

A

Acidic amino acids: Those whose side chains can carry a negative charge at certain pH values. Typically aspartic acid, glutamic acid.

Active site: Usually applied to catalytic site of an enzyme or where chemical transformations take place.

Alignment: Tabulation of genetically related protein sequences arranged (using introduction of sequence gaps if necessary) so as to maximise visual resemblance.

Allostery: Property of some proteins by which spatially separate functional sites in the molecule can communicate with each other via movements transmitted through the structure.

Alpha-helix. Frequently highly stable protein chain conformation stabilised by hydrogen bonding between peptide groups and having side chains projecting outwards in a spiral (3.6 residues per turn). One of the classic "secondary" structures. Has end to end dipole because of alignment of peptide carbonyl groups.

Amino acid: Monomer containing an amino group and a carboxyl group that is polymerised to form peptide and protein chains. Typically, peptide/protein-forming amino acids have the amino and carboxyl groups attached to the same carbon atom (the alpha carbon) and are designated alpha amino acids. A variable substituent on the alpha carbon generates different amino acids with different chemical properties.

Amino acid analyser. Automated machine that determines amino acid composition of a protein.

Amino acid sequencer: Automated machine that determines linear order of amino acids in protein chain (i.e. the protein's primary structure).

Amino acid synthesiser: Automated machine capable of synthesising polypeptide chains by solid phase method.

Amino acid side chain: That portion of the amino acid that extends beyond the α -carbon and gives the amino acid its unique chemical character.

Amino acid residue: the portion of the amino acid that remains after incorporation into a polypeptide chain. Includes the α -carbon and the nitrogen/carbonyl moieties.

Aminopeptidase. Enzyme that cleaves amino acid residues from a protein chain commencing at the N-terminal

Antibodies: Proteins involved in recognising bacteria, viruses, toxins, etc. as part of the immune defence system. Circulate in blood/lymph systems. Also called immunoglobulins.

Apoenzyme: Enzyme lacking correct cofactor/coenzyme and hence unable to function correctly.

Aromatic amino acids: Those with aromatic side chains, typically tyrosine, phenylalanine, tryptophan, histidine.

Autologue: Protein derived from the same organism's tissues or genome as another protein.

B

Basic amino acids: Those whose side chains can carry a positive charge at certain pH values. Typically lysine, arginine, histidine.

Beta-conformation: Stable "zig-zag" conformation of protein chain which can align with sections of chain in same conformation and cross-link efficiently via hydrogen bonds between the peptide linkages. One of the classic "secondary" structures.

Beta sheet: Two or more sections of protein chain in beta conformation that have become aligned and linked by hydrogen bonding. Effectively form a twisted sheet with side chains projecting above and below plane of sheet. Anti-parallel sheet arises when neighbouring chains are aligned in opposite directions. Parallel sheet has the neighbouring chains running in the same direction.

Beta turn: Defined 90 or 180 degree turn in a protein chain stabilised by a hydrogen bond across the turn between peptide units. Several types and frequently initiated by the presence of proline and glycine.

Binding site: Locality on a protein's surface that attracts another molecule of whatever type and leads to a docking of the two. The attraction will be based on a combination of geometrical fit and chemical complementarity. It can be either weak or strong, short-lived or long-lived. The binding may or may not translate into an action such as catalysis or structural adjustment.

Bioinformatics: As applied to proteins, typically the use of theoretical computational techniques to classify, group, align, store, access and interpret protein sequence/3D data. Major application is in the prediction of protein shape, properties and function from primary sequence data alone.

C

C-terminus: That end of a protein chain that carries a free -COOH group.

Carboxypeptidase: Enzyme that cleaves amino acids from a protein chain commencing at the C-terminal

Chromatography: General term for related techniques to purify the components of peptide/protein mixtures according to molecular size, polarity, charge, recognition properties, etc.

Circular dichroism: Valuable technique for finding balance of secondary structures in a protein molecule and for measuring conformational responses to changing conditions.

Coenzyme: A small organic molecule that facilitates an enzyme's catalytic action by providing a chemical property not achievable by amino acid residues. Precursors are needed in diet as vitamins. E.g. NAD, FAD

Collagen: Structural protein found in bones, muscles, tendons.

Common amino acids: The set of 20 usually used by organisms to construct proteins.

Conformational change: Large, medium or small adjustment to the molecular shape/dynamics of a protein that may follow a change in environment or the specific recognition of another molecule.

Core: Used to describe the interior regions of some protein molecules, typically a region of densely packed hydrophobic amino acid residues that mutate infrequently and make a substantial contribution to molecular stability.

Covariation: Phenomenon observed in aligned homologous sequence sets wherein side chains at different sequence positions have apparently changed synchronously.

Cryo-electron microscopy: Very low temperature variant of electron microscopy to immobilise and visualise systems in dynamic motion at more normal temperatures. E.g. study of membrane proteins by freeze-fracture of membrane lipid bilayer.

D

D-amino acid. Isomeric form of amino acid rarely seen in natural peptides/proteins except in some bacterial cell walls.

Denaturation: Loss of a protein's typical conformation and activity through extreme alterations to its environment (e.g. changes in pH, temperature, etc.).

Disulfide bridge: Covalent linkage between 2 cysteine side chains formed when they become in close proximity in the fold and there is an oxidising environment. Play a critical role in holding together the folds of many proteins and enhancing the stability under hostile conditions.

Domain: Partly autonomous substructure in a protein molecule, with self-contained secondary structure/stabilising features and sometimes wholly self-originating functional contributions.

E

Electron Microscopy: Very high resolution form of microscopy, capable of visualising the smallest organisms and the organisation of macromolecular protein complexes.

Electrophoresis: Experimental technique by which a mixture of peptides/proteins can be separated on the basis of their molecular charge.

Electrospray mass spectrometry: Useful method for analysing the composition of peptide mixtures according to their molecular weight.

Endopeptidase: Enzyme that can cleave peptide bonds within the body of a protein chain.

Enzyme: Proteins with ability to catalyse the chemical reactions necessary in living organisms. Typically, enzymes are no smaller than 100 amino acid residues.

Epitope: Part of a protein's surface area liable to be recognised and bound to by an appropriate antibody. A linear epitope is composed of a single chain segment. An assembled epitope is formed from more than one chain segment.

Essential amino acids: Those required in the human diet.

Evolution: Tendency of proteins to have their sequence/conformation/function altered over time as a result of genetic mutation. The natural mechanism by which new protein functions and variations are produced and selected.

Exopeptidase: Enzyme that cleaves amino acid residues from one or other end of a protein chain

F

Fibrous protein: General name given to elongated, insoluble, fibre-like proteins that perform predominantly structural roles in organisms. E.g. keratin, collagen.

Fold: The arrangement, direction and course of the protein chain in 3D space. Characteristic for given protein families and usually highly conserved during evolution.

Folding pathway: Stages in which a newly synthesised protein chain acquires its final 3D shape.

Functional Genomics: The development and application of global (i.e. genome-wide or biosystem-wide) experimental approaches for monitoring and quantifying gene function, using the molecular structure/function knowledge and reagents provided by *structural genomics*.

G

Genetic code: Fixed correspondence between triplet base codons in mRNA and the amino acid type chosen to be incorporated in the growing protein chain during the process of translation.

Globular protein: General name given to water-soluble, roughly spherical proteins that may be easily denatured. E.g. enzymes, immunoglobulins.

H

Hairpin: Chain reversal between two adjacent beta strands.

Helix 3.10: Similar to alpha-helix but more tightly wound (3 residues per turn) and less commonly observed in natural proteins.

Helix pi: Similar to alpha-helix but more loosely wound (5 residues per turn) and rarely observed in natural proteins.

Homologue: As applied to proteins, a primary sequence whose resemblance to a previously characterised sequence is strong enough to suggest a common genetic ancestry

HPLC: High Performance Liquid Chromatography. Versatile method for the separation of peptide/protein mixtures, usually on basis of relative hydrophobicity. Notable for being successful with very small amounts of material.

I

Imidazole: Side chain group of histidine.

Imino acid: Contains secondary instead of primary amine function. e.g. proline

Indole: Side chain group of amino acid tryptophan.

Induced fit: Process of mutual, complementary adjustment that may take place when a protein docks with another molecule. May ensure selective recognition and strong binding. In the cases of some enzymes, may trigger correct catalytic geometry.

Inhibitor: Molecule capable of preventing an enzyme working, frequently by binding in or near the catalytic site and blocking the usual interaction with substrate.

Isoelectric point (pI): The pH at which a peptide/protein carries no overall charge.

J

K

Keratin: Structural protein found in skin, hair, horns, hooves, feathers, fur.

L

L-amino acid. Isomeric form of amino acid found throughout natural peptides and proteins.

Loop: Section of protein chain connecting classical secondary structures like helix and sheet. Often of irregular structure, surface exposed and containing hydrophilic residues.

M

Main chain: The fundamental peptide-linked “backbone” of a protein molecule.

Metalloenzyme: Enzyme which requires one or more metal atoms to be incorporated into its structure before it can show catalytic activity. E.g. carboxypeptidase A.

Molecular chaperone: A molecule (often a protein itself) that may assist another protein in the attainment and/or performance of the latter’s function. E.g. assistance in reaching the final folded state or guidance to an interactive site.

Molecular modelling: Computer simulation of protein molecular structure, variously designed to predict and display shape, calculate minimum energy conformations and dynamic ranges, predict recognition sites, binding orientations, etc.

Motif: A frequently observed 3D pattern of assemblage of secondary structures seen within different protein molecules, sometimes called a “supersecondary” structure.

N

N-terminus/terminal: That end of a protein chain that carries the free α -amino group.

NMR: Nuclear Magnetic Resonance. Technique for determining protein 3D structures in solution, detecting secondary structures and measuring conformational changes.

O

Oligopeptide: An imprecise term used to describe a peptide of around 6-10 residues.

One letter Amino Acid code: Preferred shorthand for recording an amino acid sequence in which each residue is represented by one letter of the alphabet.

A Ala; C Cys; D Asp; E Glu; F Phe; G Gly; H His; I Ile; K Lys; L Leu; M Met; N Asn; P Pro, Q Gln; R Arg; S Ser; T Thr, V Val; W Trp, Y Tyr.

Orthologue: As applied to proteins, a primary sequence from one organism that closely resembles a primary sequence from another organism because the proteins involved are products of an equivalent gene.

P

Paralogue: Term for a protein whose primary sequence resembles that of a second protein because it is translated from a *duplicate* of the gene for the second protein. Typically the case where two similar sequences are isolated from the same organism.

Peptide: Short chain of amino acid residues linked by peptide bonds.

Peptide bond: Formed by the condensation of an α -amino group of one amino acid with an α -carboxyl group of a second to eliminate water and covalently link the two together. Notable because new nitrogen to carbon bond is not free to rotate, generating cis- or trans-planar (most common) linkages. The fixed stereochemistry of the trans-planar peptide bond is a major determinant of polypeptide chain conformation.

Peptidase: Enzyme capable of hydrolysing certain peptide bonds in a polypeptide chain.

Primary structure: Linear sequence of amino acid residues in peptide/protein chain. By convention, sequences written with free N-terminal at left.

Prediction: As applied to proteins, attempts to predict aspects of shape, function and mechanism from basic information, such as knowledge of primary structure. Necessary because the production of gene/protein sequence data is faster and cheaper than direct experimental investigation.

Prosthetic: A non-protein group incorporated into some protein structures to add special chemical properties. E.g. metal atoms, oxygen binding groups.

Protein: long heteropolymers of amino acid residues linked by peptide bonds formed by the condensation of α -amino and α -carboxyl groups.

Protein engineering: The non-natural production of new protein variants, either as a result of modifying existing proteins or their genes, or de novo design.

Proteinase: Enzyme capable of hydrolysing certain peptide bonds in protein chains.

Proteome: An all-embracing term (covering quantity, range and type) to describe the population of protein molecules output by the genome of an organism, or, in the context of a multicellular organism, the population of protein molecules output by a specific type of tissue or differentiated cell.

Proteomics: The use of quantitative techniques to measure changes in the population of proteins output by an organism's genome during normal biological processes, disease states or in response to drugs. General aim is to relate gene expression to the characterisation and control of biological processes.

Q

Quaternary structure. Situation where a large protein is composed of 2 or more independent chains. Chains may be identical or different.

R

Ramachandran plot. Graph of permitted bond rotation angles either side of α -carbon for an amino acid residue contained within a polypeptide chain. Predicts that α -helix and β -conformation will be the most energetically favoured polypeptide chain conformations. One angle of rotation is called phi (α -carbon to nitrogen) and the other is psi (α -carbon to carbonyl carbon).

Random coil: "Wastebasket" classification for protein chain conformations that are not of the classic types. Are not truly "random" and may be highly controlled and specified.

Recognition site: Surface patch or pocket that is able to attract, bind and discriminate between other molecules.

S

Salt-link: Strong interaction between oppositely charged amino acid side chains (e.g. lysine/glutamic acid). May contribute to holding a protein fold together.

Secondary structure: Protein chain conformations that are energetically favoured as a consequence of the available bond rotation angles in the polypeptide backbone and hydrogen bonding between peptide bonds. Most common regular conformations are α -helix and β -structure.

Secondary structure prediction: Result of applying one of the many theoretical methods available to predict sections of helix, sheet and turn in a protein chain from sequence information alone. Typically 60-70% accurate.

Site-specific mutagenesis Alteration of a protein primary structure by artificial means (usually via the gene) to test mechanistic hypotheses or to alter activity.

Specificity: Frequently applied to enzymes as a descriptor of the limited range of substrates upon which the enzyme will act.

Structural Genomics: As arising from the genome sequencing projects, the drive to elucidate the 3D structure, function and mechanism of the proteins produced from the genes in order to apply the knowledge. Goals include: the determination of the shapes and actions of *all* human proteins, or of the *complete* sets of proteins in particular functional classes, such as *all* enzymes or *all* cell-surface receptors.

Substrate: Typically the molecule(s) that will be transformed by the catalytic action of an enzyme.

Supersecondary structure: See motif.

T

Tertiary structure: 3D shape formed by folding and packing together of the various secondary structures that the protein chain adopts

Topology: Shape and folding motif of protein chain in 3D space.

U

V

W

X

X-ray crystallography: Important technique for the deduction of protein 3D structure. Requires crystals of pure protein.

Y

Z

Zwitterion: Natural state of free amino acids in which α -amino group is protonated giving a positive charge and a carboxyl group is deprotonated giving a negative charge.