

Perspective of the Chemical Signature of Life: The Structure and Function of Proteins

Luka Atwok Opio Awata^{1, *}, Adelaide Pingdewinde Ouedraogo², Nathan Aliel Kachiguma³, Tighankoumi Junior Gmakouba⁴, Isaac Kumaga Asante⁵

¹Directorate of Research, Ministry of Agriculture and Food Security, Juba, South Sudan

²Environmental Institute of Agricultural Research (INERA), Ouagadougou, Burkina Faso

³Department of Agricultural Research Services (DARS), Lunyangwa Agriculture Research Station, Mzuzu, Malawi

⁴Department of Root and Tuber Crops, Togolese Agricultural Research Institute (ITRA), Lomé, Togo

⁵West Africa Center for Crop Improvement (WACCI), College of Basic and Applied Sciences, University of Ghana, Legon, Ghana

Email address:

lukatwok11@gmail.com (L. A. O Awata), aouedraogo@wacci.ug.edu.gh (A. P. Ouedraogo),

nkachiguma@wacci.ug.edu.gh (N. A. Kachiguma), tgmakouba@wacci.ug.edu.gh (T. J. Gmakouba),

asanteisaack57@gmail.com (I. K. Asante)

*Corresponding author

To cite this article:

Luka Atwok Opio Awata, Adelaide Pingdewinde Ouedraogo, Nathan Aliel Kachiguma, Tighankoumi Junior Gmakouba, Isaac Kumaga Asante. Perspective of the Chemical Signature of Life: The Structure and Function of Proteins. *American Journal of Chemical and Biochemical Engineering*. Vol. 4, No. 2, 2020, pp. 18-30. doi: 10.11648/j.ajcbe.20200402.11

Received: July 19, 2020; **Accepted:** August 3, 2020; **Published:** December 11, 2020

Abstract: Genes are the sources of information used for creating amino acids which are then assembled to form protein structures (molecules). Together, the various protein structures function in different catalytic and structural activities that are responsible for establishing the phenotypes we see. Although both gene and proteins are equally involved in the biological functions that determine phenotypes, considerable amount of time has been portioned, by geneticists and breeders alike, for dissection of gene architecture and its characteristics comparative to proteins. Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, they transport and store other molecules such as oxygen, they provide mechanical support and immune protection, they generate movement, they transmit nerve impulses, and they control growth and differentiation. Indeed, much of this text will focus on understanding what proteins do and how they perform these functions. Understanding protein structure and its functions is instrumental for advancing molecular sciences. This review attempts to shed some light on structure of proteins, relationship between amino acid sequence and DNA base sequence, hierarchical nature of protein structure as well as the relationship between protein structure and its functions. The information synthesized could provide an insight into the complex nature of proteins and its importance in the perception of biological sciences.

Keywords: Protein Structure, Amino Acid, Polypeptide, Gene

1. Introduction

Understanding protein structure and its functions is critical for enhancing breeding tools and molecular genetics. Enzymatic activity of each protein follows from its primary sequence of amino acids. Together, the various protein products of a cell undertake the catalytic and structural activities that are responsible for establishing its phenotype [1]. Of course, in addition to sequences that code for proteins,

DNA also contains certain sequences whose function is to be recognized by regulator molecules, usually proteins. Here, the function of the DNA is determined by its sequence directly, not via any intermediary code. Both types of region, genes expressed as proteins and sequences recognized as such, constitute genetic information [4, 2]. Each gene represents a particular polypeptide chain. The sequence of nucleotides in DNA is important not because of its structure per se, but because it codes for the sequence of amino acids that

constitutes the corresponding polypeptide (protein) [1]. Proteins can be divided into two broad classes with regard to localization: those that are not associated with membranes and those that are associated with membranes. Each class can further be subdivided, depending on whether the protein associates with a particular structure in the cytosol or type of membrane [4]. Here, we reviewed the structure of the basic unit of proteins (amino acid), relationship between amino acid sequence and DNA base sequence, hierarchical nature of protein structure, and the relationship between protein structure and its functions.

2. Amino Acid (The Basic Unit of Proteins)

2.1. Structure of Amino Acid

Protein is made up of units of amino acids. There are 20 common amino acids found in nature, each containing a unique side chain of particular chemical compound called *R*-group [1]. Side chains do not participate in polypeptide formation and are thus free to interact and react with their environments. Also bound to the main alpha (α) carbon of each amino acid are a carboxyl functional group, amino functional group and a hydrogen atom [2, 5, 6]. The α -carbon in an amino acid is considered to be asymmetrical because it is connected to four different atoms or group of atoms. Amino acid exists in two isomeric forms called *D*-amino acid (amine group is fitted to the dextro or right) and *L*-amino acid (amine group is fitter to the levo or left) (Figure 1). The *L*-amino acids are the only common type found in proteins of most living organisms and they play important roles as chemical signature of life [6, 7]. At neutral *pH* values inside cells, both functional groups are ionized i.e the carboxyl group loses a hydrogen ion and becomes negatively charged while the amino group gains one hydrogen ion and becomes positively charged [4, 8].

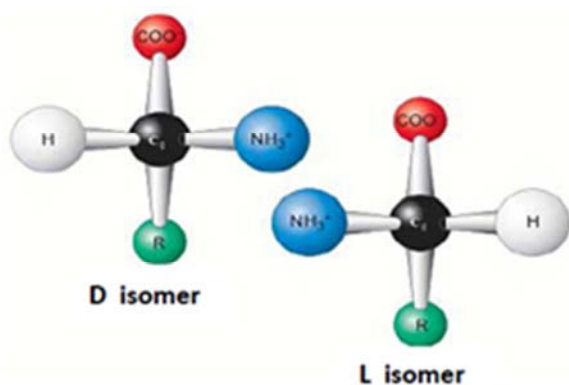
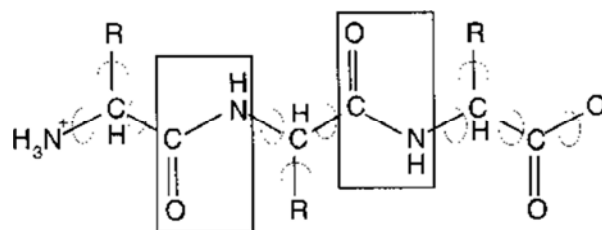


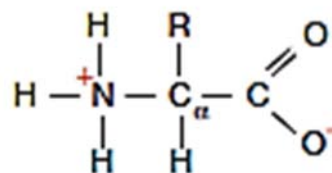
Figure 1. Two isomeric forms (*D* and *L* isomers) of amino acid [6].

The functional groups are important in determining the three-dimensional structure and function of protein molecules. They allow amino acid molecules to join to one another (polymerization) through the peptide bonds and form polypeptide or α -chain. A polypeptide is an assembly of amino acid molecules while protein is formed of

polypeptides [2, 8, 9]. The peptide bonds of the polypeptides are rigid planar units formed due to reaction between amino group of one amino acid and the carboxyl group of another. The peptide bond possesses no rotational freedom due to the partial double-bond character of the carbonyl-amino amide bond [7, 9]. However, rest of the bonds around the carbon atoms are true single bonds with considerable freedom of movement (Figure 2).



a. Amino acid molecule



b. Polypeptide chain

Figure 2. Amino acids linked by rigid peptide bonds to form polypeptide [5].

2.2. Classification of Amino Acid

Amino acids are grouped differently in various literatures with regards to their forms, functions and compositions, etc. For example, they can be classified based on position of “-NH₂”, nutritional requirement, composition of “-R” and on their metabolic fate [2, 5]. There are four common systems by which amino acids can be classified [8].

2.2.1. Classification Based on Polarity

1 Hydrophilic amino acids with electrically charged side chains

Five amino acids are under this class: *arginine*, *histidine* and *lysine* (with positive charges), and *aspartic acid* and *glutamic acid* (with negative charges). These amino acids, due to their electrically charged (+/-) side chains, are hydrophilic and highly attractive to water and positively charged ions (Figure 3) [5, 9].

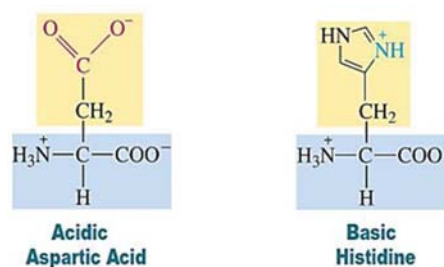


Figure 3. Some hydrophilic amino acids with electrically charged side chains [10].

2 Hydrophilic amino acids with polar and uncharged side chains

These are amino acids that contain relatively polar constituents that make them hydrophilic in character. They include *serine, threonine, asparagine, glutamine and tyrosine*. The amino acids are usually found in hydrophilic regions of a protein molecule, especially at or near the surface where they can be hydrated with the surrounding aqueous environment [8]. *Asparagine, threonine, and serine* are often post-translationally modified with carbohydrate in *N*-glycosidic (*asparagine*) and *O*-glycosidic linkages (*threonine* and *serine*) (Figure 4).

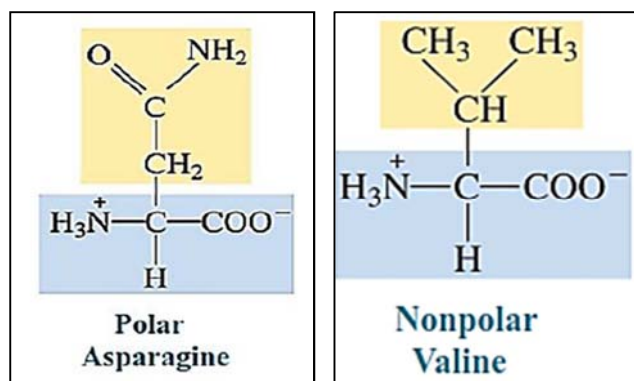


Figure 4. Asparagine (polar) and Valine (nonpolar) amino acids with uncharged side-chains [10].

Though these side chains are enzymatically derivatized in nature, the hydroxyl and amide portions have relatively the same nucleophilicity as that of water and are therefore, difficult to modify with common reagents under aqueous conditions [5, 9].

3 Hydrophobic amino acids with non-polar side chains

Seven amino acids fall under this category and their side chain are nonpolar hydrocarbons or very slightly modified hydrocarbons. In aqueous conditions of the cell, these hydrophobic side chains may cluster together in the interior of protein due to their hydrophobicity [4]. These amino acids include *alanine, isoleucine, leucine, methionine, phenylalanine, tryptophan and valine* (Figure 5).

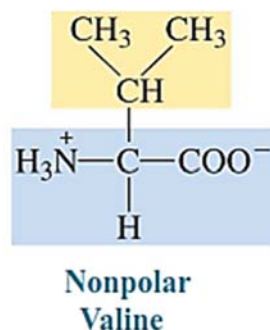


Figure 5. Valine nonpolar amino acid [10].

2.2.2. Classification Based on R-group

Depending on the side-chain (*R*-group) substituent, an amino acid can be classified into eight groups including

aromatic amino acids, aliphatic amino acids, sulphur amino acids, hydroxyl amino acids, cyclic amino acids, carboxyl amino acids, amine and amide amino acids [5] (Table 1). Amino acids that include an aromatic ring *phenylalanine, tryptophan*, and *tyrosine* with a benzene ring and *histidine* with an imidazole ring) are classified as aromatic amino acids.

Table 1. Classification of amino acids based on *R*-group.

SN	Group	Amino acid
1	Aromatic	Trp (W), Phe (F), Tyr (Y), His (H)
2	Aliphatic	Gly (G), Ala (A), Val (V), Leu (L), Ile (I)
3	Sulphur	Met (M), Cys ©
4	Hydroxyl	Ser (S), Thr (T), Tyr (Y)
5	Cyclic	Pro (P)
6	Carboxyl	Asp (D), Glu (E)
7	Amine	Lys (K), Arg (K)
8	Amide	Asn (N), Gln (Q)

2.2.3. Classification of Amino Acid Based on pH Level

Based on their side chains, the polar hydrophilic amino acids are classified into acidic (*glutamic* and *aspartic*), basic (*lysine, histidine* and *arginine*) and neutral (*tyrosine, serine, threonine, cysteine, glutamine* and *asparagine*) amino acids depending on their *pH* levels [5, 11].

2.2.4. Special cases Amino Acids

Three amino acids are considered in this class and they include *cysteine, glycine* and *proline*. Each of the three acids has special characteristics though their side chains are generally hydrophobic [5]. The *cysteine* side chain, with terminal $-SH$ group, can react with another *cysteine* side chain to form a covalent bond called a disulfide bridge ($-S-S-$) which is useful for polypeptide folding. However, when *cysteine* is not in a disulfide bridge, its side chain is hydrophobic. The *glycine* side chain consists of single hydrogen atom and is small enough to fit into tight corners in the interior of a protein molecule [8]. *Proline* however, differs from the rest because it has a modified amino group lacking hydrogen on its nitrogen and thus, limits its hydrogen-bonding activities. Also, the ring structure of *proline* limits rotations about carbon atom [12, 13]. Therefore, *proline* is usually found in bends or loops in protein structures.

2.2.5. Essential Amino Acids

Ten of the 20 natural occurring amino acids are considered essentials because they cannot be synthesized in animal body for which external supplementations are required. These include *tryptophan, histidine, arginine, leucine, isoleucine, lysine, valine, methionine, phenylalanine* and *threonine* [10].

2.3. Properties of Amino Acid

The side chains of amino acid confer different chemical, physical and structural properties to the final peptide or protein.

2.3.1. Acid-base Properties of Amino Acid

Amino acids can undergo an intramolecular acid-base reaction (proton transfer). The transfer of the H from the $-COOH$ group to the $-NH_2$ group forms a dipolar ion, an

ionized amino acid that has both (+) and (-) charges known as zwitterion or hybrid [5, 10]. Under acidic conditions (low pH), the zwitterion accepts proton on the basic -COO^- group while the -NH_3^+ group remains positively charged. In basic solution (high pH) however, the zwitterion loses proton from the acidic -NH_3^+ group while -COO^- group remains negatively charged (Figure 6).

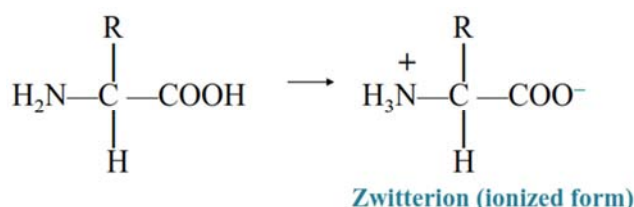


Figure 6. Acid-base properties of amino acid [10].

2.3.2. Peptide Bonds

A peptide bond links successive amino acids in a polypeptide chain and releases one molecule of water (Figure 7). The peptide bond possesses two extraordinarily important properties that facilitate folding of a polypeptide into a particular protein structure [14]. First, as a consequence of the partial double-bond character of the peptide bond

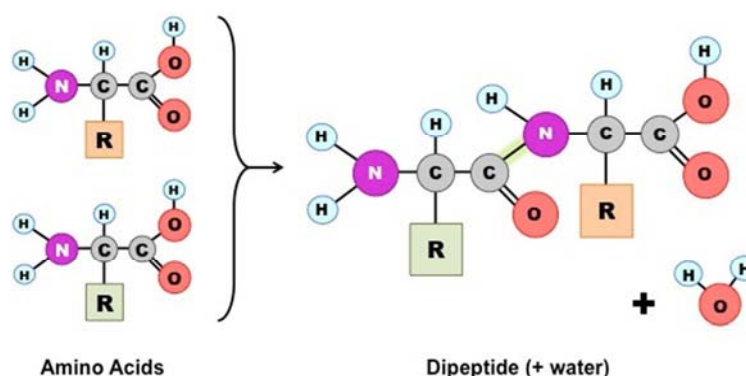


Figure 7. Peptide bonding of two molecules of amino acid (3).

2.3.3. Optical Properties of Amino Acid

Except for histidine, most aromatic amino acids produce reflection (fluoresce) when excited with ultraviolet light. For example, the intrinsic fluorescence emission of a folded protein is due to excitation of *tryptophan*, *tyrosine* and *phenylalanine* residues [5]. The optical properties of the aromatic amino acids such as fluorescence and second-harmonic generation are strongly influenced by the local environments [12].

2.3.4. Chirality of Amino Acids

Chirality is a property of asymmetry and is important in several branches of science. The word *chirality* is derived from the Greek word (*kheir*), "hand", a familiar chiral object [5]. An object or a system is chiral if it is distinguishable from its mirror image; that is, it can not be superimposed onto it (Figure 8).

between the carbonyl carbon and nitrogen, the unit of peptide bond links successive amino acids in a polypeptide chain. Secondly, many hydrogen bonds are formed in a polypeptide due to interactions between amide hydrogen and the carboxyl oxygen atoms of various amino acids in the protein [4, 10]. This provides stabilizing forces to the protein structure. Hydrogen bonds provide the energy required for formation of the three-dimensional structure of a protein. The bonds are formed between different AA depending on their positions [12]. Hydrogen bonds are generally weak however, they still collectively, provide strong force due to their great number within a protein structure.

Since the linkages connecting the amino acid residues consist of single bonds, each polypeptide is expected to undergo constant conformational changes caused by rotation around the single bonds. However, most polypeptides spontaneously fold into a single biologically active form [10]. Peptide bonds have partial double-bond characteristics and are thus, resonance hybrids. The bonds contain both rigid and planar (flat) structures [6]. The C—N bonds, joining each two AA, are shorter than other types of C—N bonds. Because of the rigidity of the peptide bond, almost one-third of the bonds in the backbone chain cannot rotate freely [6]. Consequently, this limits the number of conformations.



Figure 8. Chirality characteristics of amino acids (<https://www.seekpng.com/ipng/u2e6t4o0r5r5u2o0>).

Conversely, a mirror image of an *achiral* object, such as a sphere, cannot be distinguished from the object. Human

hands are perhaps the most universally recognized example of chirality: the left hand is a non-super-imposable mirror image of the right hand. The amino acids are all chiral, with the exception of glycine, because its side-chain is H.

2.3.5. Color (Buret) Reaction and 'Ninhydrin' Test of Amino Acid

Presence of AA can be determined through chemical reactions. For example, peptide containing two or more peptide bonds will react with Cu^{2+} in an alkaline solution to form violet blue colour (Buret test) [15]. Another simplest way to identify an amino acid is where the alpha amino group of amino acid reacts with the oxidizing agent *ninhydrin* (triketohydrindene hydrate). In presence of *ninhydrin*, amino acids undergo oxidative deamination and release ammonia, carbon dioxide, aldehyde and a reduced form of ninhydrin [6]. The ammonia liberated from the alpha amino group of amino acid reacts with the ninhydrin and its reduced product to form a blue substance *diketohydrin* or Ruhemann's purple (Ninhydrin test). However, when *proline* is used, a yellow color is produced because *proline* contains of α -imino group instead of α -amino group [6].

3. Relationships Between Amino Acid, DNA Sequence and Proteins

3.1. DNA Transcription

The process by which a gene gives rise to a protein is called gene expression. Each gene contains a continuous stretch of deoxyribonucleic acids (DNA) whose length is directly related to the number of amino acids in the protein it represents. Gene expression occurs by a two-stage process in eukaryotes: transcription and translation [16]. For simple eukaryotes e.g bacteria, both processes occur in the same single unit and mature mRNA is immediately produced. In higher eukaryotes however, transcription occurs in the nucleus where pre-mRNA is produced. The pre-mRNA is then transported into the cytoplasm where it is processed into mature mRNA, followed by translation into proteins [17]. Processing is a key step because majority of higher eukaryotes contain internal regions (introns) that do not code for proteins [1]. Therefore, processing (splicing) is required to remove the introns from the pre-mRNA so as to generate mature mRNA that has a continuous open reading frame.

The DNA strand that synthesizes mRNA via complementary base pairing is called the template strand or antisense strand (Figure 9). The other DNA strand bears the same sequence as the mRNA (except for possessing T instead of U), and is called the coding strand or sense strand [6]. Transcription generates several types of

single-stranded RNA. Three principal classes of RNA are involved in the synthesis of proteins: messenger RNA (*mRNA*), transfer RNA (*tRNA*) and ribosomal RNA (*rRNA*). An mRNA contains a series of codons that interact with the anti-sense codons of aminoacyl-tRNAs so that a corresponding series of amino acids are incorporated into polypeptide chains [8].

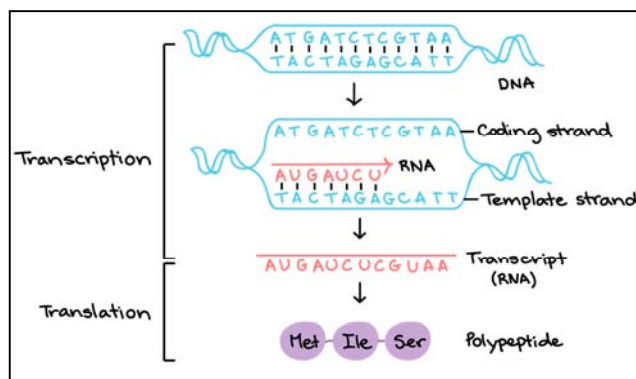


Figure 9. Transcription of DNA (www.khanacademy.org/science/biology/gene-expression).

3.2. Translation Process

Translation is the process by which the sequence of mRNA is read in triplets to give the series of amino acids which are used to synthesize the corresponding proteins. Translation converts the nucleotide sequence of mRNA into the sequence of amino acids comprising a protein [1]. The entire length of an mRNA is not translated, but each mRNA contains at least one coding region that is related to a protein sequence by the genetic code [18]. Each nucleotide triplet (codon) of the coding region represents one amino acid (Table 2).

The genetic code is a polynucleotide chain which contains four types of bases which are usually referred to by their initial letters. During transcription, DNA is transcribed into ribonucleic acid (RNA). DNA carries the genetic information, while the RNA is used to synthesize proteins [1]. DNA contains Adenine (A), Guanine (G), Cytosine (C), Thymine (T). These bases are collectively referred to as nucleotides [5]. Adenine and Guanine are purines are found in DNA and RNA while Cytosine and Thymine are pyrimidines and are found in DNA. In RNA however, the Thymine is replaced by Uracil. The only difference between Uracil and Thymine is the presence of a methyl substituent [1]. The DNA strand consists of nucleotide triplets (codons) meaning three nucleotides encode one amino acid ([19]. During translation, the strand is read in the direction from 5' to 3'. Each of 4 possible nucleotides can occupy each of the three possible positions of the codon, so that there are $4^3=64$ possible trinucleotide sequences [18]. Each of these codons has a specific meaning in protein synthesis.

Table 2. The genetic code and related amino acids [18].

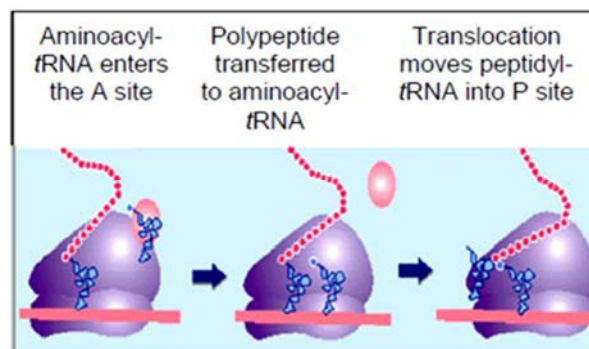
		Second letter											
		U			C			A			G		
First letter	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U	Third letter		
		UUC		UCC		UAC		UGC		C			
		UUA	Leu	UCA		UAA	Stop	UGA	Stop	A			
		UUG		UCG		UAG	Stop	UGG	Trp	G			
	C	CUU	CCU	CAU	Pro	His	CGU	U					
		CUC	CCC	CAC		CGC	Arg	C					
		CUA	CCA	CAA		CGA	A						
		CUG	CCG	CAG		Gln	CGG	G					
	A	AUU	ACU	AAU	Thr	Asn	AGU	Ser	U				
		AUC	ACC	AAC		AGC	C						
		AUA	ACA	AAA		AGA	A						
		AUG	Met	ACG		Lys	AGG	Arg	G				
	G	GUU	GCU	GAU	Ala	Asp	GGU	U					
		GUC	GCC	GAC		GCC	C						
		GUA	GCA	GAA		GGA	Gly	A					
		GUG	GCG	GAG		Glu	GGG	G					

The 64 codons represent amino acids and three (3) of these codons (UAA, UGA, and UAG) cause termination of protein synthesis (Figure 9). Because there are more codons (61) than there are amino acids (20), almost each amino acid is represented by more than one codon [1]. The only exceptions are methionine and tryptophan which are represented by single codons. Codons that have the same meaning are called synonyms and tend to be similar in sequence. Often the base in the third position of a codon is not significant, because the four codons differing only in the third base represent the same amino acid. The reduced specificity at the last position is known as third-base degeneracy [20]. The tendency for similar amino acids to be represented by related codons minimizes the effects of mutations. It increases the probability that a single random base change will result in no amino acid substitution or in one involving amino acids of similar character. For example, a mutation of CUC to CUG has no effect, since both codons represent leucine; and a mutation of CUU to AUU results in replacement of leucine with isoleucine, a closely related amino acid [21].

Translation of mRNA into proteins involves three processes: Initiation, elongation and termination. Initiation process begins where amino acids enter the protein synthesis pathway through the aminoacyl-tRNA synthetases, which provide the interface for connection with nucleic acid (Figure 10). The RNA polymerase binds to a sequence of DNA called the promoter, which is located near a gene. The mRNA is translated into a protein sequence by tRNA and rRNA. The interpretation of a codon requires base pairing with the anti codon of the corresponding aminoacyl-tRNA. The ribosome provides the environment for controlling the interaction between mRNA and aminoacyl-tRNA.

Elongation then follows where the RNA polymerase uses one stand of DNA (template DNA) and produces RNA molecule (transcript) that elongates from 5' to 3' direction. The RNA transcript differs from the non-template DNA (coding) strand only in that it carries uracil (U) instead of thymine (T).

The enzymes (synthetases) sort the tRNAs and amino acids into corresponding sets. Each synthetase recognizes a single amino acid and all the tRNAs that should be charged with it. Usually, each amino acid is represented by more than one tRNA. Translation is accomplished by a complex apparatus that includes both protein and RNA components. The translation process is done in the ribosome, a large complex that includes some large ribosomal RNAs (*rRNAs*) and many small proteins [22]. Lastly, the transcript is dislodged from RNA polymerase by terminators (sequences) and translation stops.

**Figure 10.** Translation process involving mRNA, ribosome and aminoacyl-tRNA (Source: [21]).

4. Hierarchical Nature of Protein Structure

Proteins are molecules formed of one or more polypeptides folded and coiled into different and complex three-dimensional conformations. The polypeptides are constructed from same set of 20 monomers (amino acids) connected in a specific sequences [7]. The way how protein performs its tasks is determined by its three-dimensional structure [3, 6]. Therefore, a protein which is not in its proper

shape or conformity cannot function efficiently. Polypeptides can be conformed or assembled into four major structures of protein and these include primary, secondary, tertiary and quaternary structures.

4.1. Primary Structure of Protein

The primary structure of protein is the linear arrangement of amino acids. After translation of mRNA, a polypeptide with defined sequence of amino acids is produced and this is called primary structure of protein [2, 6]. The primary structure may consist of 200 to 1000 amino acids joined by the peptide bonds between the carboxyl (COOH) group of one amino acid and the amide (NH₂) group of the other (Figure 11).

Repeated amide N, α carbon (C α), and carbonyl C atoms of amino acid residues form the backbone of a protein molecule on which various side-chain (R) groups are arranged [7, 9]. The sequence of a protein is conventionally written with its N-terminal amino acid on the left and its C-terminal amino acid on the right. Peptide bonds (yellow) link the amide nitrogen atom (blue) of one amino acid with the carbonyl carbon atom (gray) of an adjacent AA, resulting in linear polymers (polypeptides), depending on their length [6]. The side chains or R-groups (green), extending from the α -carbon atoms (black) are the major determinants of protein properties (Figure 11). At physiological pH values, the amino terminal is positively charged while the carboxyl terminal is negatively charged. In addition to peptide bonds, other types of covalent linkages such as disulphide bonds may also occur in the primary structure.

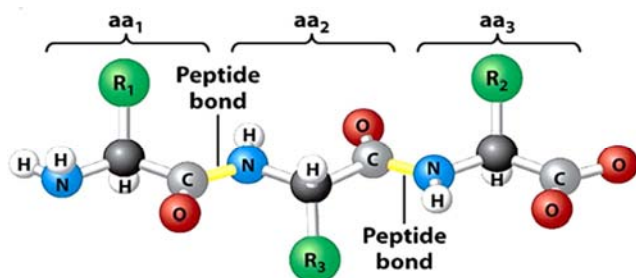


Figure 11. Primary structure of protein showing AA joint by peptide bonds [6].

4.2. Secondary Structure of Protein

To become a functional unit the primary structure, supported by the amino acid side chains and chaperones, quickly folds into a regular, repeated and compact three-dimensional shape called secondary structure [16, 23]. A single polypeptide may show different types of secondary structure depending on its sequence. The molecular forces responsible for the secondary structures are the non-covalent interactions between the various amino acids and the water molecules, normally referred to as hydrogen bonds [2, 9]. Therefore, if the stabilizing monovalent interactions are not present, the polypeptide will randomly fold to form a random coil structure. However, if stabilizing hydrogen bonds are formed between some amino acid molecules along the

polypeptide, parts of the polypeptide chain fold into well defined, periodic secondary structures known as α -helix, β -sheet and short U-shaped turns [4, 6].

4.2.1. Alpha (α) Helix Structure

One of the most common secondary structure patterns is called alpha (α) helix. The α -helix is stabilized by H-bonds between backbone carbonyl oxygen (C=O) and amide nitrogen (N-H) atoms that are oriented parallel to the helix axis. The carbonyl oxygen of one amino acid residue makes a hydrogen bond with the hydrogen atom of another amino group along the chain as discussed above. This regular pairing pulls the polypeptide backbone into a cylindrical spiral structure of 3.6 residues per turn, forming a coiled ribbon shape [6]. Right-handed and left-handed α -helices can exist though right-handed α -helices are the most common. Left-handed helices can be found in connective tissue such as collagen, and are also known to have unusual amino acid composition. Sometimes, α -helix coils around each other to produce a coiled-coil shape, a framework common for structural proteins such as nails and skin (Figure 12).

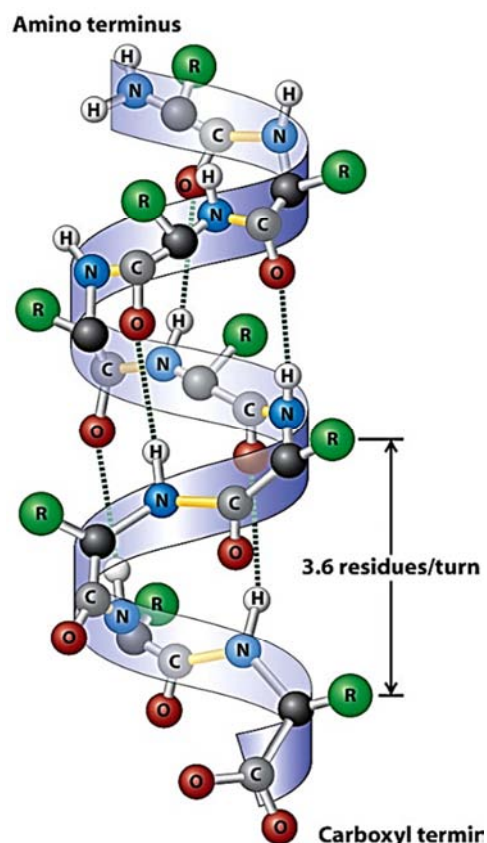


Figure 12. Alpha (α)-helix of the primary structure of protein [6].

The hydrophobicity and hydrophilicity of the α -helix is determined by the R-groups since the polar groups (amino and carboxyl) are already involved in the hydrogen bondings in the α -helix [9].

4.2.2. Beta (β) Sheet Structure

The β -sheet is another type of secondary structure

consisting of short (5-8 amino acids) polypeptide segments [9, 10]. The hydrogen bond interactions between backbone atoms in adjacent β -strands within same polypeptide chain or between different polypeptide chains lead to formation of a β -sheet. The planarity of the peptide bond allows β -sheet to become pleated (β -pleated sheet). The β -strands are directionality defined by orientation of the peptide bonds. Therefore, in a pleated sheet, adjacent β -strands are arranged in same (parallel) or opposite (anti parallel) to each other. In both arrangements, the side chains are projected from both sides of the sheet. The β -sheets can form the floor for binding activities in some proteins [8]. Also, hydrophobic core of other proteins contains numerous β sheets. The β -pleated sheets are predominant in the fibrous proteins. The fibroin protein of natural silk and spider webs contains β -pleated sheets as the main structural components and has high, tensile strength. Most β -strands in proteins are 5 to 8 amino acids long. The β -turns consist of 3-4 amino acids that form tight bends [10]. Glycine and proline are common β -turns. Longer connecting segments between β -strands are called loops. Alpha helix and β -sheet are the most predominant types accounting for about 60 percent of the polypeptide chain folding; the remainder of the molecule is just in random coils and turns. Thus, α -helices and β -sheets are the major internal supportive structures in proteins (Figure 13).

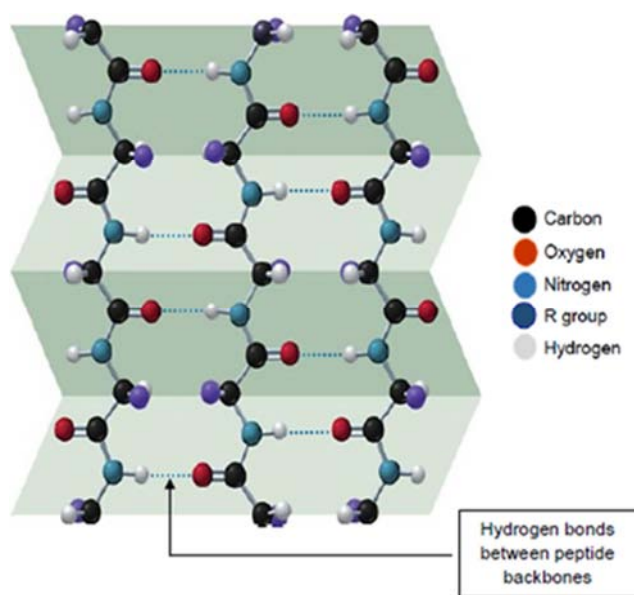


Figure 13. β -pleated sheet structure [10].

4.2.3. Beta (β) Turns Structure

Turns are few molecules (3-4) of amino acids located on the surface of protein and form sharp bends that redirect the polypeptide chain back towards the interior. These short U-shaped secondary structures are stabilized by hydrogen bond between their terminal amino acids. These types of structures are common in glycine and proline. Lack of large side chains in glycine and the presence of built-in bend allow the polypeptide backbone to fold into a tight U-shape (Figure 14). In case of longer polypeptides, the bends are also longer resulting into loops which, unlike turns, can be formed in

many different ways.

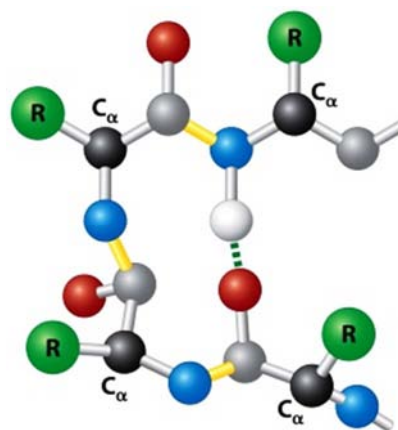


Figure 14. β -turns structure [6].

4.2.4. Random Coil Structure

The structure pattern consists of unspecified coils, loops, and sheets, which don't have any constant pattern. They are usually caused by a mixture of amino acid sequence, which cannot form either α -helix or β -sheet due to the noncovalent interactions within the polypeptide segments. Like α -helices and β -sheets, random coils are also found in most of the globular proteins.

4.2.5. Motifs of Secondly Structure

The variation in protein structure plays important roles in its function and regulation. The secondary structure elements can also undergo particular combinations to form motifs or folds such as helix-loop-helix, zinc-finger and coiled coil (amphipathic) which have importance in various protein functions. Motifs are considered to be evolutionarily conserved collections of secondary structure elements with defined conformations. This is because their structures and functions remain the same from protein to protein [4, 6]. Various motifs exist but in this review, emphases are made on the above three categories including the domains of tertiary structure (Figure 15).

Helix-loop-helix motif is a Ca^{2+} binding motif characterized by the presence of certain hydrophilic amino acids at invariant positions within the loop. Oxygen atoms in the invariant amino acids bind a Ca^{2+} ion through ionic bonds. The helix-loop-helix (EF hand motif) is commonly found in many calcium-binding proteins. Zinc-finger motif consists of α -helix and two β -strands arranged in anti parallel directions, which combined to form a finger-like bundle held together by a zinc ion, through reactions between the α -helix and the two β -sheets. Zinc-finger motif is commonly found in proteins that bind RNA and DNA. Coiled coil motif is common among fibrous proteins where proteins group themselves into oligomers. In these proteins, each polypeptide contains α -helical segments in which the hydrophobic amino acids are arranged in a regular pattern of repeated heptad sequence (repeating pattern of seven amino acids). In the heptad, a hydrophobic amino acid such as valine, alanine or methionine is fixed at position 1 while a leucine residue is fixed at position

4. Because hydrophilic side chains extend from one side of the helix and hydrophobic side chains extend from the opposite side, the overall helical structure becomes amphipathic (containing both hydrophobic and hydrophilic groups). The amphipathic character of these α -helices can allow 2-4 α -helices to wind around each other and form a coiled coil motif.

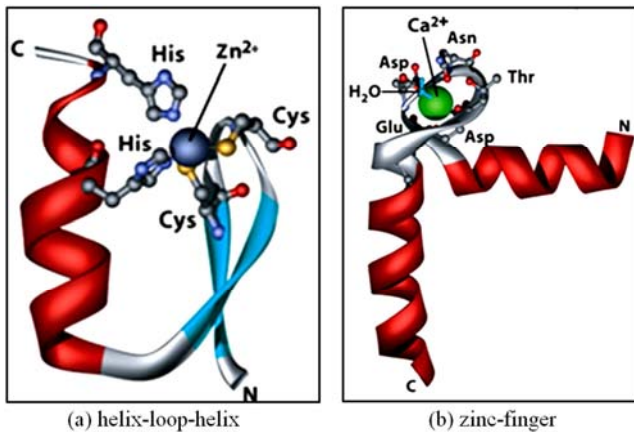


Figure 15. Some motif structures of protein [6].

4.3. Tertiary Structure of Protein

A combination of elements of secondary structure: α -helix, β -sheet, β -turn and random coil, due to the overall conformation, results into tertiary structure of protein. The tertiary structure is stabilized by hydrophobic interactions between the nonpolar side chains, hydrogen bonds between polar side chains, and peptide bonds [12, 13]. These forces are generally weak and therefore, tertiary structure of protein is not rigidly fixed and undergoes continual and minor fluctuation (Figure 16).



Figure 16. Tertiary structure of proteins is determined by hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages [6].

In tertiary structure, larger proteins (=15,000MW) are

subdivided into specific regions called domains. The domains are interconnected by segments of polypeptide chains and each contains a globular and a fibrous domains. The fibrous (structural) domain consists of 100-150 amino acids in most combinations of motifs [6, 8]. Domains can be characterized based of abundance of certain amino acids e.g proline-rich domain, an acidic domain, etc. They can also be grouped according to the functional terms for example, a specific region (s) of a protein may be responsible for its catalytic activities e.g a kinase domain (Figure 17). Like motifs of secondary structure, domains of tertiary structure are incorporated as modules into different proteins [6]. The modular approach to protein architecture is particularly easy to recognize in large proteins, which tend to be mosaics of different domains and thus can perform different functions simultaneously.

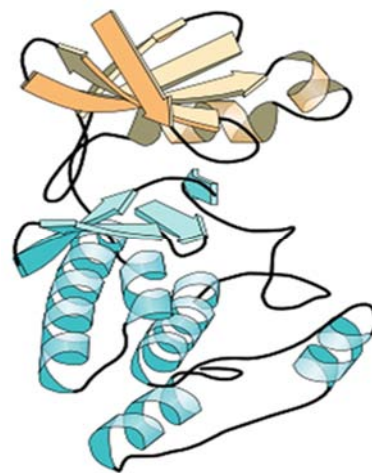


Figure 17. Kinases functional domain-the structure is divided into two globular domains (beige and blue) with the catalytic cleft in between [7].

4.4. Quaternary Structure of Protein

This structure contains several polypeptide chains (multimeric or subunits) which are important for regulation of their functions [12] (Figure 18). For example, hemagglutinin is a trimer of three identical subunits held together by noncovalent bonds. Various multimeric proteins can be formed by any number of identical or different subunits [2, 6].

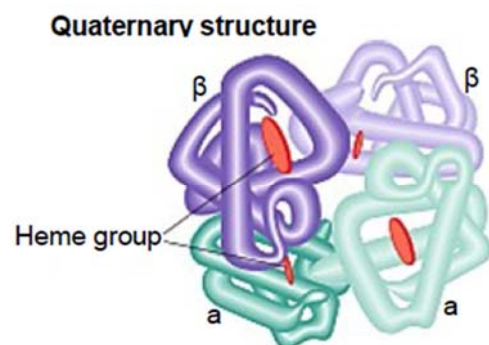


Figure 18. Quaternary structure of hemoglobin made up of four polypeptide subunits [10].

The subunits are assembled after reaching their tertiary states [12]. The main molecular interactions involved in the assembly of subunits and formation of quaternary structures include hydrophobic and electrostatic attractions, in addition to the weak Van der Waal's attractions [12]. The highest level of protein structure is where proteins are associated into larger (macromolecular) assemblies (10-100 polypeptide chains)

Examples of macromolecular assemblies include capsid that encases the viral genome, bundles of cytoskeletal filaments that support and give shape to plasma membrane, RNA polymerase, transcription factors, promoter-binding

protein, etc. Studies show that function derives from the three-dimensional structure is specified by amino acid sequence [6]. For example, sequencing of myoglobin and hemoglobin subunits revealed that many identical or chemically similar amino acids are found in identical positions throughout the primary structures of both proteins. This confirms the previous evidence of relationship between amino acid sequence, three-dimensional structure and function of proteins [6, 8]. The summary of protein structure is illustrated below (Figure 19).

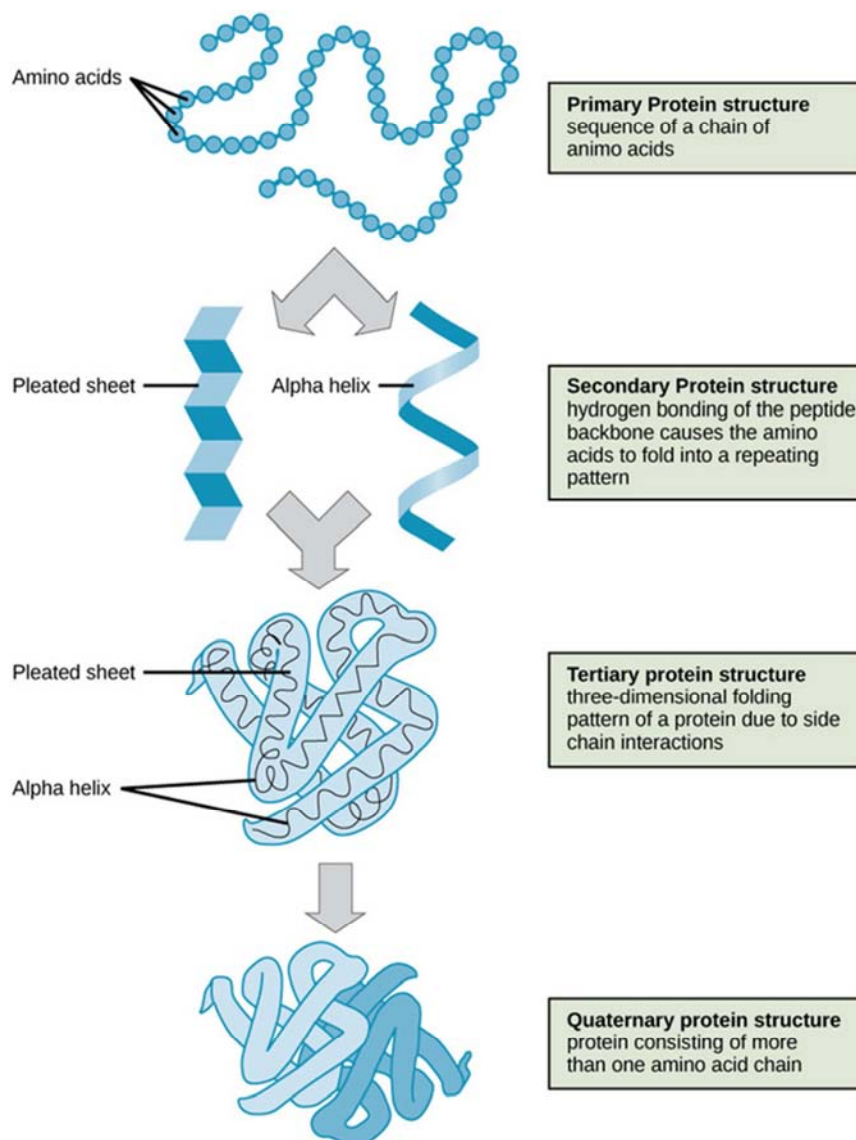


Figure 19. Four levels of protein structure (<https://courses.lumenlearning.com/introchem/chapter/protein-structure>).

5. Relationship Between Protein Structure and Its Functions

Proteins are the molecules in the cell that carry out the activities encoded by genes. They are the group of enzymes responsible for catalyzing a wide range of intracellular and

extracellular chemical reactions [24]. Therefore, proteins are grouped into specialized classes based on their importance. These include structural proteins, transport proteins, regulatory proteins, signaling proteins and moto proteins respectively (Figure 20). It is important to note that understanding the functional structure of protein is the recognition that most proteins have moving parts which are capable of transmitting various forces and energy in an orderly

manner [2, 8]. Though simple structure proteins are able to carry out various cellular activities, critical and complex cell

processes such as synthesis of nucleic acids and proteins are carried out by large macromolecular assemblies.

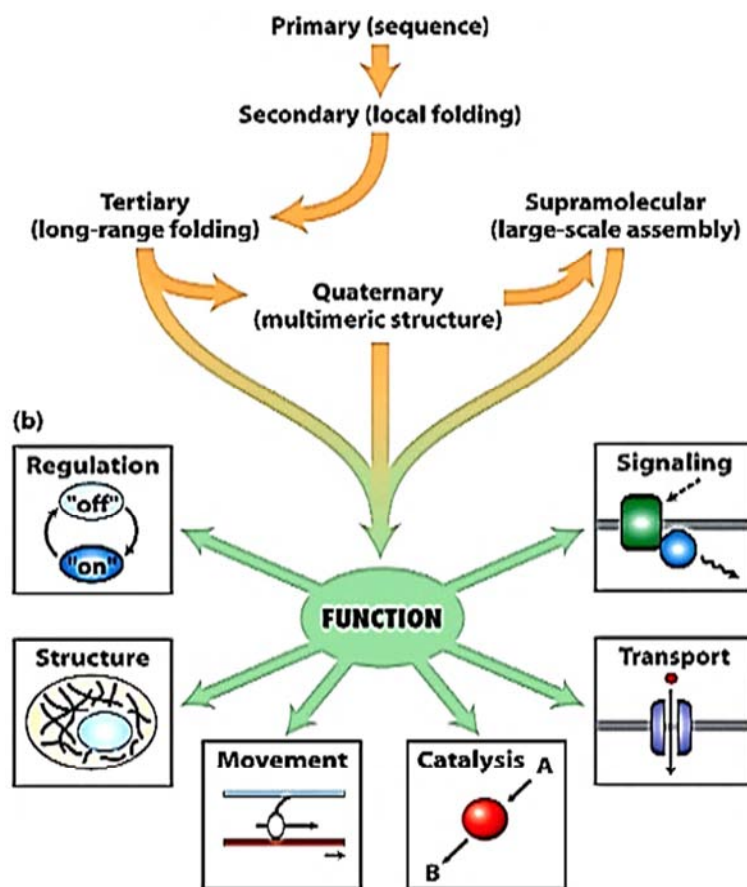


Figure 20. Molecular structure and functions of protein [6].

5.1. Binding Proteins

The most fundamental of these is binding, which underlies all the other biochemical functions of proteins. Enzymes must bind substrates, as well as cofactors that contribute to catalysis and regulatory molecules that either activate or inhibit them [6]. Structural proteins are, at their simplest, assemblages of single type of protein molecules bound together for strength or toughness. In more complex cases they bind to other types of molecules to form specialized structures such as the actin-based intestinal microvilli or the spectrin-based mesh that underlies the red blood cell membrane and helps maintain its integrity as the cells are swept round the body [8, 24]. Protein switches such as the GTPases depend on both binding and catalytic functions of proteins. Their switching properties rely fundamentally on the binding and the hydrolysis of GTP that they catalyze. They must also bind the molecules with which they interact when GTP is hydrolyzed, plus the regulatory molecules that activate GTP hydrolysis and that exchange GDP for GTP to enable the cycle to start again (Figure 21).

Functions of all proteins, whether signaling or transport or catalysis, depend on its ability to bind other molecules or ligands [10]. The bound ligand could be a small molecule or a macromolecule, and binding is usually very specific.

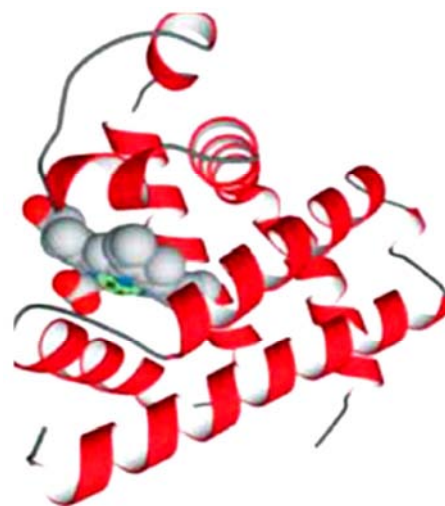


Figure 21. Myoglobin binds oxygen molecule to the iron atom in its heme group (shown in grey with the iron in green) [6].

Ligand binding is enhanced by noncovalent interactions between ligand and protein surface [8]. These are the same types of bonds that are involved in stabilizing folded proteins as well as in interactions between protein subunits. Specificity arises from the complementarity of shape and charge

distribution between the ligand and its binding site on the protein surface, and from the distribution of donors and acceptors of hydrogen bonds. Changes in the conformation of a protein may accompany binding or a possibility of binding to occur. Alternatively, even a small change in the structure of ligand or protein can disrupt binding [12].

Specific recognition such as TATA binding of other molecules is central to protein functions. The molecule that is bound (the ligand) can be as small as the oxygen molecule that coordinates activities with the heme-group of myoglobin, or as large as the specific DNA sequence (TATA box) which is bound and distorted by the TATA binding protein [19]. Specific binding is governed by shape complementarity and polar interactions such as hydrogen bonding. The TATA binding protein binds a specific DNA sequence and serves as a platform for a complex that initiates transcription of genetic information [6].

5.2. Catalysis Proteins

Almost all chemical reactions in a living cell are catalyzed by protein enzymes [10, 24]. The catalytic efficiency of enzymes is remarkable and reactions can be accelerated by as much as 17% over the simple buffer catalysis [2]. Many structural features contribute to the catalytic power of enzymes. For example, holding reacting groups together in an orientation favorable for reaction (proximity), binding the transition state of the reaction more tightly than ground state complexes (transition state stabilization), acid-base catalysis, and so on (Figure 22).

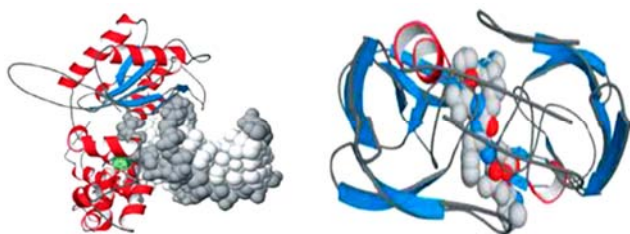


Figure 22. Catalytic proteins [10].

5.3. Switching Proteins

Proteins are flexible molecules and their conformation can change in response to changes in *pH* or ligand binding. Such changes can be used as molecular switches to control cellular processes [5]. One example, which is critically important for the molecular basis of many cancers, is the conformational change that occurs in the small GTPase when GTP is hydrolyzed to GDP [8]. The GTP-bound conformation is an "on" state that signals cell growth while the GDP-bound structure is the "off" signal.

5.4. Structural Proteins

Protein molecules are the major structural elements of living systems. As a result, protein can perform various functions depending on specific association of the protein subunits with themselves as well as with other proteins,

carbohydrates, etc [6]. This enables even the complex systems like actin fibrils to assemble spontaneously (Figure 23).



Figure 23. Silk and actin structures are important for structural functions of protein (a) silk structure (b) actin structure [7].

Structural proteins are also important sources of biomaterials such as silk, collagen, and keratin [2, 24]. Proteins are the most versatile macromolecules of the cell and therefore, "protein function" may mean the biochemical function of the molecule in isolation; or the cellular function it performs as part of an assemblage or complex with other molecules; or the phenotype it produces in the cell or organism. Major examples of the biochemical functions of proteins include binding, catalysis, operating as molecular switches, and serving as structural components of cells and organisms [6].

Proteins may bind to other macromolecules, such as DNA of polymerases or gene regulatory proteins, or to proteins in the case of a transporter or a receptor that binds a signaling molecule. This function exploits the ability of proteins to present structurally and chemically diverse surfaces that can interact with other molecules with high specificity [24]. Catalysis requires not only specific binding to substrates and in some cases to regulatory molecules, but also specific chemical reactivity [2]. Structural proteins may be as strong as silk or as tough and durable as keratin, the protein component of hair, horn and feathers; or they may have complex dynamic properties that depend on nucleotide hydrolysis [2, 6]. Transport is another important role of structural protein where some proteins carry various substances such as oxygen, ions, and so on, to the target sites. Likewise, hormones are vital in information transfer [8].

Another important aspect of protein structure is the relationship between the amino acid sequence, the three-dimensional structure, and function of proteins. This allows the prediction of structure and function of proteins, based on comparison with the sequences of known protein [23]. As a result, sequence comparison has expanded substantially in recent years as genomes of more and more organisms have been sequenced [25]. In addition, similarities and differences in amino acid sequences of proteins are used in the biological classification (homology). For example, proteins that have a common ancestor are referred to as homologs. The main evidence for homology among proteins, and hence their common ancestry, is the similarity in their sequences or structures [25].

Thus, homologous proteins are considered to belong to a "family" and their lineage can be traced for the comparisons of their sequences. The folded three-dimensional structures of homologous proteins are similar even if parts of their primary

structure show little evidence of homology [23]. The relatedness among homologous proteins is most easily visualized by a tree diagram based on sequence analyses. For example, the amino acid sequences of globins from bacteria, plants, and animals suggest that they evolved from an ancestral monomeric, oxygen-binding protein [8]. With the passage of time, the gene for this ancestral protein slowly changed, initially diverging into lineages leading to animal and plant globins [6]. Subsequent changes gave rise to myoglobin, a monomeric oxygen-storing protein in muscle, and to the α - and β -subunits of the tetrameric hemoglobin molecule of the circulatory system [26].

6. Conclusion

A functional protein consists of one or more polypeptides that have been precisely twisted, folded, and coiled into a unique conformation. It is the order of amino acid that determines what the three-dimensional conformation will be. In almost every case, the function of protein depends on its ability to recognize and bind to some other molecule. For example, antibodies bind to particular foreign substances that fit their binding sites. Enzymes recognize and bind to specific substrates, facilitating a chemical reaction. Neurotransmitters pass signals from one cell to another by binding to receptor sites on proteins in the membrane of the receiving cell. Much has been done to understand the structure and functions of proteins however, the function of proteins at molecular level remains a challenge. Development and implementation of reliable computational methods can provide good options for better predictions of function for proteins of unknown function.

References

- [1] Shu J. J., A new integrated symmetrical table for genetic codes, BioSystems, 2017. 151: p. 21-26.
- [2] Pierce, P., Genetics: A Conceptual Approach, 2nd Editio, H. W Freeman and Company, New York, USA, 2006.
- [3] Griffiths, A., et al., Introduction to Genetic Analysis, 8th Ed., W. Freeman and Company, New York, USA, 2005.
- [4] Hermanson, G., Bioconjugate Techniques, Secon Ed., Academic Press, Illinois, USA, 2008.
- [5] Lodish, H., et al., Molecular Cell Biology, 5th Editio, W. H. Freeman and Company, New York, USA, 2004.
- [6] Weaver, R. F., Molecular Biology, 2nd Editio, The McGraw-Hill Companies, Kansas, USA, 2004.
- [7] Sadava, D., et al., Life: The Science of Biology, 8th Editio, W. H. Freeman, Sinauer, USA, 2008.
- [8] Schleif, R., Genetics and Molecular Biology, 2nd Editio, The Johns Hopkins University Press Baltimore, London, UK, 1993.
- [9] Jeff, H., B. Gregory, J. Lewis, Becker's World of the Cell, Pearson Education, Inc., New Jersey, USA, 2018.
- [10] Nieto, C., et al., An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon, 2006. p. 452-462.
- [11] Janin, J., R. P. Bahadur, P. Chakrabarti, Protein-protein interaction and quaternary structure, Q. Rev. Biophys, 2008. 41: p. 133-180.
- [12] Dimmock, N. J., A. J. Easton, K. N. Leppard, Introduction to Modern Virology, Sixth Edit, Blackwell Publishing Ltd., Malden, MA, USA, 2007. doi: 10.1007/s13398-014-0173-7.2.
- [13] Pandey, G., et al., Computational Approaches for Protein Function Prediction: A Survey, Access. (2006). doi: 10.1.1.120.385.
- [14] Silver, M., J. Denburg, and J. Steffens, Spectrophotometric Deteemination of the Kietics of the Pepsin-Catalyzed Hybrolsis of Certain Dipeptide Substrates, J. Am. Chem. Soc., 1965. 87: p. 882-886.
- [15] Yura, K., and M. Go, Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles, BMC Plant Biol., 2008. 8: p. 1-11.
- [16] Soll, D., and U. RajBhandary, tRNA Structure, Biosynthesis, and Function, 1st Editio, American Society for Microbiology, Washington, DC., 1995.
- [17] Nirenberg, M. W., and P. Leder, The effect of trinucleotides upon the binding of sRNA to ribosomes, Science, 1964. 145: p. 1399-1407.
- [18] Cann, A., Principles of molecular virology, 4th Editio, Elsevier Academic Press, 2005.
- [19] Rould, A., Structure of E. coli glutamyl-tRNA synthetase complex with tRNAGln and ATP at 2.8 Å resolution, Science, 1989. 245: p. 1135-1142.
- [20] Crick, F. h. C., Codon-anticodon pairing: the wobble hypothesis, J. Mol. Biol., 1966. 19: p. 548-555.
- [21] Schimmel, P., Parameters for the molecular recognition of tRNAs, Biochemiistry, 1989. 28: p. 2747-2759.
- [22] Hvidsten, T. R., et al., A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity, PLoS One, 2009. 4.
- [23] Lucas, W. J., Plant viral movement proteins: Agents for cell-to-cell trafficking of viral genomes, 2006. 344: p. 169-184.
- [24] Y. He, Y., et al., Comparisons of structurally homologous CheY-like proteins, PNAS, 2017. 114.
- [25] Carter, J., and V. Saunders, Virology: Principles and applications, John Wiley and Sons, Liverpool, UK, 2007.