Review Article

PROTEINS: THE INGENIOUS WORKHORSE MOLECULES, ADETAILED OVERVIEW.

Abstract

Proteins are the most common macromolecules in humans.Proteins are formed from amino acid units following the template of the genetic code in the protein coding regions of the deoxyribonucleic acid (DNA).It has protean structure and function throughout the human body.

The amino acid sequence of a protein constitutes its fundamentalor(primary) structure. Proteins' secondary, tertiary, and quaternary structures are determined by their basic structure, which represents non-covalent protein folds containing domains and motifs. Protein post-translational modification adds chemical groups to one or more amino acid units, diversifying protein structure and activities.These dictate protein characteristics and activities. Protein structure is studied via x-ray crystallography and magnetic resonance imaging. There are various bioinformatics websites and online databases that focus on proteins, protein domains, motifs, and folds, as well as their identification, annotation, classification, and structural alignment.This paper presents amino acid composition, structure, classification, physicochemical properties, domains and folds and motifs of proteins. It also explains the synthesis, post translational modification and other related characteristics of proteins. Research into proteins has aided the understanding of the etiology,diagnosis and treatment of diseases in humans such as cancer, neurodegenerative and metabolic diseases. It has also led to the discovery of drug targets and disease biomarkers.This overview aims to increase these readers' understandingof proteins by deconstructing some of the most common protein-related subjects and titles.It has provided both classical and contemporary knowledge on the subject. This shall be of immense benefits to early career researchers and both undergraduate and post graduate students of molecular biology, biochemistry other life sciences and even non-life science disciplines.

Keywords Proteins, amino acids, protein structure, protein synthesis, post translational modification of proteins.

1.INTRODUCTION

Proteins are found throughout the human body. They are the most common macromolecules in humans, accounting for more than half of the dry weight of cells (Cozzone, 2002). They are involved in almost every aspect of cell function. They are necessary for survival (W.Rose, 2019). In contrast to carbohydrates (stored as glycogen) and lipids (stored as lipid droplets in adipose tissue), proteins lack specific storage depots in the human body (A.Rose, 2019). This is why the body conserves proteins during times of food deficit, such as fasting (Cahill, 2006), by delaying or stopping tissue amino acid use (Carroll, 2015; A.Rose and Richter, 2009; Millward and Garlick, 1972). However, the amino groups of specific amino acids are still used in various metabolic reactions and must be replenished by dietary intake (Reed and Biolo, 2002). Since 1912, amino acids have been divided into nutritionally essential (indispensable) and nutritionally non-essential (dispensable) categories based on growth and nitrogen balance (Hou et al, 2015).

Nutritionally essential amino acids must be obtained through diet since their carbon skeletons cannot be produced de novo in human cells or are insufficiently generated to meet human metabolic needs. (Lopez and Mohuiddin, 2022). The amino acids are phenylalanine, valine, tryptophan, threonine, methionine, histidine, isoleucine, leucine, and lysine. A complete protein includes all of the essential amino acids. Non-essential (dispensable) amino acids can be produced de novo in human and animal cells and may not be present in the diet. They are alanine, asparagine, aspartic acid, glutamic acid, and serine.

However, Hou et al. (2015) believe that non-essential amino acids must be included in the diet for optimal growth, development, and metabolic outcomes. Furthermore, there is a subset of amino acids known as conditionally essential amino acids because they can be synthesized by human or animal cells in healthy physiological states but not in unhealthy physiological states such as starvation or in-born errors of metabolism, where they must be obtained through diet. They comprise arginine, cysteine, glutamine, glycine, proline, and tyrosine.

The central dogma of molecular biology holds that genes are transcribed into messenger RNA, which is then translated into proteins that mediate metabolic networks in creatures including humans. Metabolic networks regulate various aspects of life, including cellular electron transfer in oxidative phosphorylation, muscle tissue contraction, and hearing at the organ level (Salem, 2013). Proteins are complex organic macromolecules made up of amino acid units. Animal protein is made up of 20 naturally occurring amino acid units.

The amino acid units or residues are connected together by condensation between the amino group of one amino acid and the carboxyl group of the next amino acid unit, resulting in the loss of a water molecule to form peptide bonds that connect the amino

acid units in a linear pattern. Oligopeptides have fewer connected amino acid units than polypeptides. Proteins are polypeptides with molecular weights more than 5000 Daltons (Sumner 1988). Jons Jacob Berzelius, a Swedish scientist, created the term protein in 1838 from the Greek 'proteios', which means 'holding first place' (Katz, 2021). Proteins differ in amino acid quantities and sequences, making them species and organ specific. The proteome denotes all the proteins that are produced by an organism or a person while the genome denotes all the genes that exist in the organism or person. Proteomics systematically studies these proteins which are not static but vary with time Pharmacoproteomics is an exploitation of the great potentials of proteomics even beyond pharmacogenomics in precision medicine to develop and design pharmacotherapeutics that target a specific phenotype. Single nucleotide polymorphism may yield varying mRNAs that in turn yield varying amino acids and proteins as the phenotype (Jain, 2016).

2. AMINO ACID COMPOSITION OF PROTEINS

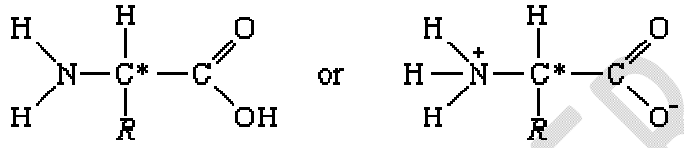


Figure 1: Amino acid Structure(Sumner 1988)

Amino acids are the building components of proteins. Each amino acid has an amino group, a carboxyl group, and a side chain that extends from the backbone of the polypeptide chain. The attached side chain determines the protein's characteristics. They have different sizes, shapes, polarities, charges, and hydrophobicities. Glycine, alanine, valine, leucine, isoleucine, and methionine all have neutral aliphatic side chains, while phenylalanine, tyrosine, and tryptophan have aromatic side chains. At neutral pH, basic side chains include lysine, arginine, and histidine; acidic side chains include glutamic acid and aspartic acid; and polar uncharged side chains include serine, threonine, glutamine, proline, and cysteine.

Cysteine has sulfur-containing side chains, which allow disulphide bonds to be formed between polypeptide chain segments or chains. Basic proteins include more basic amino acid side chains than acidic proteins. Proteins, on the other hand, are typically amphoteric, which means they can function as acids or bases due to the presence of both acidic and basic amino acid side chains. Protein ionization is pH-dependent. The amino group is positively charged at low pH, whereas the carboxyl group is unionized. At high pH, the carboxyl group is negatively charged, whereas the amino group is unionized. pH affects protein staining because it influences protein ionization.The isoelectric point is the pH at which positive and negative charges in proteins are balanced(Sumner,1988).

Selenocysteine and pyrrolysine are newly discovered naturally occurring amino acids

found in some organisms. Some amino acids are formed by modifying one of the twenty amino acids and are rarely encountered in proteins. These include hydroxylysine and hydroxyproline in collagen, the neurotransmitter gamma aminobutyric acid derived from glutamic acid decarboxylation, the neurotransmitter serotonin formed from tryptophan, gamma carboxyglutamic acid in calcium binding proteins, histamine derived from histidine decarboxylation, and epinephrine formed from tyrosine. (Garret and Grisham, 2010)

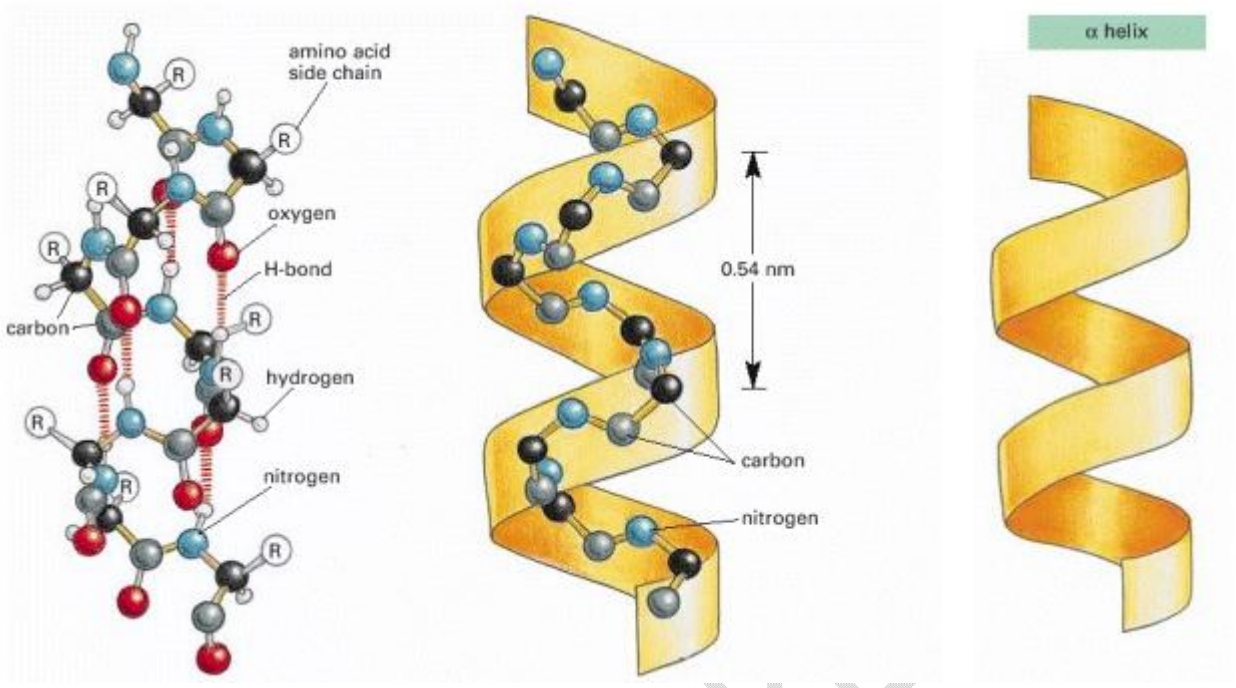
3.PROTEIN STRUCTURE

There exist a traditional “sequence – structure – function” model in protein discuss which implies that the sequence of amino acids in proteins determine their structure which in turn dictates their functions.(He et al, 2009, Wright & Dyson, 1999). This is captured in the central dogma of protein structural biology which has it that the amino acid sequence of a protein has all the information required for a protein to fold into a three-dimensional structure under favorable physiological or experimental conditions and that this structure plays a significant role in specifying the function of the protein (Li et al, 2014). The high-resolution X-ray crystallography is used to study the structure of proteins as it produces static protein structures (Henzler-Wildman & Kern, 2007). Nuclear magnetic resonance (NMR) is also usedfor this (Sun et al, 2004). However, proteins are dynamic molecules and their physical motions within cells determine their biological functions (Yang, 2014). These physical motions include the protein folding process itself that gave rise to the original folded state, the protein dynamics and molecular motions that occurs in the folded conformation and the dynamics of protein ligand interaction(Li et al, 2014). Protein dynamics may be studied by x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, Laue x-ray diffraction, cryo-electron microscopy, small angle x-ray scattering, circular dichroism (CD), fluorescence resonance energy transfer (FRET), infrared spectroscopy, Raman spectroscopy, electron paramagnetic resonance and computational approaches like molecular dynamics (MD) simulation and their derivatives like simulated annealing, replica exchange Monte Carlo technique, force field simplification models eg Gaussian network models, GO models and enhanced sampling techniques like metadynamics, umbrella sampling, targeted and steered simulations (Li et al, 2014).

Proteinstructure is grouped into four levels of increasing complexity: the primary structure, secondary structure, tertiary structure and quaternary structure. The primary structure is the simple linearly linked arrangement of amino acids found in newly synthesized polypeptide chains.In the primary structure, covalent bonds link the carboxyl group of the preceding amino acid with the amino group of the next amino acid. The first amino acid in the polypeptide chain is the N-terminus and the last amino acid with the carboxylic acid group is the C-terminus. The primary structure determines the higher orders of protein structure (Anfinsen, 1973).

The secondary structure of proteins refers to the pleating and spiraling of the primary protein structure due to hydrogen bonds between parts of the same polypeptide juxtaposed to each other (Sumner, 1988). This is the rearrangement of the primary structure of protein caused by regular hydrogen bonds between the backbone C=O group and the NH group in protein sequence. Secondary protein structure is further elucidated as the local spatial conformation of the polypeptide back bone excluding the side chains. The α-helix and β-pleats are secondary protein structures (He, et al, 2009) as well as turns (Sun et al, 2004).The α-helix is a right handled spatial helical conformation of the backbone of the protein structure (shape of a helix) whereby the C=O group of one amino acid (ith residue.) is linked to the N-H group of ith+ 4 amino acid residue in advance of the ith residue by hydrogen bonding. There are normally 3.6 amino acid residues per turn of the helix. β-pleated sheets or simply β-sheets form a group of secondary protein structure formed by extended strands of polypeptide chains linked by interchain hydrogen bonding. The β sheets or pleats are composed of two or more extended polypeptide chains called β strands that run alongside each other and are organized in parallel or antiparallel pattern. The β-pleated sheets may be parallel where the C terminus or N terminus of one end lie side by side or antiparallel where they lie in opposite directions. A mixed β-sheet is formed from both parallel and antiparallel strands. B-turns or β-bends or reverse turns are twisted polypeptide chains that connect helices and β-sheets. Loops and coils are secondary protein structure that have irregular hydrogen bonding from those in reverse turns, β-sheets and α-helices (Sun et al, 2004).

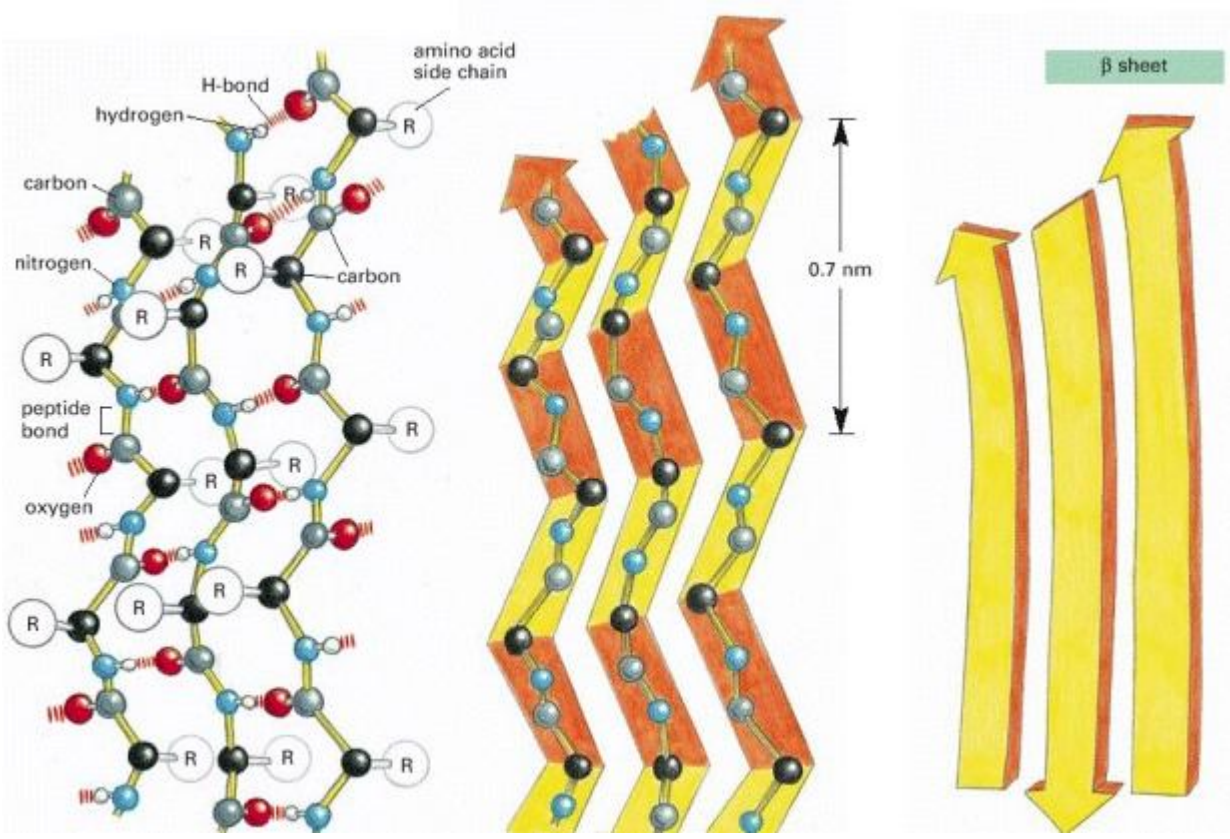
Topology refers to the pattern of connectivity between secondary protein structures, for instance, β-α-β topology where a β-strand is connected to an α-helix, then to another β-strand. The topology diagram is a two-dimensional diagrammatic representation of the secondary protein structure.



184

185 **Figure 2:** Secondary Protein Structure: Alpha Helix (Alberts et al., 2002)

186



187

Figure 3: Secondary Protein Structure: Beta Sheet (Alberts et al., 2002)

The tertiary structure of proteins is the overall three-dimensional appearance of proteins due to the interactions of the side chains and the way in which the secondary structure packs together to fold the protein.It is the three-dimensional arrangement of all the atoms that comprise the protein molecule. The tertiary structure of the protein molecule is maintained by hydrogen bonds, disulphide bonds, Van Der Waal forces and ionic (electrostatic) bonds that hold previously distant points into juxtaposition.

The quaternary protein structure is a multi-subunit protein molecule complex composed of many interactingsubunits that are folded to form the complex (He et al, 2009, Sumner, 1988). It is the exact spatial arrangement of subunits within a protein (Sun et al, 2004). It is the interaction of two or more constituent polypeptide subunits which creates the specific structural disposition of the protein molecule like the hemoglobin molecule that has two alpha and two beta subunit polypeptide chains making up its tetrameric structure.Molecular chaperones, intramolecular hydrogen bonds, electrostatic and Van der Waals forces contribute to proper folding of proteins (Upadhyay 2009).

The elucidation of protein structure has benefitted from advances in artificial intelligence. The neural network, alpha fold 2 (AF2) was developed by Jumper and his colleagues (Jumper et al., 2021) to predict the three-dimensional protein structure from its amino acid sequence to near experimental accuracy. Alpha fold 3 (AF3) was later developed by Abramson and his colleagues to predict the structure of proteins bound to other complexes like ions, ligands, nucleic acids (Abramson et al., 2024). Protein structure is essential in structure-based drug development and targeted mutagenesis (Tunyasuvunakool et al., 2021) Proteomics is the systematic study and analysis of proteins produced by cells of an organism or a human in disease and in the healthy state. This has been applied in the evaluation, monitoring, treatment and prognostication in disease management such is poly trauma (Alpantaki et al., 2007). Proteomics has been applied in precision medicine and the identification of disease biomarkers like thyroglobulin, a biomarker for follicular thyroid carcinoma (Van Eyk and Snyder, 2019).

4. PROTEIN DOMAINS, FOLDS AND MOTIFS

Domains, folds, and motifs are all components of the tertiary protein structure. Domains are discrete, compactly folded tertiary protein units that range in length from 100 to 200 residues. According to an examination of protein structures from the protein data bank, autonomously folding protein domains must have a maximum size of 200 residues. Folds are domain components characterized by the spatial organization of protein secondary structural elements into domain-like structures that comprise the domain. Motifs are similar to folds but smaller, and they are subunits of protein folds. Structural

and sequence motifs are interchangeable. Sequence motifs are well preserved structural motifs with a distinctive signature amino acid sequence.

Protein structural motifs include alpha-helix bundles, coiled coils, beta hairpins, Greek key and jelly-roll motifs, beta-sandwich motifs, mixed alpha/beta-sandwich motifs, beta-barrels, alpha/beta barrels, and beta propellers. A specific structural motif, fold, or domain may be linked to a particular protein activity (Sun et al, 2004). There are various bioinformatics websites and online databases that focus on proteins, protein domains, motifs, and folds, as well as their identification, annotation, classification, and structural alignment. PROSITE, SMART, Pfam, Interpro, TIGRFAM, Blocks, CDD, PRINTS, ProDom, RCSB PDB, MSD, MMBD, SCOP, CATH, FSSP, HOMSTRAD, VAST, and CE are among of them.

5. THE PHYSICOCHEMICAL PROPERTIES OF PROTEINS

The physicochemical properties of proteins depend on their amino acid constituents. Amino acids except glycine, exist as isomers (mirror images of themselves). The isomers are classed as ‘D’ (dextrorotatory) for rotation to the right or ‘L’ (levorotatory), for rotation to the left because they possess the property of optical rotation, which is their ability to rotate the plane of polarized light.Except for a few 'D' amino acid isomers found in bacteria, proteins are virtually entirely composed of 'L' amino acids.

Proteins that contain the aromatic amino acids, phenylalanine, tryptophan and tyrosine are intrinsically fluorescent (Lakowicz, 2006). Here we mean that they possess the ability to absorb ultraviolet electromagnetic radiation at certain wavelengths and then emit light. Protein fluorescence is applied in the spectroscopic determination of protein concentration. Fluorescent proteins are useful for live-cell imaging, whole animal imaging as well as serving as genetic labels when they are included in transgenic approaches (Kremers et al, 2011). Tryptophan also exhibits phosphorescence meaning that they emit the absorbed light for a longer duration (Garrett and Grisham, 2010). Fluorescent proteins have a unique β-barrel fold made up of 11 β-sheets surrounding an α-helix (Ormo et al., 1996). Mutations of fluorescent proteins have the potential of changing their fluorescent properties because of their unique β-barrel fold,making them capable of emitting light of various colors. This increases their utility as molecular probes(Kremers et al, 2011). Fluorescent proteins differ in their brightness, photostability, PH stability and monomeric properties(Cranfill et al., 2016).

Solubility is a physicochemical property of a solid substance, liquid, or gas known as a solute that dissolves in a liquid solvent to generate a homogeneous solution (Wedler & Freund, 2012). Protein solubility is the concentration of protein in a saturated solution that is in equilibrium with a solid phase, which can be crystalline or amorphous depending on the conditions (Cohn et al., 1943; Arakawa & Timasheff, 1985). Extrinsic and intrinsic variables regulate protein solubility (Kramer et al., 2012). Extrinsic

parameters influencing protein solubility include pH, temperature, ionic strength, and the presence of solvent additives (Colowick et al., 1962). Protein solubility is determined mostly by the amino acid concentration on its surface (Kramer et al., 2012).
The extrinsic factors affecting protein solubility are easier to manipulate than the intrinsic t54t54factors. There is inadequate knowledge of how to alter the intrinsic factors to increase protein solubility (Arakawa & Timasheff, 1985, Volkin & Middaugh, 1992).Protein solubility in aqueous solution can be quantified by adding lyophilized protein solvent and by the concentration of protein by ultrafiltration(Kramer et al., 2012). Protein solubility may be measured by using precipitants which are extraneous agents that reduce the solubility of proteins. Protein precipitants are classified into three types: salts, organic solvents, and long chain polymers (McPherson 2004).Denatured proteins have significantly higher hydrophobicity and lower solubility than native (properly folded) proteins. (Garidel 2013). Proteins are less soluble in polar solvents like ethanol and insoluble in nonpolar solvents like cyclohexane (Nick Pace et al., 2004).

Hydration is critical for protein solubility in water. Protein hydration is the process of gradually adding water to dry protein until a hydration point is achieved, after which further water addition makes no change but dilutes the protein (Rupley & Careri, 1991). Water of hydration refers to the water that is attached to the protein. The hydration shell is the water that remains with the protein at the end of hydration (Rupley & Careri, 1991). The addition of a salt, such as ammonium sulfate, can diminish the solubility of protein dissolved in water, causing the protein to precipitate. According to Haurowitz et al. (2023), the protein is salted off or precipitated.

Proteins are usually identified based on their molecular weight. Atomic force microscopy generates three-dimensional pictures of single molecules, allowing for molecular volume measurements. It can be used to calculate protein molecular weights because the molecular volume of single and multimeric proteins correlates with their molecular weights (Schneider et al., 1998). However, ultracentrifugation is the most commonly used method for measuring the molecular weight of proteins. SDS-PAGE is a method for determining protein molecular weight by comparing it to the electrophoretic migration of a protein with a known molecular weight (Jardine, 1990). Differences between the measured protein molecular weight and those predicted by its amino acid sequence can be attributable to post-translational changes of the protein. (Ahmad et al., 2005).

6. CLASSIFICATION OF PROTEINS

Proteins are conventionally classified according to structure into globular proteins. Fibrous proteins and membrane proteins (Chan& Dill, 1993). Globular proteins are sphere-like and are soluble in aqueous solutions. The hydrophobic monomers are tucked into the core of the protein while the polar monomers reside on the surface. Enzymes that catalyze nearly all biochemical reactions in living cells are globular proteins. Globular proteins have β-sheets and α-helices as their secondary structure(Chan & Dill, 1993). Fibrous proteins are long polypeptide chains that are

parallel to one another and are stabilized by cross linkages. They form support structures in humans and are important in biomechanics.They are insoluble in aqueous solutions. Collagen, a fibrous protein, is the most prevalent protein in the human body, accounting for up to 30% of the total protein pool. It possesses a triple-helix structure (Delgoda & Murray, 2017). Membrane proteins are responsible for transport and signaling across cell membranes. They can be transmembrane α-helices or β-strands (Gromiha & Ou, 2014).

Proteins may further be classified basedon their function or the biological process they partake in thus: nucleic acid binding protein, transcription factor, receptor, transporter, defense/immunity protein, isomerase, lyase, phosphatase, ligase, kinase, protease, oxidoreductase, enzyme modulator, transferase, cell adhesion molecule, extracellular matrix protein, membrane traffic protein, hydrolase, trans-membrane receptor regulatory/adaptive protein, chaperone, calcium binding protein, transfer/carrier protein, storage protein, cytoskeleton, viral protein, surfactant and structural protein (Gerret & Grisham, 2010).

The modern era of molecular biology and bioinformatics has led to classification of proteins according to structural and functional differences and similarities by adopting curated, automated or a hybrid approach (Ouzounis et al., 2003). The resources or databases for structural classification of proteins include: SCOP (structural classification of proteins); CATH (class, architecture, topology, homology); FSSP (fold classification based on structure-structure alignment of proteins); DSSP (database of secondary structure of proteins); HSSP (homology and secondary structure of proteins); Pfam (protein families); PRINTS (protein fingerprints); SMART (simple modular architecture research tool); PROSITE; TIGRFAM; PRODOM (PROTEIN DOMAIN); BLOCKS; eMOTIF; SYSTERS; ClustR (clusters of related proteins); COGS (clusters of orthologous groups); ProtoMap; MetaFam (meta database of protein families); TRIBES (Ouzounis et al., 2003). The resources or databases for functional classification of proteins include: EC (enzyme classification); YPD (yeast proteome database); SGD (saccharomyces genome database); MIPS (Munich information center for protein sequences); WIT (what is there) renamed ERGO; STRING (search tools for the retrieval of interacting genes/proteins); AllFuse; Predictome; Riley; GeneQuiz; GO (gene ontology); BioCyc; KEGG (Kyoto encyclopedia of genes and genomes); DIP (Database of Interacting Proteins); YPL.db; TRIPLE (Transposon-Insertion phenotypes, localization and expression in saccharomyces); MINT (Molecular Interaction); BIND (Biomolecular Interactions Database); PIM (Protein Interaction Manager); CellZome db (Ouzounis et al., 2003).

7. PROTEIN IDENTIFICATION BY HISTOCHEMISTRY

Histochemical identification of proteins in animal tissues can be achieved by ultraviolet (UV) absorption; direct staining methods; enzyme histochemical methods if the protein

is an enzyme; immunohistochemical methods if the protein is antigenic and autoradiography after injecting a radioactively labelled metabolic precursor (Sumner, 1988).

Proteins absorb UV radiation at peak wavelengths of 240 to 310 nm based on their amino acid content. This coincides with the peak UV absorption of nucleic acids making identification difficult. However, hemoglobin has a peak UV absorption wavelength at 416 nm. It can be quantitatively identified by microspectrophotometry (James & Tas, 1984). UV absorption at wavelength of 280nm is due to tryptophan and tyrosine in the protein sequence and their absence in the protein sequence gives erroneous results (Noble & Bailey, 2009).

The direct staining methods for identifying proteins include: methods for staining proteins ionically according to their acidic or basic properties, like the alkaline fast green method as an instance; methods for staining chemical groups on a protein by a covalent attachment, like the ninhydrin-schiff test for amino groups (Yasuma & Ichikawa, 1953); methods for fluorescent staining of proteins by covalent bonding to a fluorescent dye, like the fluorescein isothiocyanate (FITC) as an instance (Chaganti et al.,2018); Methods for staining specific amino acid residues like tryptophan using the DMAB-nitrite (dehydrocoupling of dimethylamine-borane-nitrite) method (Shori et al., 1997); Methods for staining specific proteins like collagen using picro-sirius red (Puchtler et al., 1988); methods for detecting abolition of staining following removal of proteins by specific enzyme digestion (Pearse, 1968).

8. PROTEIN SYNTHESIS

Proteins synthesis in cells uses amino acid building blocks and messenger ribonucleic acids (mRNA) as the templates. Ribonucleic acid (RNA) is important in converting the genetic information into proteins. It differs from DNA (deoxyribonucleic acid): It has only one strand and is not double stranded like DNA; It has a ribose sugar which differs from the deoxyribose sugar of DNA in having an extra hydroxyl group; It has an uracil nitrogenous base instead of the thymine of DNA, though the uracil still binds with adenine as its complementary base. Three types of RNA are involved in the conversion of genetic information to proteins: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) (Li et al.,2011).

Gene expression proceeds through two stages: transcription and translation. Transcription is the process of converting genetic information from DNA sequence to mRNA. Messenger RNA carries the genetic information from the chromosome to the site of protein synthesis.
Translation is the process of converting the genetic information in mRNA to proteins by polypeptide synthesis(Kohler et al., 2017; Artsimovitch, 2018).

Transcription is the process of copying a gene's sequence from DNA into a single-stranded mRNA molecule. Transcription has three stages: initiation, elongation, and termination.

Initiation: The RNA polymerase enzyme binds to the DNA segment that will be transcribed in a location known as the promoter and opens the double helix. The promoter region is located in front of the gene that will be transcribed and is typically made up of a string of adenine and thymine bases that act as the recognition region for the RNA polymerase. The promoter specifies which DNA strand should be transcribed and where RNA polymerase should begin transcription on the DNA. RNA polymerase's binding site recognizes only the promoter region and can only bind in front of the gene.

Elongation: The RNA polymerase attaches to the DNA promoter region and opens the double helix. It then creates the single-stranded mRNA in a 5' to 3' direction without transcribing the promoter region, analogous to DNA replication. However, RNA polymerase does not need a primer and merely transcribes one strand of DNA. The transcribed DNA strand acts as a template strand. The transcribed mRNA strand is complementary to the template strand of DNA, with the exception that it contains uracil instead of thymine.

Termination: The synthesis of the mRNA goes on until the RNA polymerase reaches the end of the gene which is recognized by a stop sequence called the termination sequence. Here transcription stops, the newly synthesized mRNA disconnects from the DNA template strand. The RNA polymerase is free to bind to another promoter region and transcribe another mRNA
(Merrick, 1992).
In the second part of the central dogma of molecular biology, the genetic information in the mRNA is decoded to make chains of amino acids to form polypeptides. Translation involves protein synthesis and is dependent on the nature of the genetic code. All proteins are made up of only 20 amino acids, which are coded for by various combinations of the four nitrogenous bases. The genes in the DNA are read in groups of three nucleotides.
Codons can be any combination of the four nitrogenous bases, resulting in a total of 64 distinct codons (43 = 64) for the 20 amino acids. Codons apply to both DNA and mRNA. In the mRNA codon, AUG codes for the amino acid methionine and is known as the'start codon' since all translations begin there. So, while all polypeptide chains begin with methionine, it may be edited out later. Three more codons (UAA, UAG, and UGA) are known as "stop codons," and they do not code for amino acids. They indicate where the translation should conclude.
The remaining sixty codons code for amino acids implying that more than one codon, in fact, many codons may code for the same amino acid eg serine and leucine each has 6 codons.

Translation has three stages like transcription: initiation, elongation, termination. Ribosome recycling follows this (Maquat et al., 2010).

Initiation: Initiation of translation starts when a ribosome recognizes a unique mRNA codon and binds to it. The ribosome in eukaryotic cells has a large subunit and a small subunit. The two subunits bind to the mRNA, clamping it between them and move along it in a 5’ to 3’ direction adding new amino acid units to the growing polypeptide chain each time it reads a codon. The ribosome synthesizes different polypeptides by reading the coding sequence on the mRNA. It must begin reading this codon sequence from the start codon AUG, in order to read the right reading frame of the mRNA. The ribosome binds to the mRNA's start codon, ensuring that the genetic code is translated using the mRNA's reading frame. To translate the genetic code into the correct amino acid sequence, the mRNA must be positioned in the ribosome in its reading frame. Once the ribosome has attached to the mRNA molecule, transfer RNA (tRNA) provides the appropriate amino acid to be added to the polypeptide. The anticodon, a three-base sequence, is found at one end of the tRNA and is complementary to the mRNA sequence. The opposite end of the tRNA transports the matching amino acid to the mRNA sequence.Each tRNA transports a unique amino acid. There are 20 amino acids and 64 codons, hence the number of tRNA molecules available varies from 20 to 64 depending on the organism. The tRNA has a clover leaf structure (Laursen et al. 2005).

Elongation: The ribosome recognizes the start codon, AUG, which also codes for methionine. All polypeptides begin with the amino acid methionine. The ribosome has two tRNA attachment sites: the 'A' (aminoacyl) and the 'P' (peptidyl). The tRNA containing anticodon corresponding to the mRNA's start codon enters the P site. The next tRNA that has the designated amino acid enters the A site. A peptide link is formed between the methionine and the following amino acid. The ribosome moves over one codon, freeing up the P site for the second tRNA and releasing the methionine-carrying-tRNA. The second tRNA goes to the P site, freeing up the A site for the third tRNA. The released tRNAs are recycled in the cytoplasm by incorporating new amino acids. This process continues until the full mRNA sequence is translated into the specified protein and the ribosome reaches a stop codon (Czworkowski & Moore, 1996).

Termination: The ribosome eventually hits one of the stop codons UGA, UAG, or UAA, which do not code for any amino acid and thus lack a corresponding tRNA. The release factor is a protein that detects stalled ribosomes and aids in the release of the polypeptide chain. The ribosome's two subunits separate from the mRNA as translation ends.

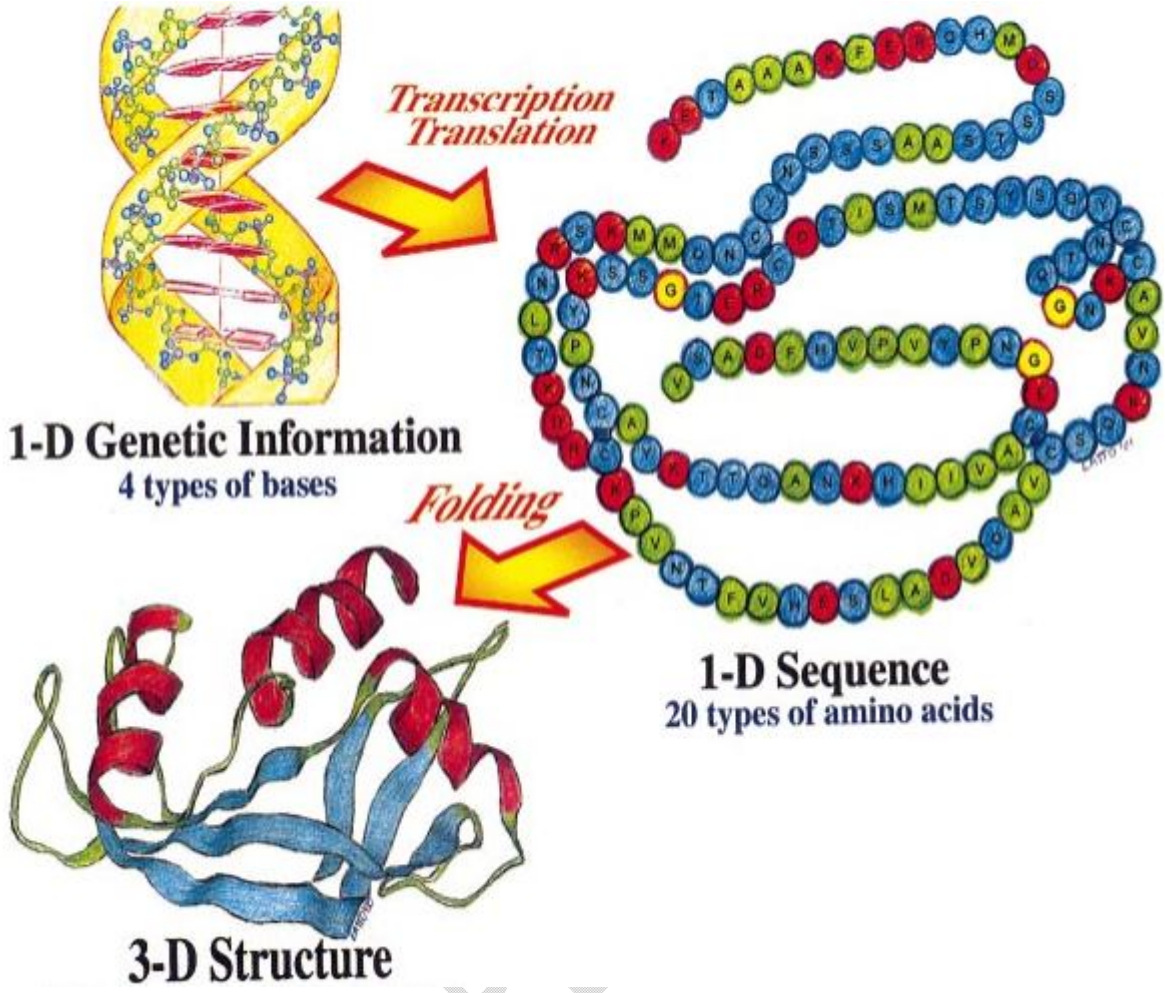


Figure 4: Protein Synthesis (Plotkin & Onuchic, 2002).

9. POST-TRANSLATIONAL MODIFICATION OF PROTEINS

Post-translational modification (PTM) of proteins are alterations in the structural content of proteins after their biosynthesis by proteolytic cleavage and incorporation of one or more modifying groups like phosphoryl, glycosyl, acetyl, methyl, etc to one or more amino acids, consequentially changing the properties and functions of the protein (Ramazi et al., 2020). Proteins can be reversibly or irreversibly modified after they have been translated (Wang I et al., 2014). A single amino acid or a group of amino acids may be modified (Huang et al., 2019). Post-translational modification of proteins may occur in the nucleus, cytoplasm, endoplasmic reticulum or golgi apparatus (Blom et al., 2004). There are over 400 PTMs of proteins (Khoury et al., 2011). Phosphorylation, acetylation and ubiquitination comprise more than 90% of PTMs reported in the dbPTM , a comprehensive database of PTM of proteins. Each amino acid has at least three different PTMs. Lysine has fifteen PTMs, the highest number of PTMs among amino acids. Other major PTMs include glycosylation, methylation, SUMOylation, palmitoylation, myristoylation, prenylation, sulfation (Ramazi & Zahiri, 2021). New PTMs

discovered by updated mass spectroscopy include ufmylation, crotonylation, succinylation and lactylation (Wu et al,. 2024). Post-translational modification of proteins may influence biological processes like signal transduction, cell cycle control, gene expression control, gene activation, DNA repair. (Wang II et al., 2015, Strumillo & Beltrao, 2015, Wei et al., 2017), The influence of PTMs of proteins on biological processes is brought about by their effects on protein folding (Del Monte & Agnetti, 2014), protein lifespan and protein-protein interaction (Marshall, 1993), protein localization(Audagnotto & Dal Peraro, 2017), enzyme function and assembly (Ryšlavá et al.,2013), protein solubility (Ramazi & Zahiri, 2021), molecular trafficking (Cundy et al., 2002, Ohtsubo & Marth, 2006), cell-cell and cell-matrix interaction(Goulabchand et al., 2014, Caragea et al., 2007), receptor activation (Ohtsubo & Marth, 2006).Post-translational modification of proteins may be studied or detected by proximity ligation assay, an immunoassay technology employed in the study of PTM of proteins (Leuchowius et al., 2011), immunoprecipitation which is used in PTM detection assays (Fuchs & Strahl, 2011) or a combination of mass spectrometry and immunoprecipitation (Larsen et al., 2006). Other experimental methods for detection of PTMs include Liquid chromatography (Welsch & Nelsestuen, 1988), radioactive chemical method (Slade et al., 2014), western blotting (Jaffrey et al., 2001) and eastern blotting (Welsch & Nelsestuen, 1988). Nanopore technology is emerging as a novel method of detecting PTMs (Zhao et al., 2024).Computational methods are being used in recent times to predict PTMs of proteins because experimental identification of PTMs is expensive(Hasan & Khatun, 2018, Sobolev et al., 2014). Aberrant PTM pathways in humans are associated with diseases like cancer, heart failure, autoimmune diseases, neurodegenerative disease like Parkinson's disease, Alzheimer's disease and Huntingtons disease (Ramazi et al., 2020).

10. CONCLUSION

Proteins are the most numerous biomolecules in animals and they are the work horse of their life processes. Proteins are polymers made up of amino acid units whose sequence form their primary structure. The protein secondary, tertiary and quaternary structures are as a result of protein folding by interplay of non-covalent bonding. The primary protein structure dictates the subsequent structural folds, properties and functions of the proteins. Proteins are synthesized in cells using the genetic code. Post-translational modification of proteins alters their structure and diversify their functions. Their pathway disruption may cause disease.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ABBREVIATIONS

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

NMR: Nuclear magnetic resonance

CD: circular dichroism

FRET: Fluorescence resonance energy transfer

MD: Molecular dynamics

D: Dextrorotatory

L: Levorotatory

mRNA: Messenger ribonucleic acid

pH: Potential of hydrogen

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SCOP: Structural classification of proteins

CATH: Class, architecture, topology, homology

FSSP: Fold classification based on structure-structure alignment of proteins

DSSP: database of secondary structure of proteins

HSSP: Homology and secondary structure of proteins

Pfam: Protein families

PRINTS: Protein fingerprints

SMART: Simple modular architecture research tool

PRODOM: Protein domain

ClustR: Clusters of related proteins

COGS: Clusters of orthologous groups

MetaFam: Meta database of protein families

EC: Enzyme classification

YPD: Yeast proteome database

SGD: Saccharomyces genome database

MIPS: Munich information center for protein sequences

WIT: What is there

STRING: Search tools for the retrieval of interacting genes/proteins

GO: Gene ontology

KEGG: Kyoto encyclopedia of genes and genomes

DIP: Database of Interacting Proteins

TRIPLE: Transposon-Insertion phenotypes, localization and expression in saccharomyces

MINT: Molecular Interaction

BIND: Biomolecular Interactions Database

PIM: Protein Interaction Manager

UV: Ultraviolet

FITC: Fluorescein isothiocyanate

DMAB-nitrite: Dehydrocoupling of dimethylamine-borane-nitrite

tRNA: Transfer RNA

rRNA: Ribosomal RNA

A: Aminoacyl

P: Peptidyl

PTM: Post-translational modification

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

REFERENCES

1. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., ... & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3.
2. Ahmad, Q. R., Nguyen, D. H., Wingerd, M. A., Church, G. M., & Steffen, M. A. (2005). Molecular weight assessment of proteins in total proteome profiles using 1D-PAGE and LC/MS/MS. *Proteome science*, *3*(1), 6. https://doi.org/10.1186/1477-5956-3-6
3. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). The shape and structure of proteins. In *Molecular Biology of the Cell. 4th edition*. Garland Science.
4. Alpantaki, K., Tsiridis, E., Pape, H. C., & Giannoudis, P. V. (2007). Application of clinical proteomics in diagnosis and management of trauma patients. *Injury*, *38*(3), 263-271.
5. Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. Science 181:223-230.
6. Arakawa, T., & Timasheff, S. N. (1985). Theory of protein solubility. *Methods in enzymology*, *114*, 49–77. https://doi.org/10.1016/0076-6879(85)14005-x
7. Artsimovitch, I. (2018). Rebuilding the bridge between transcription and translation. *Molecular microbiology*, *108*(5), 467-472.
8. Audagnotto, M., & Dal Peraro, M. (2017). Protein post-translational modifications: In silico prediction tools and molecular modeling. *Computational and structural biotechnology journal*, *15*, 307-319.

9. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, *4*(6), 1633-1649.

10. Cahill, G.F., Jr. (2006) Fuel metabolism in starvation. *Annu. Rev. Nutr.*. *26*, 1–22.

11. Caragea, C., Sinapov, J., Silvescu, A., Dobbs, D., & Honavar, V. (2007). Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC bioinformatics*, *8*, 1-13.

12. Carroll, B.; Korolchuk, V.I.; Sarkar, S. (2015) Amino acids and autophagy: Cross-talk and co-operation to control cellular homeostasis. *Amino Acids*, *47*, 2065–2088.

13. Chaganti, L. K., Venkatakrishnan, N., & Bose, K. (2018). An efficient method for FITC labelling of proteins using tandem affinity purification. *Bioscience reports*, *38*(6), BSR20181764.

14. Chan, H. S., & Dill, K. A. (1993). The protein folding problem. *Physics today*, *46*(2), 24-32.

15. Cohn, E. J., Edsall, J. T., Kirkwood, J. G., Mueller, H., Oncley, J. L., & Scatchard, G. (1943). Proteins, amino acids and peptides as ions and dipolar ions. *The American Journal of the Medical Sciences*, *205*(6), 884.

16. Colowick, S. P., Abelson, J. N., Simon, M. I., & Packer, L. (1962). *Methods in enzymology* (Vol. 5, p. 543). N. O. Kaplan (Ed.). New York: Academic Press.\

17. Cozzone, A. J. (2002). Proteins: Fundamental chemical properties. *Encyclopedia of Life Sciences. No month listed—2002. John Wiley & Sons Ltd*, 1-10.

18. Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., de Gruiter, H. M., ... & Piston, D. W. (2016). Quantitative assessment of fluorescent proteins. *Nature methods*, 557-562.

19. Crick, F. H. (1958, January). On protein synthesis. In *Symp Soc Exp Biol* (Vol. 12, No. 138-63, p. 8).

20. Cundy, T., Hegde, M., Naot, D., Chong, B., King, A., Wallace, R., ... & Cornish, J. (2002). A mutation in the gene TNFRSF11B encoding osteoprotegerin causes an idiopathic hyperphosphatasia phenotype. *Human Molecular Genetics*, *11*(18), 2119-2127.

21. Czworkowski, J., & Moore, P. B. (1996). The elongation phase of protein synthesis. *Progress in nucleic acid research and molecular biology*, *54*, 293-332.

22. Delgoda, R., & Murray, J. E. (2017). Evolutionary perspectives on the role of plant secondary metabolites. In *Pharmacognosy* (pp. 93-100). Academic Press.

23. Del Monte, F., & Agnetti, G. (2014). Protein post-translational modifications and misfolding: New concepts in heart failure. *PROTEOMICS–Clinical Applications*, *8*(7-8), 534-542.

24. Fuchs, S. M., & Strahl, B. D. (2011). Antibody recognition of histone post-translational modifications: emerging issues and future prospects. *Epigenomics*, *3*(3), 247-249.

25. Garidel, P. (2013). Protein solubility from a biochemical, physicochemical and colloidal perspective. *Am Pharm Rev*, *2*(5), 26-28.

26. Garrett R.H, Grisham C.M. (2010). Biochemistry, Fourth Edition, Brooks/Cole, Cengage Learning

27. Goulabchand, R., Vincent, T., Batteux, F., Eliaou, J. F., & Guilpain, P. (2014). Impact of autoantibody glycosylation in autoimmune diseases. *Autoimmunity reviews*, *13*(7), 742-750.

28. Gromiha, M. M., & Ou, Y. Y. (2014). Bioinformatics approaches for functional annotation of membrane proteins. *Briefings in bioinformatics*, *15*(2), 155-168.

29. Gruebele, M. (2009). Protein dynamics: from molecules, to interactions, to biology. *International journal of molecular sciences*, *10*(3), 1360-1368.

30. Hasan, M. M., & Khatun, M. S. (2018). Prediction of protein post-translational modification sites: an overview. *Ann Proteom Bioinform*, *2*, 049-57.

31. Haurowitz, F. and Koshland, . Daniel E. (2023, November 11). protein. Encyclopedia Britannica. https://www.britannica.com/science/protein

32. He, B., Wang, K., Liu, Y. *et al.* (2009). Predicting intrinsic disorder in proteins: an overview. Cell Res **19**, 929–949. https://doi.org/10.1038/cr.2009.87

33. Henzler-Wildman KA, Kern D. (2007). Dynamic personalities of proteins. Nature, 450: 964–972

34. Hou Y, Yin Y, Wu G. (2015). Dietary essentiality of “nutritionally non-essential amino acids” for animals and humans. Experimental Biology and Medicine. 240(8):997-1007. doi:10.1177/1535370215587913

35. Huang, K. Y., Lee, T. Y., Kao, H. J., Ma, C. T., Lee, C. C., Lin, T. H., ... & Huang, H. D. (2019). dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic acids research*, *47*(D1), D298-D308.

36. Jaffrey, S. R., Erdjument-Bromage, H., Ferris, C. D., Tempst, P., & Snyder, S. H. (2001). Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nature cell biology*, *3*(2), 193-197.

37. Jain, K. K. (2016). Role of proteomics in the development of personalized medicine. *Advances in protein chemistry and structural biology*, *102*, 41-52.

38. James, J., & Tas, J. (1984). Histochemical protein staining methods. Microscopy Handbook, Royal Microscopical Society/Oxford University Press, Oxford.

39. Jardine, I. (1990). [24] Molecular weight analysis of proteins. In *Methods in enzymology* (Vol. 193, pp. 441-455). Academic Press.

40. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *nature*, *596*(7873), 583-589.

41. Katz E. (2021). Electrochemical Contributions: Jons Jacob Berzelius (Jacob Berzelius, 1779-1848), Editorial. Electrochemical science Advances. 1:e2160005.https://doi.org/10.1002/elsa.202160005

42. Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports*, *1*(1), 90.

43. Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., & Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science*, *356*(6334), 194-197.

44. Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., & Scholtz, J. M. (2012). Toward a molecular understanding of protein solubility: increased negative

surface charge correlates with increased solubility. *Biophysical journal*, *102*(8), 1907–1915. https://doi.org/10.1016/j.bpj.2012.01.060

45. Kremers, G. J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W., & Piston, D. W. (2011). Fluorescent proteins at a glance. *Journal of cell science*, *124*(2), 157-160.

46. Lakowicz, J. R. (Ed.). (2006). *Principles of fluorescence spectroscopy*. Boston, MA: springer US.

47. Larsen, M. R., Trelle, M. B., Thingholm, T. E., & Jensen, O. N. (2006). Analysis of posttranslational modifications of proteins by tandem mass spectrometry: Mass Spectrometry For Proteomics Analysis. *Biotechniques*, *40*(6), 790-798.

48. Laursen, B. S., Sørensen, H. P., Mortensen, K. K., & Sperling-Petersen, H. U. (2005). Initiation of protein synthesis in bacteria. *Microbiology and molecular biology reviews*, *69*(1), 101-123.

49. Leuchowius, K. J., Weibrecht, I., & Söderberg, O. (2011). In situ proximity ligation assay for microscopy and flow cytometry. *Current protocols in cytometry*, *56*(1), 9-36.

50. Li HM, Xie YH, Liu CQ, Liu SQ. (2014) Physicochemical bases for protein folding, dynamics, and protein-ligand binding. Sci China Life Sci, 57: 287–302, doi: 10.1007/s11427-014-4617-2

51. Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., & Cheung, V. G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *science*, *333*(6038), 53-58.

52. Lopez MJ, Mohiuddin SS. (2022) Biochemistry, Essential Amino Acids. In: StatPearls. StatPearls Publishing, Treasure Island (FL). PMID: 32496725.

53. Maquat, L. E., Tarn, W. Y., & Isken, O. (2010). The pioneer round of translation: features and functions. *Cell*, *142*(3), 368-374.

54. Marshall, C. J. (1993). Protein prenylation: a mediator of protein-protein interactions. *Science*, *259*(5103), 1865-1866.

55. McPherson A. (2004). Introduction to protein crystallization. *Methods (San Diego, Calif.)*, *34*(3), 254–265. https://doi.org/10.1016/j.ymeth.2004.03.019

56. Merrick, W. C. (1992). Mechanism and regulation of eukaryotic protein synthesis. *Microbiological reviews*, *56*(2), 291-315.

57. Millward, D.J.; Garlick, P.J. (1972). The pattern of protein turnover in the whole animal and the
effect of dietary variations. *Proc. Nutr. Soc. 31*, 257–263.

58. Nick Pace, C., Trevino, S., Prabhakaran, E., & Martin Scholtz, J. (2004). Protein structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1448), 1225-1235.

59. Noble, J. E., & Bailey, M. J. (2009). Quantitation of protein. *Methods in enzymology*, *463*, 73-95.

60. Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. *Cell*, *126*(5), 855-867.

61. Ormo M., Cubitt A. B., Kallio K., Gross L. A., Tsien R. Y., Remington S. J. (1996). Crystal structure of the Aequorea victoria green fluorescent protein. *Science*273, 1392-1395

62. Ouzounis, C. A., Coulson, R. M., Enright, A. J., Kunin, V., & Pereira-Leal, J. B. (2003). Classification schemes for protein structure and function. *Nature Reviews Genetics*, *4*(7), 508-519.

63. Pearse, A. G. E. (1980). Histochemistry, theoretical and applied: 4th edn, vols 1 & 2. Churchill Livingstone.

64. Plotkin, S. S., & Onuchic, J. N. (2002). Understanding protein folding with energy landscape theory part I: basic concepts. *Quarterly reviews of biophysics*, *35*(2), 111-167.

65. Ramazi, S., Allahverdi, A., & Zahiri, J. (2020). Evaluation of post-translational modifications in histone proteins: A review on histone modification defects in developmental and neurological disorders. *Journal of biosciences*, *45*, 1-29.

66. Ramazi, S., & Zahiri, J. (2021). Post-translational modifications in proteins: resources, tools and prediction methods. *Database*, *2021*, baab012.

67. Reeds, P.J.; Biolo, G. (2002) Non-protein roles of amino acids: An emerging aspect of nutrient requirements. Curr. Opin. Clin. Nutr. Metab. Care., *5*, 43–45

68. Rose A.J. (2019). Amino Acids Nutrition and Metabolism in Health and Disease. Nutrients. 11(11), 2623; https://doi.org.10.3390/nu11112623/

69. Rose, A.J.; Richter, E.A. (2009). Regulatory mechanisms of skeletal muscle protein turnover during exercise. *J. Appl. Physiol. 106*, 1702–1711.

70. Rose, W.C., II. (1968). The sequence of events leading to the establishment of the amino acid needs of man. *Am. J. Public Health Nation's Health*. *58*, 2020–2027.

71. Rupley, J. A., & Careri, G. (1991). Protein hydration and function. *Advances in protein chemistry*, *41*, 37-172.

72. Ryšlavá, H., Doubnerová, V., Kavan, D., & Vaněk, O. (2013). Effect of posttranslational modifications on enzyme function and assembly. *Journal of proteomics*, *92*, 80-109.

73. Salem MSZ (2013) Medical Genetics: An Overview. Human Genet Embryol S5: 001. doi:10.4172/2161-0436.S5-00

74. Schneider, S. W., Lärmer, J., Henderson, R. M., & Oberleithner, H. (1998). Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflugers Archiv : European journal of physiology*, *435*(3), 362–367. https://doi.org/10.1007/s004240050524

75. Shori, D. K., Proctor, G. B., Garrett, J. R., Zhang, X. S., & Carpenter, G. H. (1997). Histochemical staining of ducts in submandibular glands by DMAB-nitrite detects stored tissue kallikreins. Biochemical Society Transactions, *25*(1), 28S-28S.

76. Slade DJ, Subramanian V, Fuhrmann J, Thompson PR. Chemical and biological methods to detect post-translational modifications of arginine. Biopolymers. 2014; 101: 133-143. Ref.: https://goo.gl/qBW8uz

77. Sobolev, B. N., Veselovsky, A. V., & Poroikov, V. V. E. (2014). Prediction of protein post-translational modifications: main trends and methods. *Russian Chemical Reviews*, *83*(2), 143.

78. Stollar, E. J., & Smith, D. P. (2020). Uncovering protein structure. *Essays in biochemistry*, *64*(4), 649–680. https://doi.org/10.1042/EBC20190042

79. Stollar, E. J., & Smith, D. P. (2021). Correction: Uncovering protein structure. *Essays in biochemistry*, *65*(2), 407. https://doi.org/10.1042/EBC-2019-0042C_COR

80. Strumillo, M., & Beltrao, P. (2015). Towards the computational design of protein post-translational regulation. *Bioorganic & medicinal chemistry*, *23*(12), 2877-2882.

81. Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. *Current protocols in protein science*, *Chapter 17*(1), 1711–17.1. https://doi.org/10.1002/0471140864.ps1701s35

82. Tuite, M. F., & Stansfield, I. (1994). Termination of protein synthesis. *Molecular biology reports*, *19*, 171-181.

83. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., ... & Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*(7873), 590-596.

84. Upadhyay, A. (2019) Structure of proteins: Evolution with unsolved mysteries, Progress in Biophysics and Molecular Biology,https://doi.org/10.1016/j.pbiomolbio.2019.04.007

85. Van Eyk, J. E., & Snyder, M. P. (2019). Precision Medicine: Role of Proteomics in Changing Clinical Management and Care. *Journal of proteome research*, *18*(1), 1–6. https://doi.org/10.1021/acs.jproteome.8b00504

86. Volkin, D. B., & Middaugh, C. R. (1992). Stability of protein pharmaceuticals. *Chemical and Physical Pathways of Protein Degradation. Plenum Press, New York, NY, USA*, 109-134.

87. Wang, Y. C., Peterson, S. E., & Loring, J. F. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell research*, *24*(2), 143-160.

88. Wang, M., Jiang, Y., & Xu, X. (2015). A novel method for predicting post-translational modifications on serine and threonine sites by using site-modification network profiles. *Molecular BioSystems*, *11*(11), 3092-3100.

89. Wedler, G., & Freund, H. J. (2012). *Textbook of Physical Chemistry* (Vol. 1). John Wiley & Sons.

90. Wei, L., Xing, P., Shi, G., Ji, Z., & Zou, Q. (2017). Fast prediction of protein methylation sites using a sequence-based feature selection technique. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(4), 1264-1273.

91. Welsch, D. J., & Nelsestuen, G. L. (1988). Amino-terminal alanine functions in a calcium-specific process essential for membrane binding by prothrombin fragment 1. *Biochemistry*, *27*(13), 4939-4945.

92. Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, *293*(2), 321-331.

93. Wu, Z., Liu, Y., Zhang, M., & Wang, D. (2024). Emerging posttranslational modifications and their roles in DNA damage response. *Genome Instability & Disease*, *5*(1), 1-16.

94. Yang LQ, Sang P, Xie YH, Tao Y, Fu YX, Zhang KQ, Liu SQ. Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. J Biomol Struct Dyn, 2014, 32: 372–393

95. Yasuma, A., & Ichikawa, T. (1953). Ninhydrin-Schiff and alloxan-Schiff staining: a new histochemical staining method for protein. *The Journal of Laboratory and Clinical Medicine*, *41*(2), 296-299.

96. Zhao, X., Qin, H., Tang, M., Zhang, X., & Qing, G. (2024). Nanopore: Emerging for detecting protein post-translational modifications. *TrAC Trends in Analytical Chemistry, 117658.*

939

940

941

942
943
944