MODULE 3 PROTEINS – STRUCTURE AND FUNCTION

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. Proteins may be structural, regulatory, contractile, or protective; they may serve in transport, storage, or membranes; or they may be toxins or enzymes. Each cell in a living system may contain thousands of different proteins, each with a unique function. Their structures, like their functions, vary greatly. They are all, however, polymers of amino acids, arranged in a linear sequence.

Types and Functions of Proteins

Protein Types and Functions

Type	Examples	Functions
Digestive Enzymes	Amylase, lipase, pepsin, trypsin	Help in digestion of food by catabolizing nutrients into monomeric units
Transport	Haemoglobin, albumin	Carry substances in the blood or lymph throughout the body
Structural	Actin, tubulin, keratin	Construct different structures, like the cytoskeleton
Hormones	Insulin, thyroxine	Coordinate the activity of different body systems
Defense	Immunoglobulins	Protect the body from foreign pathogens
Contractile	Actin, myosin	Effect muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide nourishment in early development of the embryo and the seedling

Proteins have different shapes and molecular weights; some proteins are globular in shape whereas others are fibrous in nature. For example, haemoglobin is a globular protein, but collagen, found in our skin, is a fibrous protein. Protein shape is critical to its function, and this shape is maintained by many different types of chemical bonds. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to loss of function, known as denaturation. All proteins are made up of different arrangements of the same 20 types of amino acids. The functions of proteins are very diverse because there

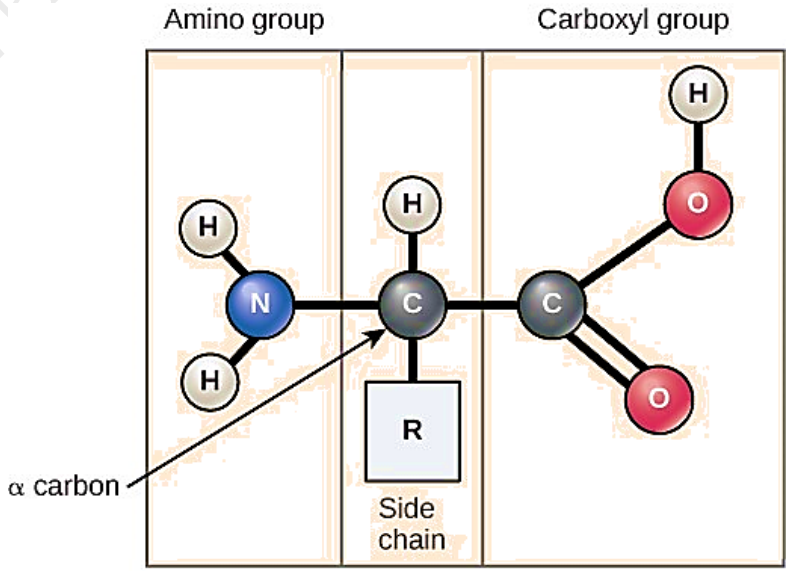
are 20 different chemically distinct amino acids that form long chains, and the amino acids can be in any order. For example, proteins can function as enzymes or hormones.

Enzymes, which are produced by living cells, are catalysts in biochemical reactions (like digestion) and are usually proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) upon which it acts. Enzymes can function to break molecular bonds, to rearrange bonds, or to form new bonds. An example of an enzyme is salivary amylase, which breaks down amylose, a component of starch.

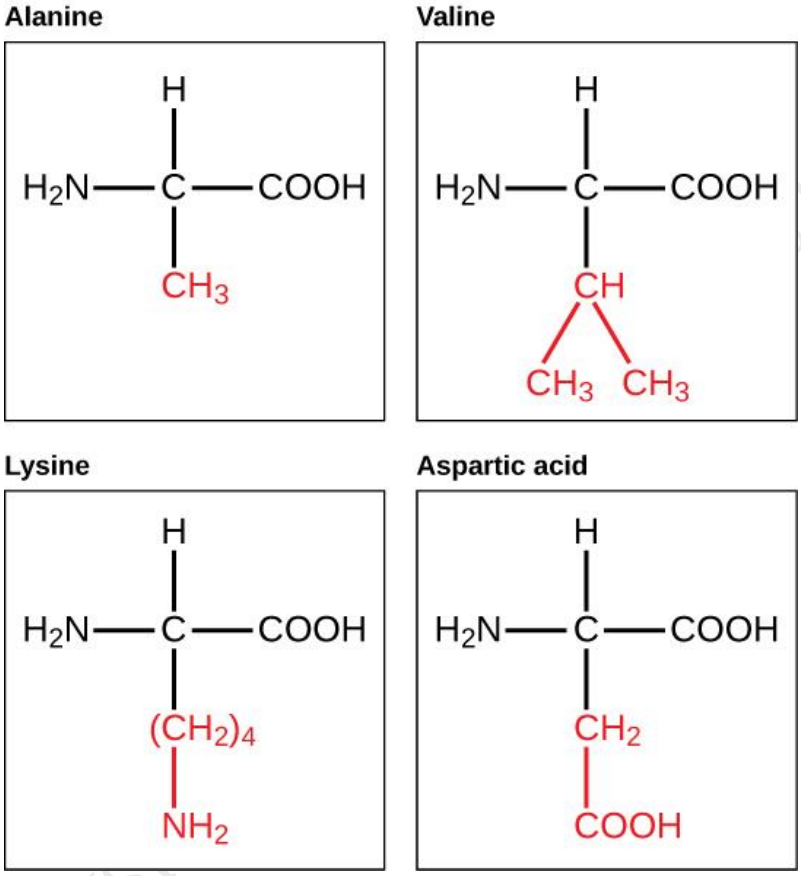
Hormones are chemical signalling molecules, usually proteins or steroids, secreted by an endocrine gland or group of endocrine cells that act to control or regulate specific physiological processes, including growth, development, metabolism, and reproduction. For example, insulin is a protein hormone that maintains blood glucose levels. Proteins have different shapes and molecular weights; some proteins are globular in shape whereas others are fibrous in nature. For example, haemoglobin is a globular protein, but collagen, found in our skin, is a fibrous protein. Protein shape is critical to its function. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to a loss of function or **denaturation** (to be discussed in more detail later). All proteins are made up of different arrangements of the same 20 kinds of amino acids.

AMINO ACIDS

Amino acids are the monomers that make up proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group (COOH), and to a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the central atom known as the R group



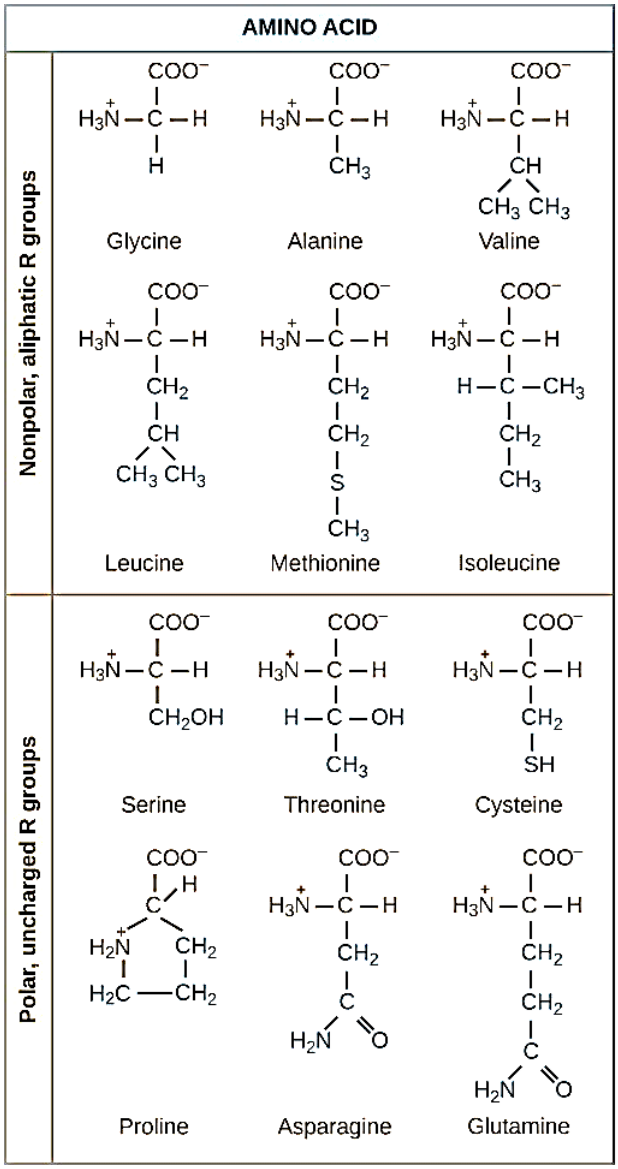
Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached. The name "amino acid" is derived from the fact that they contain both amino group and carboxyl-acid-group in their basic structure. As mentioned, there are 20 amino acids present in proteins. Ten of these are considered essential amino acids in humans because the human body cannot produce them and they are obtained from the diet. For each amino acid, the R group (or side chain) is different

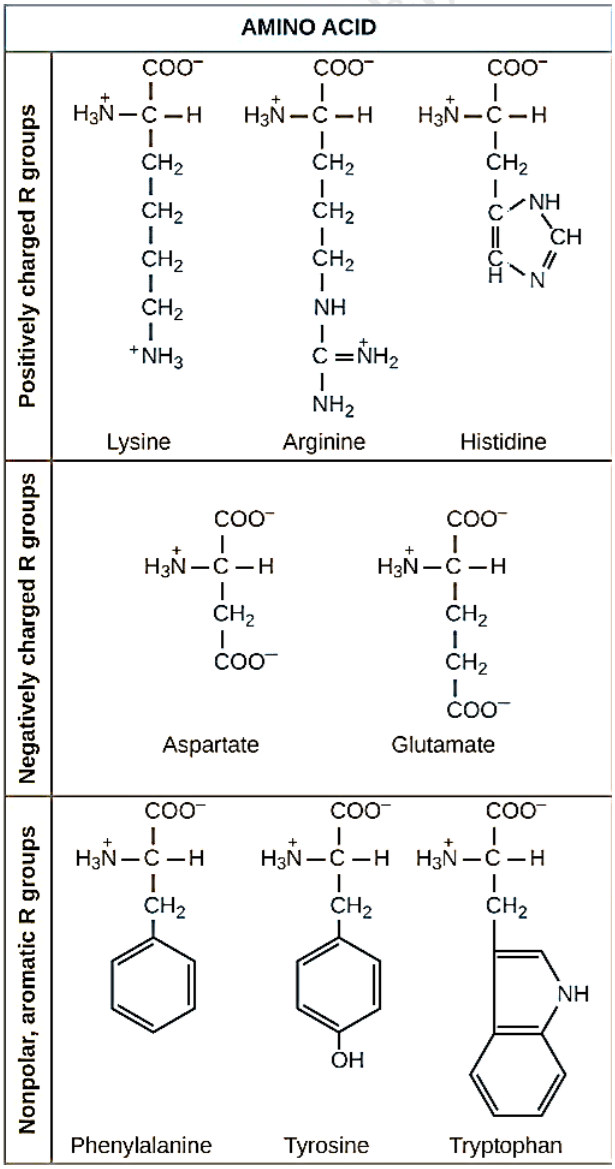


Amino acids are made up of a central carbon bonded to an amino group (–NH_2), a carboxyl group (–COOH), and a hydrogen atom. The central carbon's fourth bond varies among the different amino acids, as seen in these examples of alanine, valine, lysine, and aspartic acid. The chemical nature of the R group determines the chemical nature of the amino acid within its protein (that is, whether it is acidic, basic, polar, or nonpolar).

Essential Amino Acids (must be obtained from the diet)	**Non-Essential Amino Acids (can be synthesized by the body)**
Histidine (His, H)	Alanine (Ala, A)
Isoleucine (Ile, I)	Arginine (Arg, R) (conditionally essential, especially during growth)
Leucine (Leu, L)	Asparagine (Asn, N)

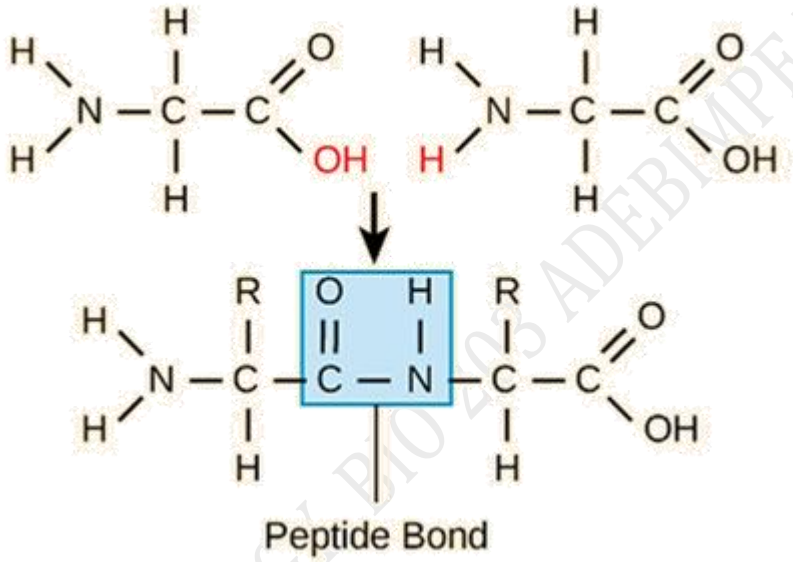
Lysine (Lys, K)	Aspartic acid (Asp, D)
Methionine (Met, M)	Cysteine (Cys, C) (conditionally essential)
Phenylalanine (Phe, F)	Glutamic acid (Glu, E)
Threonine (Thr, T)	Glutamine (Gln, Q) (conditionally essential)
Tryptophan (Trp, W)	Glycine (Gly, G)
Valine (Val, V)	Proline (Pro, P)
	Serine (Ser, S)
	Tyrosine (Tyr, Y) (conditionally essential, derived from phenylalanine)





Each amino acid has a unique side chain that defines its characteristics and roles in proteins.

The sequence and number of amino acids ultimately determine a protein's shape, size, and function. Each amino acid is attached to another amino acid by a covalent bond, known as a peptide bond, which is formed by a dehydration reaction. The carboxyl group of one amino acid and the amino group of a second amino acid combine, releasing a water molecule. The resulting bond is the peptide bond. The products formed by such a linkage are called polypeptides. While the terms polypeptide and protein are sometimes used interchangeably, a **polypeptide** is technically a polymer of amino acids, whereas the term protein is used for a polypeptide or polypeptides that have combined together, have a distinct shape, and have a unique function.



Cytochrome c

Cytochrome c is an important component of the molecular machinery that harvests energy from glucose. Because this protein's role in producing cellular energy is crucial, it has changed very little over millions of years. Protein sequencing has shown that there is a considerable amount of sequence similarity among cytochrome c molecules of different species; evolutionary relationships can be assessed by measuring the similarities or differences among various species' protein sequences.

For example, scientists have determined that human **cytochrome c contains 104 amino acids**. For each cytochrome c molecule that has been sequenced to date from different organisms, 37 of these amino acids appear in the same position in each cytochrome c. This indicates that all of these organisms are descended from a common ancestor. On comparing the human and chimpanzee protein sequences, no sequence difference was found. When human and rhesus monkey sequences were compared, a single difference was found in one amino acid. In contrast, human-to-yeast comparisons show a difference in 44 amino acids, suggesting that

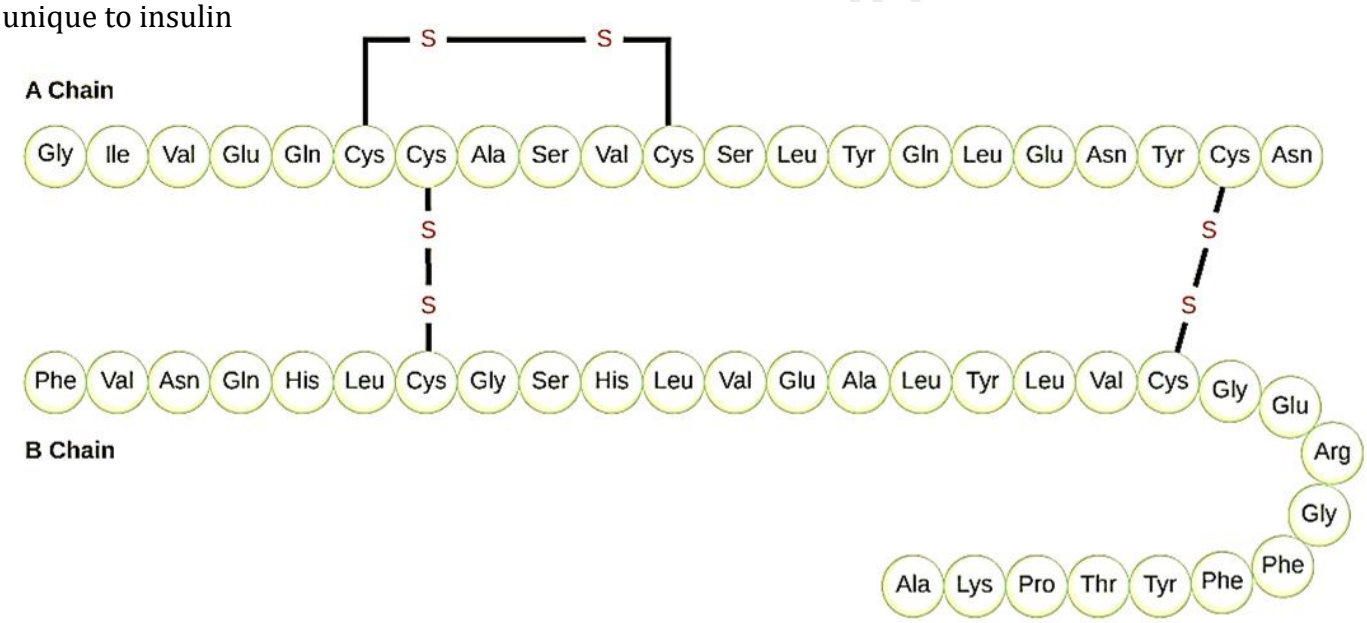
humans and chimpanzees have a more recent common ancestor than humans and the rhesus monkey, or humans and yeast.

Protein Structure

As discussed earlier, the shape of a protein is critical to its function. To understand how the protein gets its final shape or conformation, we need to understand the four levels of protein structure: **Primary, Secondary, Tertiary, and Quaternary.**

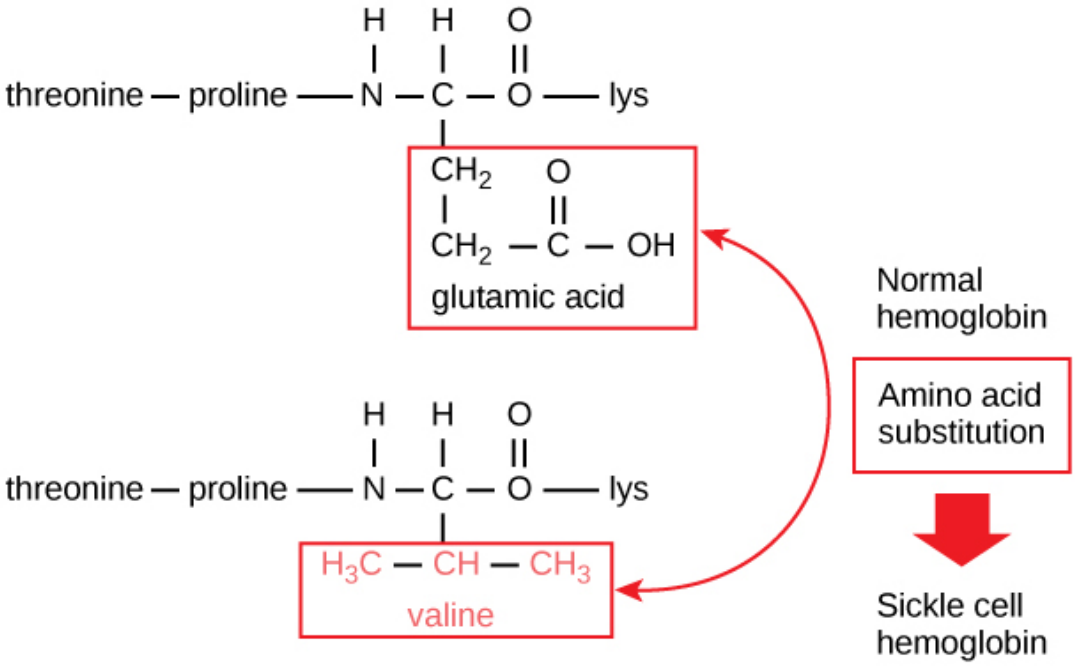
Primary Structure

The unique sequence of amino acids in a polypeptide chain is its primary structure. For example, the pancreatic hormone insulin has two polypeptide chains, A and B, and they are linked together by disulfide bonds. The N terminal amino acid of the A chain is glycine, whereas the C terminal amino acid is asparagine. The sequences of amino acids in the A and B chains are unique to insulin



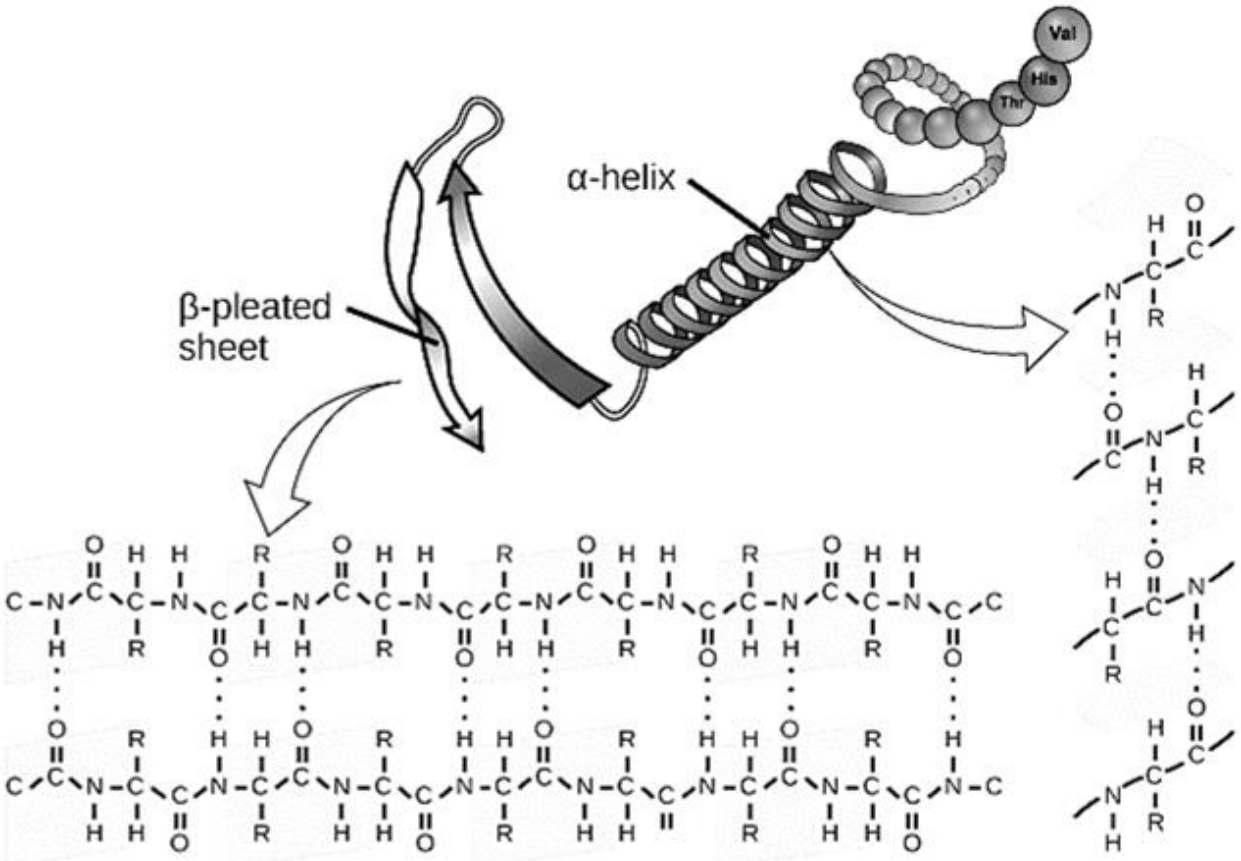
Bovine serum insulin is a protein hormone made of two peptide chains, A (21 amino acids long) and B (30 amino acids long). In each chain, primary structure is indicated by three-letter abbreviations that represent the names of the amino acids in the order they are present. The amino acid cysteine (cys) has a sulfhydryl (SH) group as a side chain. Two sulfhydryl groups can react in the presence of oxygen to form a disulfide (S-S) bond. Two disulfide bonds connect the A and B chains together, and a third helps the A chain fold into the correct shape. Note that all disulfide bonds are the same length, but are drawn different sizes for clarity.

The unique sequence and number of amino acids in a polypeptide chain is its primary structure. The unique sequence for every protein is ultimately determined by the gene that encodes the protein. Any change in the gene sequence may lead to a different amino acid being added to the polypeptide chain, causing a change in protein structure and function. **William Warrick Cardozo** showed that sickle-cell anaemia is caused by a change in protein structure as a result of gene encoding, meaning that it is an inherited disorder. In sickle cell anaemia, the haemoglobin β chain has a single amino acid substitution, causing a change in both the structure and function of the protein. What is most remarkable to consider is that a haemoglobin molecule is made up of two alpha chains and two beta chains that each consist of about 150 amino acids. The molecule, therefore, has about 600 amino acids. The structural difference between a normal hemoglobin molecule and a sickle cell molecule—that dramatically decreases life expectancy in the affected individuals—is a single amino acid of the 600.



Secondary Structure

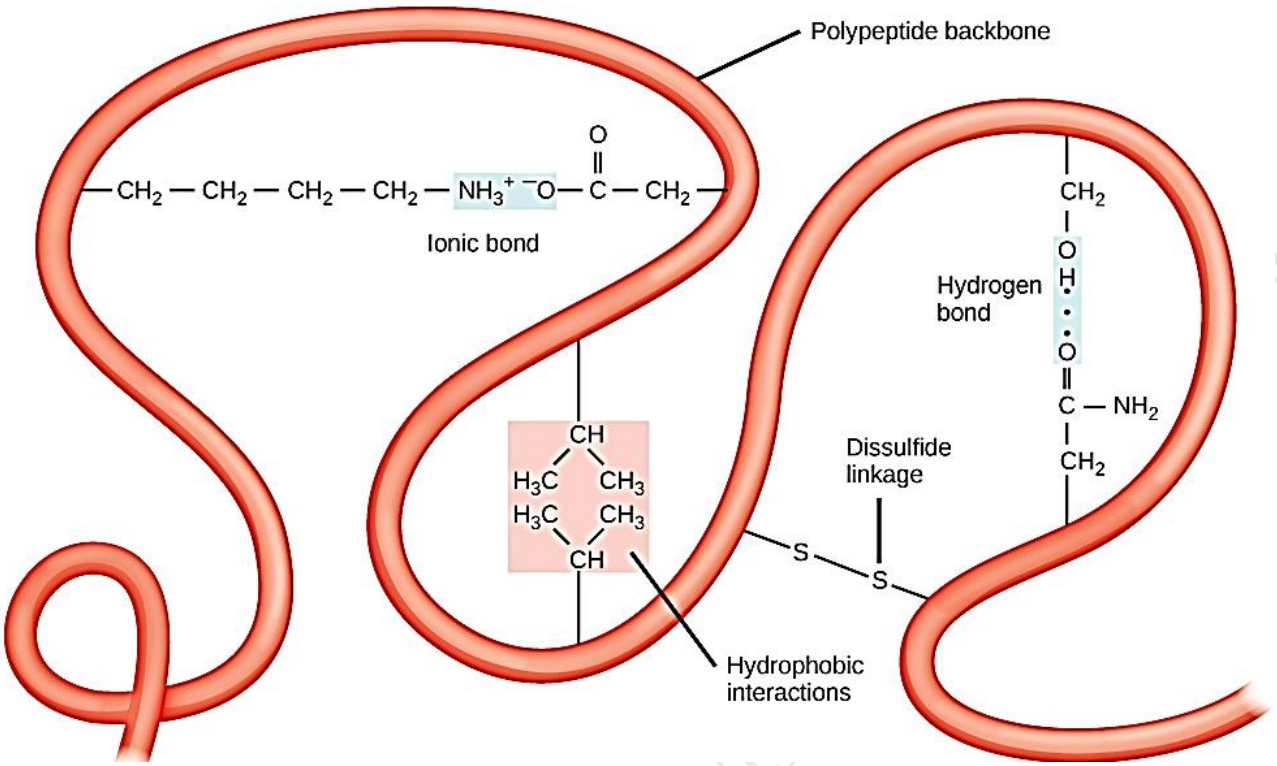
The local folding of the polypeptide in some regions gives rise to the secondary structure of the protein. The most common are the α-helix and β-pleated sheet structures (Figure 3.4.7). Both structures are the α-helix structure—the helix held in shape by hydrogen bonds. The hydrogen bonds form between the oxygen atom in the carbonyl group in one amino acid and another amino acid that is four amino acids farther along the chain.



The α-helix and β-pleated sheet are secondary structures of proteins that form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an α-helix, while others have a propensity to form a β-pleated sheet.

Tertiary Structure

The unique three-dimensional structure of a polypeptide is its tertiary structure (Figure 3.4.8). This structure is in part due to chemical interactions at work on the polypeptide chain. Primarily, the interactions among R groups creates the complex three-dimensional tertiary structure of a protein. The nature of the R groups found in the amino acids involved can counteract the formation of the hydrogen bonds described for standard secondary structures. For example, R groups with like charges are repelled by each other and those with unlike charges are attracted to each other (ionic bonds). When protein folding takes place, the hydrophobic R groups of nonpolar amino acids lay in the interior of the protein, whereas the hydrophilic R groups lay on the outside. The former types of interactions are also known as hydrophobic interactions. Interaction between cysteine side chains forms disulfide linkages in the presence of oxygen, the only covalent bond forming during protein folding.



The tertiary structure of proteins is determined by a variety of chemical interactions. These include hydrophobic interactions, ionic bonding, hydrogen bonding and disulfide linkages.

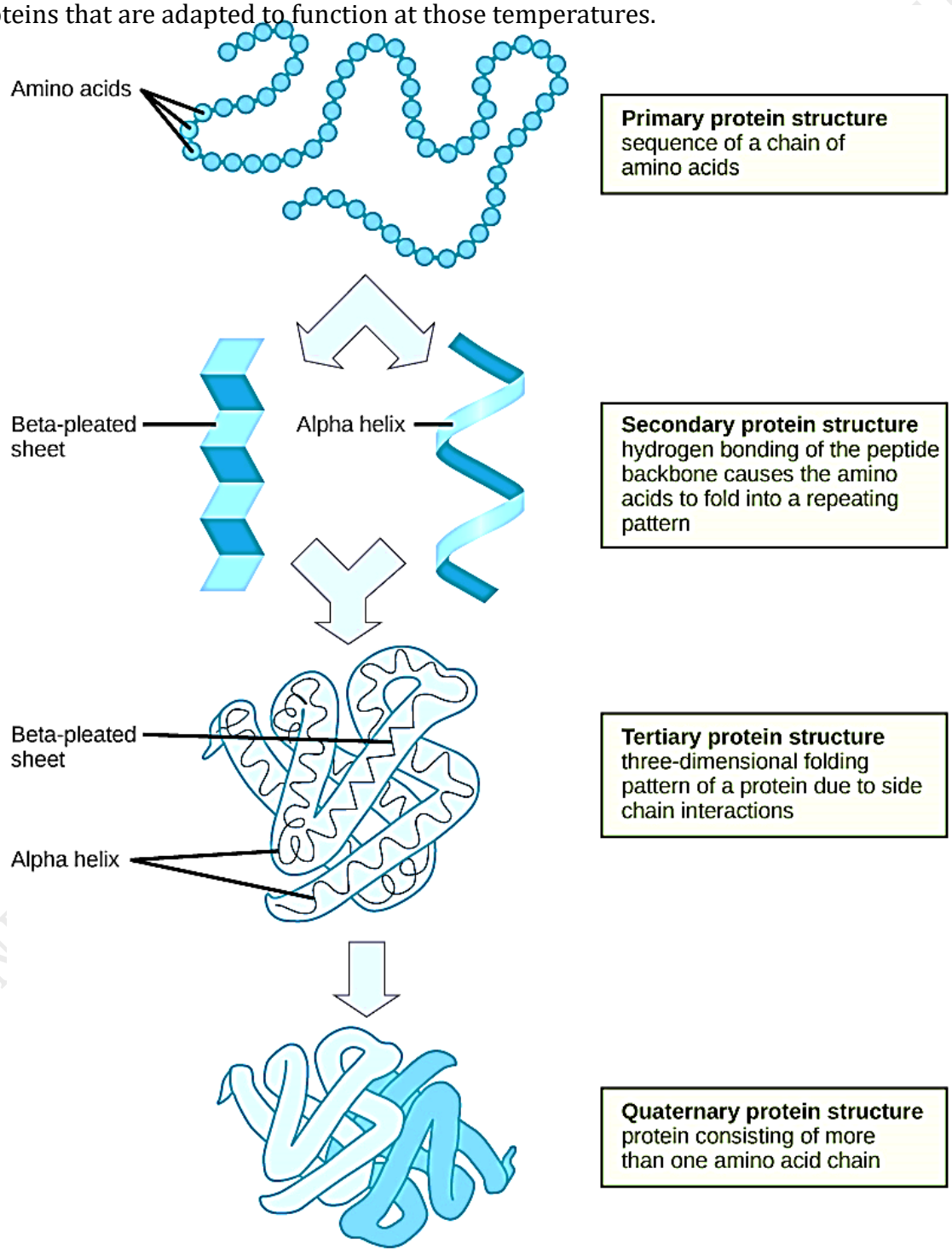
All of these interactions, weak and strong, determine the final three-dimensional shape of the protein. When a protein loses its three-dimensional shape, it may no longer be functional

Quaternary Structure

In nature, some proteins are formed from several polypeptides, also known as subunits, and the interaction of these subunits forms the quaternary structure. Weak interactions between the subunits help to stabilize the overall structure. For example, insulin (a globular protein) has a combination of hydrogen bonds and disulfide bonds that cause it to be mostly clumped into a ball shape. Insulin starts out as a single polypeptide and loses some internal sequences in the presence of post-translational modification after the formation of the disulfide linkages that hold the remaining chains together. Silk (a fibrous protein), however, has a β-pleated sheet structure that is the result of hydrogen bonding between different chain.

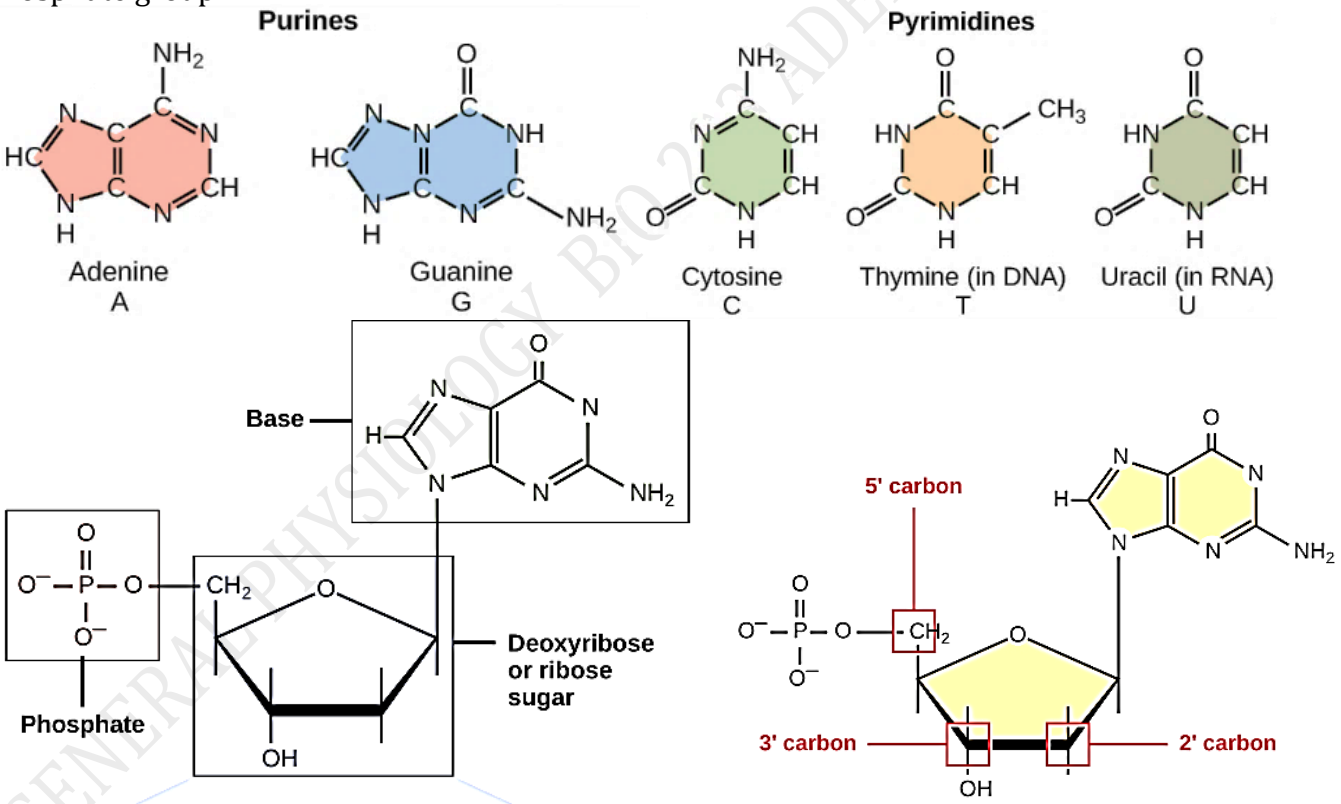
Each protein has its own unique sequence and shape held together by chemical interactions. If the protein is subject to changes in temperature, pH, or exposure to chemicals, the protein structure may change, losing its shape in what is known as denaturation as discussed earlier.

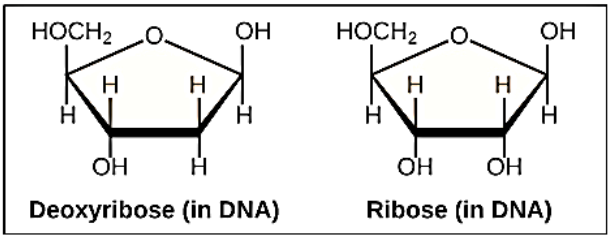
Denaturation is often reversible because the primary structure is preserved if the denaturing agent is removed, allowing the protein to resume its function. Sometimes denaturation is irreversible, leading to a loss of function. One example of protein denaturation can be seen when an egg is fried or boiled. The albumin protein in the liquid egg white is denatured when placed in a hot pan, changing from a clear substance to an opaque white substance. Not all proteins are denatured at high temperatures; for instance, bacteria that survive in hot springs have proteins that are adapted to function at those temperatures.



MODULE 5 NUCLEIC ACIDS – DNA AND RNA

Nucleic acids are key macromolecules in the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell. The two main types of **nucleic acids** are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals. The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus, but instead use an RNA intermediary to communicate with the rest of the cell. Other types of RNA are also involved in protein synthesis and its regulation. DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a polynucleotide, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to a phosphate group.

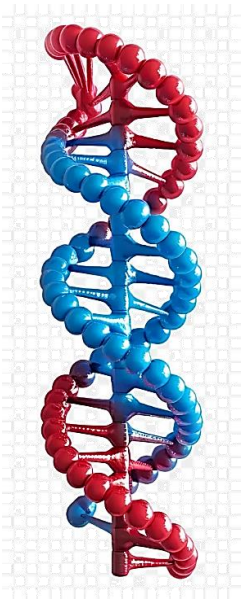


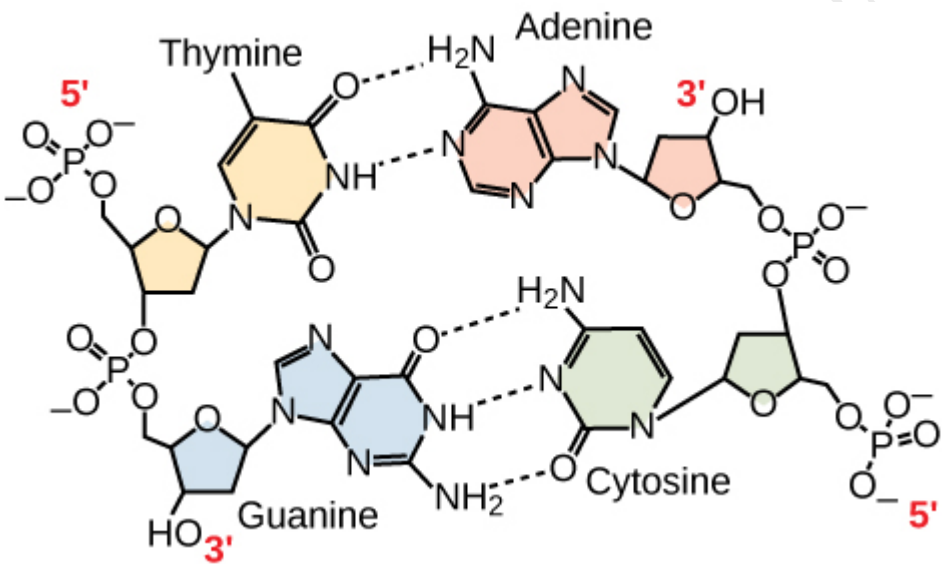


A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and a phosphate group.

DNA Double-Helical Structure

DNA has a double-helical structure. It is composed of two strands, or polymers, of nucleotides. The strands are formed with bonds between phosphate and sugar groups of adjacent nucleotides. The strands are bonded to each other at their bases with hydrogen bonds, and the strands coil about each other along their length, hence the "double helix" description, which means a double spiral.





The double-helix model shows DNA as two parallel strands of intertwining molecules. In a double stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. The phosphate backbone is located on the outside, and the bases are in the middle. Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine.

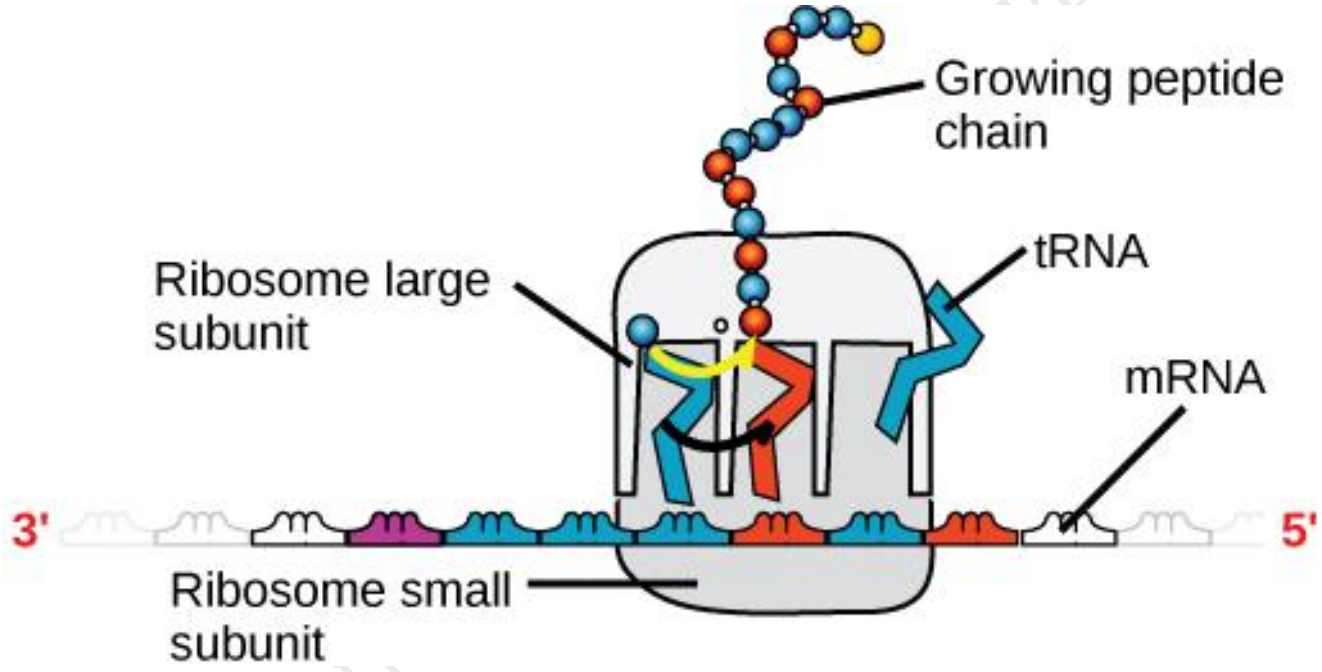
The alternating sugar and phosphate groups lie on the outside of each strand, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, and these bases pair; the pairs are bound to each other by hydrogen bonds. The bases pair in such a way that the distance between the backbones of the two strands is the same all along the molecule.

RNA (Ribonucleic acid)

Ribonucleic acid, or RNA, is mainly involved in the process of protein synthesis under the direction of DNA. RNA is usually single-stranded and is made of ribonucleotides that are linked

by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and the phosphate group.

There are four major types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and microRNA (miRNA). The first, mRNA, carries the message from DNA, which controls all of the cellular activities in a cell. If a cell requires a certain protein to be synthesized, the gene for this product is turned "on" and the messenger RNA is synthesized in the nucleus. The RNA base sequence is complementary to the coding sequence of the DNA from which it has been copied. However, in RNA, the base T is absent and U is present instead. If the DNA strand has a sequence AATTGCGC, the sequence of the complementary RNA is UUAACGCG. In the cytoplasm, the mRNA interacts with ribosomes and other cellular machinery



A ribosome has two parts: a large subunit and a small subunit. The mRNA sits in between the two subunits. A tRNA molecule recognizes a codon on the mRNA, binds to it by complementary base pairing, and adds the correct amino acid to the growing peptide chain.

The mRNA is read in sets of three bases known as codons. Each codon codes for a single amino acid. In this way, the mRNA is read and the protein product is made. Ribosomal RNA (rRNA) is a major constituent of ribosomes on which the mRNA binds. The rRNA ensures the proper alignment of the mRNA and the ribosomes; the rRNA of the ribosome also has an enzymatic activity (peptidyl transferase) and catalyzes the formation of the peptide bonds between two aligned amino acids. Transfer RNA (tRNA) is one of the smallest of the four types of RNA, usually 70–90 nucleotides long. It carries the correct amino acid to the site of protein synthesis.

It is the base pairing between the tRNA and mRNA that allows for the correct amino acid to be inserted in the polypeptide chain. microRNAs are the smallest RNA molecules and their role involves the regulation of gene expression by interfering with the expression of certain mRNA messages.

Features of DNA and RNA.

	DNA	RNA
Function	Carries genetic information	Involved in protein synthesis
Location	Remains in the nucleus	Leaves the nucleus
Structure	Double helix	Usually single-stranded
Sugar	Deoxyribose	Ribose
Pyrimidines	Cytosine, thymine	Cytosine, uracil
Purines	Adenine, guanine	Adenine, guanine

Even though the RNA is single stranded, most RNA types show extensive intramolecular base pairing between complementary sequences, creating a predictable three-dimensional structure essential for their function.

As you have learned, information flow in an organism takes place from DNA to RNA to protein. DNA dictates the structure of mRNA in a process known as transcription, and RNA dictates the structure of protein in a process known as translation. This is known as the Central Dogma of Life, which holds true for all organisms; however, exceptions to the rule occur in connection with viral infections.

Denaturation of protein:

Protein structures have evolved to function in particular cellular environments. Conditions different from those in the cell can result in protein structural changes, large and small. A loss of three-dimensional structure sufficient to cause loss of function is called denaturation. The denatured state does not necessarily equate with complete unfolding of the protein and randomization of conformation.

Most proteins can be denatured by heat, which affects the weak interactions in a protein (primarily hydrogen bonds) in a complex manner. If the temperature is increased slowly, a protein's conformation generally remains intact until an abrupt loss of structure (and function) occurs over a narrow temperature range. The abruptness of the change suggests that unfolding is a cooperative process: loss of structure in one part of the protein destabilizes other parts. The effects of heat on proteins are not readily predictable.

Proteins can be denatured not only by heat but by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea and guanidine hydrochloride, or by detergents. Each of these denaturing agents represents a relatively mild treatment in the sense that no covalent bonds in the polypeptide chain are broken. Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; extremes of pH alter the net charge on the protein, causing electrostatic repulsion and the disruption of some hydrogen bonding. The denatured states obtained with these various treatments need not be equivalent.

Plasma proteins:

Plasma is present in the blood and constitutes the liquid part of the blood. It is made up of 90% water which is required for hydration of body tissues. 7% of plasma is composed of proteins and looks like a pale yellow liquid. The total concentration of plasma protein in blood is 7-7.5 g/dl.

There are three main groups of plasma protein:

1. **Albumin:** It comprises of 60% of overall plasma protein and hence albumins are the highly abundant plasma proteins. These are produced by the liver and helps in transportation of different components in blood, along with drugs. Albumins also help to maintain water balance in the body and contribute to osmotic pressure.

2. **Fibrinogen:** These comprise of merely 4% of overall plasma proteins. This plasma protein is created by liver and its only function is to make clots and stop bleeding. It is very sticky and is a fibrous coagulant present in blood which produces thrombin. This thrombin then gets transformed into fibrin which is the major protein in blood clot.
3. **Globulin:** These comprise of 36% of overall plasma protein and include protein carriers, enzymes, gamma globulin and antibodies. Globulins are fractionized into alpha, beta and gamma and although all the globulins are manufactured in liver, gamma globulins are especially produced by lymphocytes also termed as plasma cells. Plasma globulins are categorized into four types based on its properties:
   i. **α1-Globulin:** This fraction includes several complex proteins containing carbohydrates and lipids. These are α1-glycoprotein and α-lipoproteins. The normal serum level of α1-globulin is 0.42 g/dl.
   ii. **α2-Globulins:** This fraction also contains complex proteins such as α2-glycoproteins, plasminogen, prothrombin, haptoglobulin, ceruloplasmin (transports Cu) and α2-macroglobulin. The normal serum value of this fraction is 0.67 g/dl.
   iii. **β-Globulins:** This fraction of plasma proteins contain different β-lipoproteins which are very rich in lipid content. It also contains transferrin which transports non-heme iron in plasma. The normal serum value of β-globulins is 0.91 g/dl.
   iv. **γ-Globulins:** These are also called as Immunoglobulins and possess antibody activity. On the basis of their electrophoretic mobility they are classified as IgG, IgA and IgM.

Plasma protein functions:

The functions of plasma proteins include:

1. Osmotic effect of plasma protein maintains fluid as well as electrolyte balance.
2. The plasma protein maintains viscosity of blood.
3. Performs the important function of clotting.
4. Responds with inflammation in case of wound or injury.
5. The gamma globulins act as antibodies and protect our body from infection.
6. Plasma protein also maintains acid base balance.

Amino acids sequences of proteins:

Protein sequencing denotes the process of finding the amino acid sequence, or primary structure of a protein. Sequencing plays a very vital role in Proteomics as the information obtained can be used to deduce function, structure, and location which in turn aids in identifying new or novel proteins as well as understanding of cellular processes. Better understanding of these processes allows for creation of drugs that target specific metabolic pathways.

The enzymes that achieve amino acids sequencing and break down proteins can be divided into exopeptidases and endopeptidases. Endopeptidases break peptide links (the bonds which hold amino acids together) in the middle of polypeptide chains. Exopeptidases break the peptide links of the amino acids at the ends of the chains.

It is often desirable to know the unordered amino acid composition of a protein prior to attempting to find the ordered sequence, as this knowledge can be used to facilitate the discovery of errors in the sequencing process or to distinguish between ambiguous results. Knowledge of the frequency of certain amino acids may also be used to choose which protease to use for digestion of the protein. A generalized method often referred to as amino acid analysis for determining amino acid frequency is as follows:

1. Hydrolyse a known quantity of protein into its constituent amino acids.
2. Separate and quantify the amino acids in some way.

Hydrolysis: Hydrolysis is done by heating a sample of the protein in 6 M hydrochloric acid to 100–110°C for 24 hours or longer.

Separation: The amino acids can be separated by ion-exchange chromatography or hydrophobic interaction chromatography.

Quantitative analysis: Once the amino acids have been separated, their respective quantities are determined by adding a reagent that will form a coloured derivative. If the amounts of amino acids are in excess of 10 nmol, ninhydrin can be used for this; it gives a yellow colour when reacted with proline, and a vivid purple with other amino acids. The concentration of amino acid is proportional to the absorbance of the resulting solution.

***N*-terminal Residue Identification:** *N*-terminal residue identification encompasses a technique which chemically determines which amino acid forms the *N*-terminus of a peptide chain. This information can be used to aid in ordering of individual peptide sequences that were generated using other sequencing techniques that fragment the peptide chain. The *N*-terminal residue identification can be achieved using the following reagents:

1- **Sanger reagent:** 1-fluoro-2,4-dinitrobenzene, or Sanger's reagent (commonly called dinitrofluorobenzene or DNFB) reacts with the amine group in amino acids to produce dinitrophenyl-amino acids. These DNP-amino acids are moderately stable under acid hydrolysis conditions that break peptide bonds. The DNP-amino acids can then be recovered, and the identity of those amino acids can be discovered through chromatography.

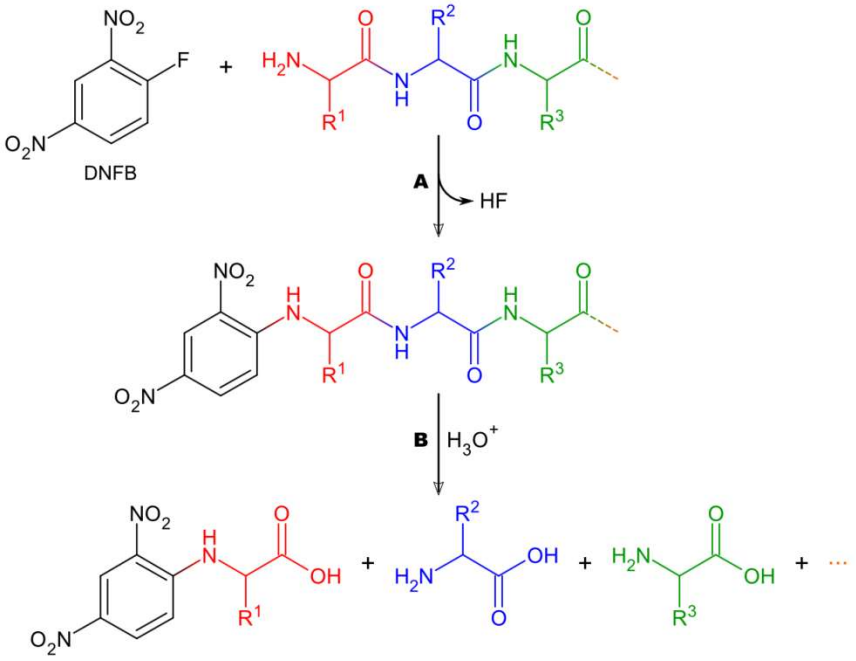


Figure 2: Sanger's method of peptide end-group analysis: A) complex of *N*-terminal end with Sanger's reagent (DNFB), B) total acid hydrolysis of the dinitrophenyl peptide.

2- **Dansyl chloride reagent:** Dansyl chloride or 5-(Dimethylamino) Naphthalene-1-Sulfonyl chloride is a reagent that reacts with the free amino groups of peptides and proteins to produce stable blue- or blue-green–fluorescent sulfonamide adducts.

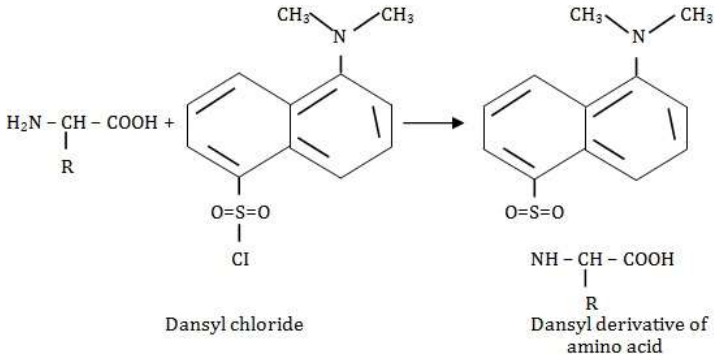


Figure 3: Dansyl chloride complex with amino acid.

3- **Edman reagent:** Edman degradation, developed by Pehr Edman, is a method of sequencing amino acids in a peptide. In this method, Phenyl isothiocyanate is reacted with an uncharged *N*-terminal amino group, under mildly alkaline conditions, to form a cyclical phenyl thiocarbamoyl derivative. Then, under acidic conditions, this derivative of the terminal amino acid is cleaved as a thiazolinone derivative. The thiazolinone amino acid is then selectively extracted into an organic solvent and treated with acid to form the more stable phenyl thiohydantoin (PTH)-amino acid derivative that can be identified using chromatography or electrophoresis.

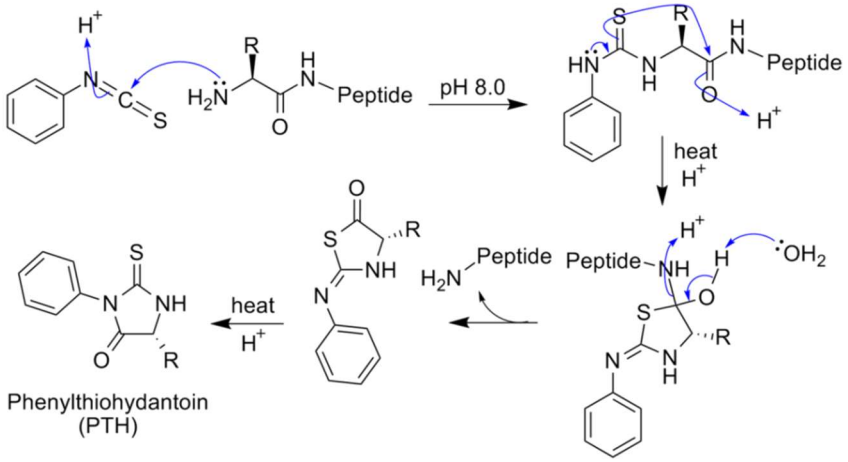


Figure 4: Formation of PTH from the reaction of Phenyl isothiocyanate with peptide.

***C*-terminal residue identification:** The number of methods available for *C*-terminal amino acid analysis is much smaller than the number of available methods of *N*-terminal analysis. The most common method is to add carboxypeptidases to a solution of the protein, take samples at regular intervals, and determine the terminal amino acid by analysing a plot of amino acid concentrations against time. *C*-terminal sequencing would greatly help in verifying the primary structures of proteins predicted from DNA.

Digestion into peptide fragments using endopeptidases: Peptides longer than about 50-70 amino acids long cannot be sequenced reliably by the Edman degradation. Because of this, long protein chains need to be broken up into small fragments that can then be sequenced individually. Digestion is done either by endopeptidases such as trypsin or pepsin or by chemical reagents such as cyanogen bromide. Different enzymes give different cleavage patterns, and the overlap between fragments can be used to construct an overall sequence. The most common endopeptidases are:

1- **Trypsin**: cuts after Arg or Lys, unless followed by Proline.
2- **Chymotrypsin**: cuts after Phe, Trp, or Tyr, unless followed by Pro.
3- **Pepsin:** cuts before Leu, Phe, Trp or Tyr, unless preceded by Pro.
4- **Cyanogen bromide (CNBr):** cuts after methionine residues.

Example: Unknown peptide was treated with trypsin first and again with CNBr. Giving the following sequences:

A) By trypsin (1) Asn-Thr-Trp-Met-Ile-Lys (2) Val-Leu-Gly-Met-Ser-Arg (3) Gly-Tyr-Met-Gln-Phe.
B) By CNBr (4) Gln-Phe (5) Val-Leu-Gly-Met (6) Ile-Lys-Gly-Tyr-Met (7) Ser-Arg-Asn-Thr-Trp-Met.

1- Deduce the sequence of the original peptide.
2- Give the sequence of the reaction of Phenyl isothiocyanate with peptide (5).
3- Give the products of peptide (1) if treated with chymotrypsin.