perhaps much of it has no function. Noncoding DNA is sometimes called "junk DNA". Some parts of the DNA contain more genes than other parts. The gene-rich portions are rich in G and C while the junk DNA is rich in A and T. The light bands on chromosomes are gene-rich regions. 10-25% of eukaryotic DNA consists of sequences of 5 to 10 nucleotides repeated 100,000 to 1,000,000 times. This type of DNA probably does not code for proteins. A large proportion of this type of DNA is found at the tips of the chromosomes and at the centromere.

DNA polymerase is not capable of initiating the synthesis of DNA; it can only elongate a strand that has already been started. Normally, an RNA primer functions to begin the process, allowing DNA polymerase to attach and finish synthesizing the strand. A DNA polymerase molecule will then replace the RNA nucleotides with DNA nucleotides. This

is not a problem for primers that are not located on the 3' end of a DNA strand because DNA polymerase extends the DNA strand that is already there. RNA Primers located on the 5' end of a DNA strand cannot be replaced because DNA polymerase cannot begin at the end of a strand. It can only add to an existing strand. The new DNA strand is shorter than the template strand. As a result of the inability of DNA polymerase to initiate synthesis, the DNA molecule becomes shorter with each cell division. Human chromosomes have the sequence "TTAGGG" repeated 100 to 1500 times at each end of the DNA strand. These repetitive sequences do not contain any genetic information. Each time a cell divides, 50 to 500 of these repeats are lost, making the DNA shorter. Short telomeres may prevent a cell from dividing. The length of telomeres, therefore, may limit the number of times a cell can divide. Telomerase is an enzyme that restores the length of telomeres. This enzyme is normally not found in somatic (body) cells but is found in germ cells.

Some genes are present in many identical or very similar copies called *multigene families*. Multiple copies of identical genes usually code for ribosomal RNA, ribosomal proteins, and histones. Hundreds to thousands of copies of these genes result in faster production of ribosomes and histones. Copies of similar genes probably evolved from the same ancestral gene. The globin genes are an example. Hemoglobin is composed of two alpha chains and two beta chains. These two genes probably evolved from the same ancestral gene and, in turn, gave rise to a family of various alpha globin genes and also a family of various beta globin genes.

## 16.4. Prokaryotes

The DNA of prokaryotes is not condensed into chromosomes as in eukaryotes. Their chromosome consists of a single loop of DNA. Replication begins at a single origin and proceeds in both directions. The rate of replication is approximately 1,000,000 base pairs per minute. It typically requires 40 minutes. Another round of replication may begin before the previous one has finished, thus allowing cells to divide every 20 minutes.