

## Lecture 2: Molecular Genetics

1. The physical carrier of inheritance
2. The Structure of DNA
3. DNA Replication
4. One gene - one protein hypothesis
5. The Structure of hemoglobin
6. Transcription: making an RNA copy of a DNA sequence
7. The Genetic code: translation of RNA code into protein
8. Protein biosynthesis
9. Protein structure
10. Functions of proteins

**Molecular Biology** is the branch of biology that deals with the formation, structure, and function of macromolecules essential to life.

Biologically important macromolecules are DNA, RNA, and proteins.

The field of molecular biology involves many other areas of biology such as biochemistry and genetics.

It deals with the manipulation of DNA so that it can be sequenced or mutated.

The history of deoxyribonucleic acid (DNA) research begins with **Friedrich Miescher**, a Swiss biologist who in 1868 carried out the first carefully thought out chemical studies on the nuclei of cells. Using the nuclei of pus cells obtained from discarded surgical bandages, Miescher detected a phosphorus-containing substance that he named *nuclein*. He showed that nuclein consists of an acidic portion, which we know today as DNA, and a basic protein portion now recognized as histones, a class of proteins responsible for the packaging of DNA. Later he found a similar substance in the heads of salmon sperm cells. Although he separated the nucleic acid fraction and studied its properties, the covalent structure of DNA did not become known with certainty until the late 1940s.

In 1928, **Frederick Griffith**, a British medical officer and geneticist, made a series of unexpected observations while performing an experiment with the disease-causing bacteria pneumococcus and laboratory mice.

Experiment. He studied the difference between two strains of bacteria *Streptococcus pneumoniae* . a disease-causing strain and a strain that did not cause pneumonia. The pneumonia-causing strain (the *S-strain*) was surrounded by a capsule. The other strain (the *R-strain*) did not have a capsule and also did not cause pneumonia. Griffith was able to induce a non-pathogenic strain of the bacterium *S. pneumoniae* to become pathogenic. Griffith injected the different strains of bacteria into mice. The S-strain (virulent) killed the mice; the R-strain (avirulent) did not. He further noted that if heat killed S-strain was injected into a mouse, it did not cause pneumonia. When he combined heat-killed S with live R and injected the mixture into a mouse (remember neither alone will kill the mouse) that the mouse developed pneumonia and died. Bacteria recovered from the mouse had a capsule and killed other mice when injected into them! Griffith referred to a *transforming factor* that caused the non-pathogenic bacteria to become pathogenic.

**Conclusion:** from killed avirulent bacteria to the alive virulent bacteria a *transforming factor* is transmitted which turns the avirulent strains into the virulent ones. Griffith's experiments first demonstrated that genetic information could be transferred !!!

In 1944, group of American scientists **Oswald Avery, Colin MacLeod, and Maclyn McCarty** published the 1st experiment to demonstrate that DNA was the genetic material. They repeated the Griffith's experiments themselves and tried to purify the transforming principle initially without success.

However, a new procedure for removing protein from solution developed by M.G. Sevag led to the demonstration that DNA was the transforming principle. They discovered that DNA taken from a virulent (disease-causing) strain of *S. pneumoniae* permanently transformed a non-virulent (or inactive) form of the organism into a virulent form.

In 1952, **Alfred Hershey and Martha Chase** showed by means of radioactive isotope tracer experiments that when a bacterial virus (*bacteriophage T2*) infects its host cell (the bacterium *Escherichia coli*), it is the DNA of the T2 virus, and not its protein coat, which enters the host cell and provides the genetic information for replication of the virus.

Experiment. In a first experiment, they labelled the DNA of phages with radioactive Phosphorus-32 (the element Phosphorus is present in DNA but not present in any of the 20 amino acids from which proteins are made). They allowed the phages to infect *E. coli*, then removed the protein shells from the infected cells with a blender and a centrifuge. They found that the radioactive tracer was visible only in the bacterial cells and not in the protein shells.

In a second experiment, they labelled the phages with radioactive Sulfur-35 (Sulfur is present in the amino acids Cysteine and Methionine, but not in DNA). After separation, the radioactive tracer then was found in the protein shells, but not in the infected bacteria, confirming that the genetic material which infects the bacteria is DNA.

**Conclusion:** Since the radioactive S remained outside the cell while the radioactive P was found inside the cell, therefore DNA was the physical carrier of heredity

Hershey shared the 1969 Nobel Prize in Physiology or Medicine for his discoveries concerning the genetic structure of viruses.+

In the late 1940s, **Erwin Chargaff**, an Austrian biochemist who emigrated to the United States during the Nazi era, and his colleagues at Columbia University analyzed the nitrogenous bases in many different forms of life and discovered two rules (now Chargaff's rules, or rules of base pairing, or nucleotide pairing) which state that in DNA:

1. The ratio of pyrimidines to purines must be 1:1, with the amount of the adenine (A) equal to the thymine (T), and guanine (G) equal to cytosine (C).
2. The composition of DNA varies from one species to another.

Despite proof that DNA carries genetic information from one generation to the next, the structure of DNA and the mechanism by which genetic information is passed on to the next generation remained the single greatest unanswered question in biology until 1953.

Many scientists were interested in deciphering the structure of DNA, among them were **Francis Crick**, an English physicist, **James Watson**, an American geneticist, working at the University of Cambridge in England, **Rosalind Franklin** and **Maurice Wilkins** at King's College (London, England).

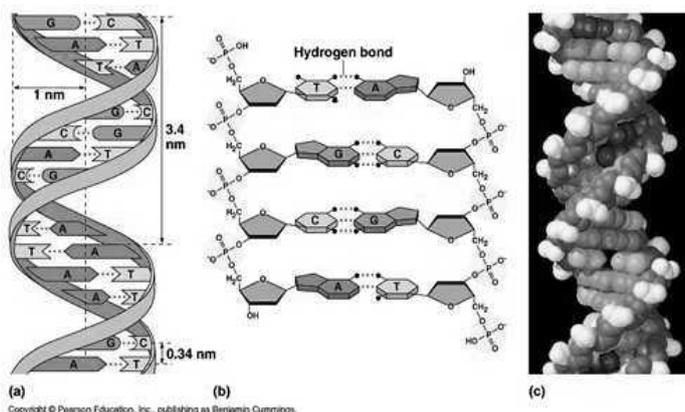
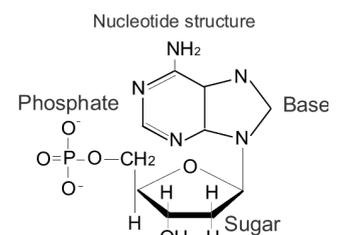
F.Watson and J.Crick gathered all available data in an attempt to develop a model of DNA structure. The data known at the time was that DNA was a long molecule, proteins were helically coiled (as determined by the work of Linus Pauling), Chargaff's base data, and the X-ray diffraction data of R.Franklin and M.Wilkins

R.Franklin, English biophysicist and X-ray crystallographer, took X-ray diffraction photomicrographs of crystalline DNA extract, the key to the puzzle. Franklin's colleague Maurice Wilkins shared her data, without her knowledge, with J.Watson and F.Crick. It was Watson and Crick who put all the pieces of the puzzle together from a variety of sources including R.Franklin's results, to build their ultimately correct and complete description of DNA's structure. Their model for the structure of DNA appeared in the journal *Nature* in April, 1953, alongside Franklin's own report.

For their outstanding work in discovering the double helical structure of DNA, Watson and Crick shared the 1962 Nobel Prize for Physiology and Medicine with Maurice Wilkins. Sadly, Rosalind Franklin, whose work greatly contributed to this key discovery, died before this date, and the rules do not allow a Nobel Prize to be awarded posthumously.

## The Structure of DNA

DNA is a polymer. The monomer units of DNA are nucleotides, and the polymer is known as a "polynucleotide." Each nucleotide consists of a 5-carbon sugar (deoxyribose), a nitrogen containing base attached to the sugar, and a phosphate group. There are four different types of nucleotides found in DNA, differing only in the nitrogenous base. The four nucleotides are given one letter abbreviations as shorthand for the four bases. Adenine (A) and guanine (G) are purines. Purines are the larger of the two types of bases found in DNA. Cytosine (C) and thymine (T) are pyrimidines.



The deoxyribose sugar of the DNA backbone has 5 carbons and 3 oxygens. The carbon atoms are numbered 1', 2', 3', 4', and 5' to distinguish from the numbering of the atoms of the purine and pyrimidine rings. The hydroxyl groups on the 5'- and 3'- carbons link to the phosphate groups to form the DNA backbone. Deoxyribose lacks a hydroxyl group at the 2'-position when compared to ribose, the sugar component of RNA.

DNA is a normally double stranded macromolecule. Two polynucleotide chains, held together by weak thermodynamic forces,

form a DNA molecule. Two DNA strands form a helical spiral, winding around a helix axis in a right-handed spiral. The two polynucleotide chains run in opposite directions to each other and are therefore anti-parallel. The sugar-phosphate backbones of the two DNA strands wind around the helix axis like the railing of a spiral staircase. The bases of the individual nucleotides are on the inside of the helix, stacked on top of each other like the steps of a spiral staircase.

### **Functions of DNA**

1. **DNA is carrier of genetic information**

The function is provided by the fact of existence of genetic code.

2. **Reproduction and transfer of genetic information in the generations of cells and organisms**

The function is provided by the process of DNA replication.

3. **Realization of genetic information in synthesis of proteins and any other substances**

The function is provided by the processes of transcription and translation

### DNA Replication

DNA was proven as the hereditary material and Watson et al. had deciphered its structure. What remained was to determine how DNA copied its information and how that was expressed in the phenotype.

*DNA replication* is the process of copying a double-stranded DNA molecule to form two double-stranded molecules.

The process of DNA replication is a fundamental process used by all living organisms as it is the basis for biological inheritance.

In 1957, two American molecular biologists - **Matthew Meselson** and **Franklin W. Stahl** designed an experiment to determine the method of DNA replication. Three models of replication were considered likely:

1. *Conservative replication*
2. *Semi-conservative replication*
3. *Dispersive replication*

Meselson and Stahl gave experimental evidence of *semi-conservative replication* predicted by J. Watson and F. Crick after their discovery of the double-helix structure of DNA in 1953.

The Meselson-Stahl Experiment. The experiment involved growing a culture of *Escherichia coli* in a medium containing  $^{15}\text{NH}_4\text{Cl}$  (ammonium chloride labeled with the heavy isotope of nitrogen). Cells were then transferred to normal medium (containing  $^{14}\text{NH}_4\text{Cl}$ ) and samples taken after 20 minutes (1 cell division) and 40 minutes (2 cell divisions). DNA was extracted from each sample and the molecules analyzed by density gradient centrifugation. After 20 min the entire DNA contained similar amounts of  $^{14}\text{N}$  and  $^{15}\text{N}$ , but after 40 minutes two bands were seen, one corresponding to hybrid  $^{14}\text{N}$ - $^{15}\text{N}$ -DNA, and the other to DNA molecules made entirely of  $^{14}\text{N}$ . The banding pattern seen after 20 minutes enables conservative replication to be discounted because this scheme predicts that after one round of replication there will be two different types of double helix, one containing just  $^{15}\text{N}$  and the other containing just  $^{14}\text{N}$ . The single  $^{14}\text{N}$ - $^{15}\text{N}$ -DNA band that was actually seen after 20 minutes is compatible with both dispersive and semiconservative replication, but the two bands seen after 40 minutes are consistent only with semiconservative replication. Dispersive replication continues to give hybrid  $^{14}\text{N}$ - $^{15}\text{N}$  molecules after two rounds of replication, whereas the granddaughter molecules produced at this stage by semiconservative replication include two that are made entirely of  $^{14}\text{N}$ -DNA.

**Conclusion:** as predicted by Watson and Crick, DNA strands serve as templates for their own replication.

In DNA replication, the two strands of a helix separate and serve as templates for the synthesis of new strands (nascent strands), so that one helix gives rise to two identical daughter helices. Replication is template process and it can be divided into three stages: *initiation* (start), *elongation* (formation of a chain), *termination* (end).

Nucleotides are added only to the 3' end of a growing nascent chain; therefore, the nascent chain grows only from the 5' to 3' direction. DNA replication has two requirements that must be met:

1. DNA template
2. Free 3' -OH group

### Proteins of DNA Replication

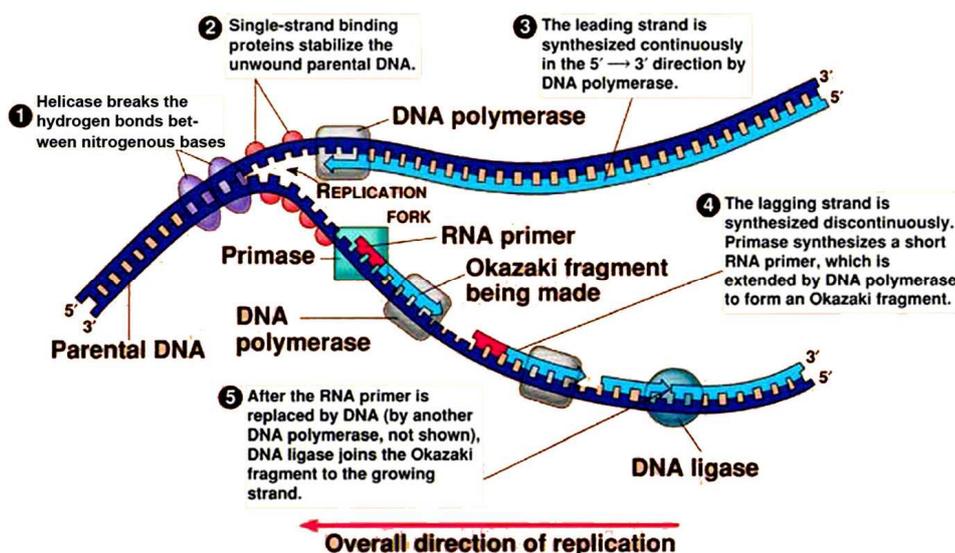
DNA exists in the nucleus as a condensed, compact structure. To prepare DNA for replication, a series of proteins aid in the unwinding and separation of the double-stranded DNA molecule. These proteins are required because DNA must be single-stranded before replication can proceed.

1. **poisomerases** alter the supercoiling of double-stranded DNA and facilitate the untwisting of supertwisted DNA

2. **Helicase** breaks the hydrogen bonds between nitrogenous bases of two opposite strands of a double helix and then separates these strands, exposing two templates. It utilizes the energy of nucleotide hydrolysis to unwind nucleic acid duplexes
3. **DNA-polymerase** forms a new DNA chain in the 5' to 3' direction
4. **DNA-ligase** links up the short DNA segments (known as Okazaki fragments), creating phosphodiester bonds between adjacent Okazaki fragments after replacing the RNA primers with deoxyribonucleosides. The final replication product does not have any nicks because DNA ligase forms a covalent phosphodiester linkage between 3'-hydroxyl and 5'-phosphate groups.
5. **RNA-primase** The requirement for a free 3' hydroxyl group is fulfilled by the RNA primers that are synthesized at the initiation sites by these enzymes. Primase catalyzes the polymerization of RNA building blocks (A, U, G, C) into the primer
6. **RNA-primer** is a short chain of RNA that is formed on the DNA template of lagging strand
7. **Stabilizing proteins (SSB proteins, DNA single-stranded binding proteins)** stabilize the unwound parental DNA. Replication is 100 times faster when these proteins are attached to the single-stranded DNA.

### A General Model for DNA Replication

The DNA molecule is unwound and prepared for synthesis by the action of DNA topoisomerase, DNA helicase and the single-stranded DNA binding proteins. A free 3'OH group is required for replication, but when the two chains separate no group of that nature exists. RNA primers are synthesized, and the free 3'OH of the primer is used to begin replication.



The replication fork moves in one direction, but DNA replication only goes in the 5' to 3' direction. This paradox is resolved by the use of *Okazaki fragments*. These are short, *discontinuous replication* products that are produced off the *lagging strand*. This is in comparison to the continuous strand that is made off the *leading strand*. The final product does not have RNA

stretches in it. These are removed by the 5' to 3' exonuclease action of DNA-polymerase I. The final product does not have any gaps in the DNA that result from the removal of the RNA-primer. These are filled in by the action of DNA-polymerase I. DNA polymerase does not have the ability to form the final bond. This is done by the enzyme DNA ligase.

One gene - one protein hypothesis

**George Beadle** and **Edward Tatum**, two American geneticists, during the late 1930s and early 1940s established the connection between genes and metabolism. They used X-rays to cause mutations in strains of the mold *Neurospora*. These mutations affected a single genes and single enzymes in specific metabolic pathways. Beadle and Tatum proposed the "one gene - one enzyme hypothesis" for which they won the Nobel Prize in Physiology or Medicine 1958 "for their discovery that genes act by regulating definite chemical events"

The Structure of hemoglobin

**Linus Pauling**, American chemists, Nobel Prizer in Chemistry 1954, started studying the disease *sickle-cell anemia*. The disease is caused by a recessive allele (h) in which defective hemoglobin is made, ultimately causing pain and death to those individuals homozygous recessive (hh) for the trait.

Pauling reasoned that if Beadle and Tatum were correct, there should be a slight (but detectable) difference between the structure of a normal (HH) and sickle cell (hh) hemoglobin due to genetic differences. Heterozygotes (Hh, also sampled by Pauling) make both normal and "sickle cell" hemoglobins. Pauling used electrophoresis to separate hemoglobin molecules. Later, Vernon Ingram discovered that the normal and sickle-cell hemoglobins differ by only 1 (out of a total of 300) amino acid.

### Transcription: making an RNA copy of a DNA sequence

**Gene** . a specific sequence of nucleotides in DNA that is located on a chromosome and that is the functional unit of inheritance controlling the transmission and expression of one or more traits

Although humans contain a thousand times more DNA than do bacteria, the best estimates are that humans have only about 20 times more genes than do the bacteria. This means that the vast majority of eukaryotic DNA is apparently nonfunctional.

Genes that are expressed usually have **introns** that interrupt the coding sequences. A typical eukaryotic gene, therefore, consists of a set of sequences that appear in mature mRNA (called **exons**) interrupted by introns.

**Exon** . the coding part of a gene. Exon contains the code for producing protein and is copied and spliced together with other such sequences to form messenger RNA (mRNA). Exons are separated by introns. Exons are not spliced out from the transcribed RNA and are retained in the final messenger RNA (mRNA) molecule.

The term "exon" was coined by Walter Gilbert in 1978.

The typical human gene contains an average of 8 exons. Internal exons average 145 nucleotides (nt) in length, and introns average more than 10 times this size and can be much larger. **Intron** noncoding part of a gene that is initially transcribed into the primary RNA transcript and is removed by splicing.

**Transcription** is the synthesis of RNA on DNA template. It is the process by which the information contained in a section of DNA is transferred to a newly assembled piece of RNA. Transcription takes place in a nucleus.

The basic components of RNA are the same than for DNA. Both nucleic acids are sugar-phosphate polymers and both have nitrogen bases attached to the sugars of the backbone- but there are several important differences. They differ in composition:

1. The sugar in RNA is ribose, not the deoxyribose in DNA
2. The pyrimidine base uracil is present in RNA instead of thymine. Adenine and Uracil for a base pair formed by two hydrogen bonds.

They also differ in size and structure:

1. RNA molecules are smaller (shorter) than DNA molecules,
2. RNA is single-stranded, not double-stranded like DNA.

Another difference between RNA and DNA is in function. DNA has only one function - **STORING GENETIC INFORMATION** in its sequence of nucleotide bases. But there are three main kinds of ribonucleic acid, each of which has a specific job to do.

The process of transcription is in principle similar to DNA replication but there are important differences (see the table below)

	<b>Transcription</b>	<b>DNA Replication</b>
<b>principal enzyme</b>	RNA polymerase	DNA polymerase
<b>nucleotides</b>	A C G U	A C G T
<b>pairings</b>	C-G .....A-U	C-G.....A-T
<b>strand "copied"</b>	lower strand (3'-5')	both strands
<b>regions "copied"</b>	specific regions	all

Like the replication process it is template process. This process involves 3 stages:

1. Initiation
2. Elongation
3. Termination

Transcription begins on the area of DNA that contains the gene. Each gene has three regions:

- 1) **Promoter** - DNA sequence where RNA polymerase enzyme attaches and initiates transcription
- 2) **Coding region** - has the information on how to construct the protein
- 3) **Termination sequence** - signals the end of the gene

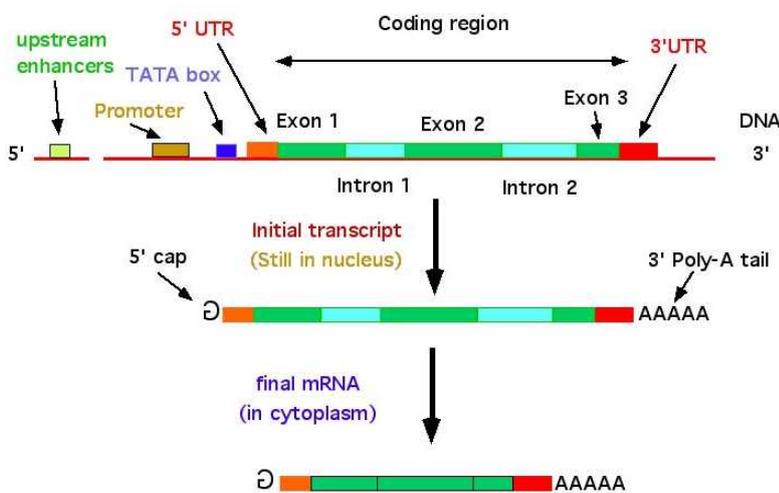
RNA Polymerase enzyme is responsible for reading the gene, and building the mRNA strand

**Initiation.** After RNA Polymerase binds to the promoter, the DNA double helix unwinds and unzips, and the enzyme initiates RNA synthesis at the start point on the template strand.

**Elongation.** The RNA polymerase moves downstream, unwinding the DNA and elongation the RNA transcript 5' → 3'. In the wake of transcription, the DNA strands re-form a double helix

**Termination.** The polymerase transcribes a short DNA terminator sequence, which signals the end of the transcription unit. The **primary RNA transcript** is released, and the enzyme detached from the DNA

All the primary RNA transcripts must undergo **processing** steps to produce functional RNA molecules for export to the cytosol.



### Events of processing:

1. **mRNA Cap** is added to 5' end (replaces the role of the RBS)
2. **Poly-A tail** is added to 3' end of message (AAUAA is the poly-A processing signal)
3. Removal of introns (more on this later) and **splicing** the exons together are among the essential steps in synthesizing mRNA

**Significance:** Alternative splicing of exons allows one gene to make several different mRNAs, depending on which exons are included in the final message. Hence, one "gene"

may code for a large number of different products.

### Genetic Code – Code of Life

Genetic Code is a set of rules, which maps DNA sequences to proteins in the living cell, and is employed in the process of protein synthesis.

Molecular biologists cracked the code of life in the early of 1960s, a decade after Watson and Crick's work, when a series of elegant experiments disclosed the amino acid translation of each of the RNA codons.

The first codon was deciphered in 1961 by American biochemists **Marshall Nirenberg** and his German colleague **Heinrich Matthaei**.

*In the experiment*, an extract from *E.coli* cells that could make protein even when no intact living cells were present was prepared.

By adding an artificial form of RNA poly-U to this extract in each of 20 test-tubes (each tube having a different "tagged" amino acid) Nirenberg and Matthaei were able to determine that the codon UUU (the only one in poly-U) coded for the amino acid phenylalanine

This showed that RNA controlled the production of specific types of protein.

M.Nirenberg won Nobel Prize "for their interpretation of the genetic code and its function in protein synthesis+in 1968

The scientists who pondered the mystery of the genetic code in the late 1950s came up with many creative theories. One main problem to work out was how many bases would be in each code word (later known as a codon).

According to Russian theoretical physicist and astronomer **George Gamow** the code consists of at least three bases. To code for the 20 essential amino acids a genetic code must consist of at least a 3-base set (triplet) of the 4 bases. If one considers the possibilities of arranging four things 3 at a

time (4×4×4), we get 64 possible code words, or codons (a 3-base sequence on the mRNA that codes for either a specific amino acid or a control word)

## Protein Synthesis

*Translation, or protein biosynthesis* is the process during which a base sequence is translated into an amino acid sequence (a polypeptide).

The most important components of protein biosynthesis: mRNA, tRNA, ribosomes, amino acids, energy (ATP, GTP), magnesium ions (are needed for association of ribosome subunits), aminoacyl-tRNA synthetases enzymes.

### Main types of RNA

1. mRNA messenger RNAs - the long form of RNA (has thousands of nitrogenous bases) that are variable in size. They carry information specifying amino acid sequences of proteins from DNA to ribosomes. The codon and the tRNA's anticodon are complementary to one another. The mRNA leaves the nucleus and goes into the cytoplasm. In bacteria, messenger RNAs are short-lived (minutes) while in eukaryotes they have varying half lives (minutes to hours).
2. tRNAs - transfer RNAs (translators). They are small cytoplasmic RNAs that bind amino acids covalently and deliver them to the ribosome for protein synthesis. Less abundant than rRNAs, and similar in size.
3. rRNAs - Ribosomal RNAs. Not translated into protein, but they play catalytic and structural roles in formation of ribosomes. They are the most abundant RNAs in the cell, present in ribosomes and in eukaryotes also in the nucleolus (where rRNA genes are transcribed in the nucleus).

Ribosomes are the organelle (in all cells) where proteins are synthesized. They are the translation factories, made of a large subunit and a small subunit. They consist of two-thirds rRNA and one-third protein.

The ribosome has enzymes and ATP molecules to help it link one amino acid to another one.

Ribosome has 3 main sites: P, E, A

- **A** site holds the tRNA carrying the next amino acid to be added to the polypeptide chain.
- **P** site attached to the growing polypeptide
- **E** site - exit site where tRNA leaves

As all template processes translation has 3 stages (steps): *initiation, elongation, termination*.

During a preparatory step, acylation, tRNA molecules are linked to their respective amino acids, forming complexes called aminoacyl-tRNAs.

The overhanging 3' end of tRNA contains an **CCA-** sequence, which binds the **-COOH** group of an amino acid.

This step occurs in the *cytoplasm* and relies on enzymes called *aminoacyl-tRNA synthetases*, which recognizes the specific shapes of tRNA's and those of amino acids.

### **Initiation**

1. The leading AUG codon associates with the UAC anticodon of an aminoacyl-tRNA which bears the amino acid Methionine
2. The mRNA strand binds to the small ribosomal subunit / tRNA-Met complex.
3. The large subunit binds to the complex, sandwiching the mRNA between the small and large subunits in such a way as to position the attached tRNA at the P site (peptidyl binding site)

### **Elongation**

1. A **new** charged **tRNA** binds to the **A-site** of the ribosome (an ATP is used). In this way, two tRNA's are side by side on the ribosome.
2. The preceding amino acid (Met at the start of translation, at the P-site) is covalently linked to the incoming amino acid (at the A-site) with a **peptide bond** by an enzyme **peptidyl transferase**.
3. The initiator tRNA is released from the P-site.
4. The ribosome moves one codon downstream - a process called **translocation**. This shifts the more recently-arrived tRNA, with its attached peptide, to the P-site and opens the A-site for the arrival of a new aminoacyl-tRNA.
5. This last step is promoted by another protein **elongation factor** (named **EF-G**) and the energy of another molecule of **GTP**.

## Termination

The multiple elongation steps repeat until a **stop** codon, e.g., UAG, is encountered. At this stage, a **releasing factor** is bound to the ribosome, which prevents further binding and dissociates the ribosome-mRNA complex.

Ribosome releases mRNA, tRNA, and the polypeptide chain. End product of translation is a polypeptide structure.

A typical protein consisting of 200-500 amino acids can be produced in this way in about **half a minute**. This rather slow rate is substantially accelerated by simultaneous translation on ~ **10 ribosomes** bound to the same mRNA strand, that is many ribosomes can jump on one mRNA. A lot of polypeptide chains is created.

## Protein structure

The word **PROTEIN** comes from Greek language (*prota*) which means "of primary importance". This name was introduced by *Jons Jakob Berzelius* in 1838 for large organic compounds with almost equivalent empirical formulas. This name was used because the studied organic compounds were primitive but seems to be very important for animal nutrition.

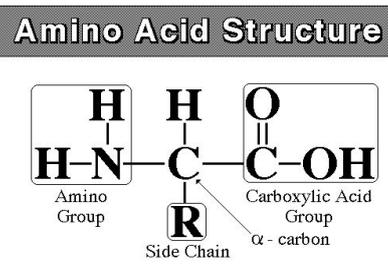
Amino acids, the building blocks of proteins, made up of carboxyl group, amino group, and unique side group. The amino group (NH<sub>2</sub>) of one amino acid joins with the acid group of the next amino acid (C=O. OH).

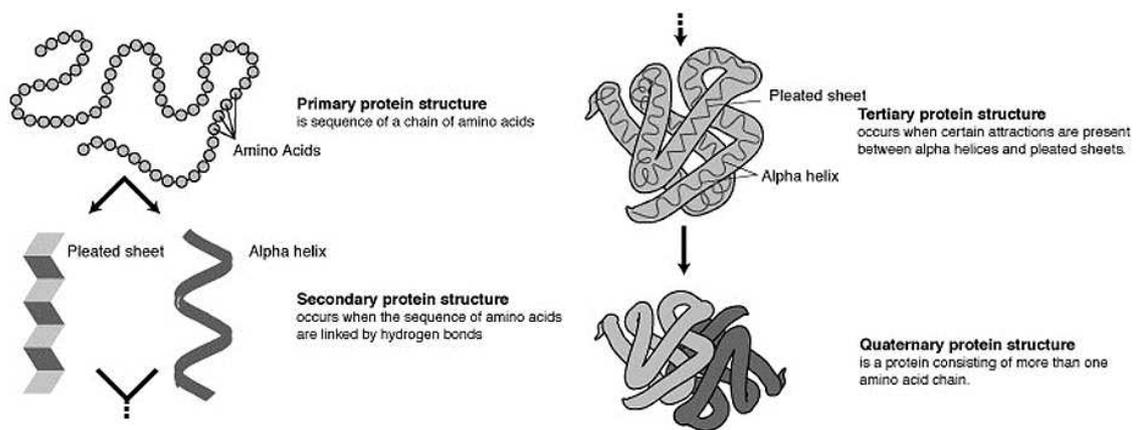
20 different amino acids have different chemical properties

**Biochemistry refers to four distinct aspects of a protein's structure:**

1. **Primary structure** - the amino acid sequence of the peptide chains. The primary structure is held together by *covalent* or *peptide* bonds, which are made during the process of protein biosynthesis or translation. These peptide bonds provide rigidity to the protein. The two ends of the amino acid chain are referred to as the C-terminal end or carboxyl terminus (C-terminus) and the N-terminal end or amino terminus (N-terminus) based on the nature of the free group on each extremity.
2. **Secondary structure** - highly regular sub-structures (alpha helix and strands of beta sheet) which are locally defined, meaning that there can be many different secondary motifs present in one single protein molecule. The various types of secondary structure are defined by their patterns of *hydrogen bonds* between the main-chain peptide groups. However, these hydrogen bonds are generally not stable by themselves, since the water-amide hydrogen bond is generally more favorable than the amide-amide hydrogen bond. Thus, secondary structure is stable only when the local concentration of water is sufficiently low, e.g., in the molten globule or fully folded states.
3. **Tertiary structure** - three-dimensional structure of a single protein molecule; a spatial arrangement of the secondary structures. It also describes the completely folded and compacted polypeptide chain. Similarly, the formation of molten globules and tertiary structure is driven mainly by structurally non-specific interactions, such as the rough propensities of the amino acids and hydrophobic interactions. However, the tertiary structure is fixed only when the parts of a protein domain are locked into place by structurally specific interactions, such as ionic interactions (salt bridges), hydrogen bonds and the tight packing of side chains. The tertiary structure of extracellular proteins can also be stabilized by disulfide bonds, which reduce the entropy of the unfolded state; disulfide bonds are extremely rare in cytosolic proteins, since the cytosol is generally a reducing environment.
4. **Quaternary structure** - complex of several protein molecules or polypeptide chains, usually called protein subunits in this context, which function as part of the larger assembly or protein complex.

In addition to these levels of structure, a protein may shift between several similar structures in performing its biological function. In the context of these functional rearrangements, these tertiary or quaternary structures are usually referred to as chemical conformation, and transitions between them are called conformational changes.





## Protein Functions

- **Structural proteins** - These proteins are maintain structures of other biological components, like cells and tissues. *Collagen, elastin, -keratin, sklerotin, fibroin* - these proteins are involved into formation of the whole organism body. *Bacterial proteoglycans* and *virus coating proteins* also belongs to this group of proteins.
- **Motor proteins** can convert chemical energy into mechanical energy. *Actin* and *myosin* are responsible for muscular motion. Sometimes it is difficult to make a strict separation between structural and motor proteins.
- **Enzymes** - proteins that catalyze chemical and biochemical reactions within living cell and outside. This group of proteins probably is the biggest and most important group of the proteins.
- **Transport proteins** transport or store some other chemical compounds and ions. Some of them are well known: *cytochrome C* - electron transport; *haemoglobin* and *myoglobin* - oxygen transport; *albumin* - fatty acid transport in the blood stream etc.
- **Hormones** - proteins that are responsible for the regulation of many processes in organisms. Hormones are usually quite small and can be classifies as peptides. Most known protein hormones are: *insulin, grows factor, lipotropin, prolactin* etc.
- **Receptors** These proteins are responsible for signal detection and translation into other type of signal. Sometimes these proteins are active only in complex with low molecular weight compounds. Very well known member of this protein family is *rhodopsin* - light detecting protein. Many receptors are transmembrane proteins.
- **Signalling proteins** - This group is involved into signalling translation process. Usually they significantly change conformation in presence of some signalling molecules. These proteins can act as enzymes. Other proteins, usually small, can interact with receptors. Classical example of this group of proteins is *GTPases*.
- **Immunoglobulin or antibodies** - proteins that involved into immune response of the organism to neutralize large foreign molecules, which can be a part of an infection. Sometimes antibodies can act as enzymes. Sometimes this group of proteins is considered as a bigger group of protective proteins with adding such proteins as *lymphocyte antigen-recognizing receptors*, antivirals agents such as *interferon, tumor necrosis factor (TNF)*. Probably the clotting of blood proteins, such as fibrin and thrombin should be classified as protective proteins as well.
- **Storage proteins.** These proteins contain energy, which can be released during metabolism processes in the organism. *Egg ovalbumin* and *milk casein* are such proteins. Almost all proteins can be digested and used as a source of energy and building material by other organisms.