ORGANIC CHEMISTRY AND BIOLOGICAL SYSTEMS - BIOCHEMISTRY

M. Lotti
Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy

Keywords : amino acid; ATP; citric acid cycle; coenzymes; coupled reactions; enzymes; folding; glycolysis; hydrophobic effect; lipid; metabolism; nucleic acid; nucleotide; oxidative phosphorylation; post-translational modification; proteins; ribosome; structure; transcription; translation

Contents

UNESCO – EOLSS SAMPLE CHAPTERS

Summary

Life is based on the use of a small set of organic molecules. Amino acids, nucleotides, sugars and lipids are the major biochemical compounds and play diversified and interconnected roles in cells. They are the building blocks of key polymers in biochemistry, proteins, nucleic acids and polysaccharides, respectively but acts as well

as energy carriers, metabolic precursors, regulatory and signaling elements. This chapter focuses on the chemical and functional features of biochemical molecules and tries to introduce them in the general context of biological systems.

1. From Molecules to Living Systems: Complexity is Obtained from Simple Building Blocks

Living organisms are wonderful machines able to perform in a highly coordinated and regulated way thousands of chemical reactions at the same time and with high specificity in a very crowded molecular environment. A look at the fundamental unit of life – the cell – reveals the presence of thousands of different compounds of various levels of complexity. Each of them has been transported, synthesized or transformed by the cell apparatus itself. Self-organization, replication, regulation and complexity are often quoted as features distinguishing living from non-living systems. How this can be achieved at the molecular level is the subject of research of biochemistry and the first question arising is, of course, whether special molecules are required to perform the enormous range of biological functions and reactions. Life however achieves an impressive diversity using only a selected set of chemical components that might be eventually combined in polymers able to drive specific biological processes and to transfer information. A closer look at the chemistry of living beings shows that cells are built up to 99% by only four elements (C, H, N, O) whereas several others occur in traces, i.e. as cofactors and coenzymes. About a thousand of different types of small organic molecules are biologically relevant, most of them are chemically related and can be classified as amino acids, nucleotides, sugars and fatty acids. Such chemical entities have properties suited to the aqueous cellular environment (water accounts for about 70% of the cell's weight), in particular solubility. Accordingly, several of the organic molecules present in cells derive form carboxylic acids and organic amines and are hydrophilic. Water also plays a crucial role in determining the structure of macromolecules. The energetic advantage of shielding from water the hydrophobic groups is the driving force in the folding of proteins and determines the structure of nucleic acids and the organization of phospholipids in membranes, as it will be described in the following sections. Out of this set of possible precursors, smaller subsets have been selected as the building blocks to form the key polymers of biochemistry: proteins are made up by 20 α-amino acids, nucleic acids (DNA and RNA) by only 8 nucleotides and polysaccharides use a small fraction of existing sugar molecules. A picture emerges in which complexity is obtained by the combination of a small number of basic modules. The principles of life at the molecular level can be summarized as follows:

i) most macromolecules are polymers in which the **sequence** of building blocks contains **information** related to function and structure. Biological macromolecules are able of **self-assembly**;
ii) the **three-dimensional structure** is important for the biochemical function;
iii) biological reactions are catalyzed by specialized proteins, the **enzymes** that control synthesis, degradation and transport of almost every chemical compound in a cell.
iv) function of macromolecules can be **regulated** in a number of ways.

In living systems, order, structure and specific interactions can be seen as the basis of

UNESCO – EOLSS SAMPLE CHAPTERS

function. It was already mentioned that important biological molecules are polymers. Polymers can be either a repetition of identical monomers, random combinations of different monomer units or they can be characterized by a specific sequence. In such cases, not only the nature of monomers but also the order in which they appear in the polymer, have functional significance. The two groups of macromolecules in which sequence carries genetically encoded specific information are nucleic acids and proteins. Both are linear polymers of covalently joined monomers. Carbohydrates, on the other hand, can form highly branched polymers, hence containing biological information of importance, for example, in molecular recognition processes. The second level of organization is spatial. Three-dimensional architecture positions close to each other functional groups far apart in the sequence, builds up active sites, forms surfaces for interactions with other molecular components. In conclusion, biological systems combine simple organic molecules in such a way that the resulting object carries new and more complex functions.

Other key rules in biochemistry concern the energetic metabolism and the need of not wasting metabolic work and materials. The two main strategies employed are:

v) Energy-releasing and energy-consuming reactions are coupled through the production of special molecules: i.e. an exoergonic reaction is coupled with the synthesis of high energy molecules, i.e. ATP, and, conversely, endoergonic reactions are possible, thanks to their coupling with the hydrolysis of high energy molecules.
vi) Cell components are obtained using a restricted number of simpler precursors which are common to different metabolic pathways. For example, the metabolism of nucleotides and amino acids is interconnected through several common intermediates.

This chapter focuses on the chemical and functional peculiarities of “the molecules of life” and aims to put them in the general context of cell life. The same molecules in fact can play different biological roles as constituents of macromolecules, intermediates of metabolism, energy transporters.

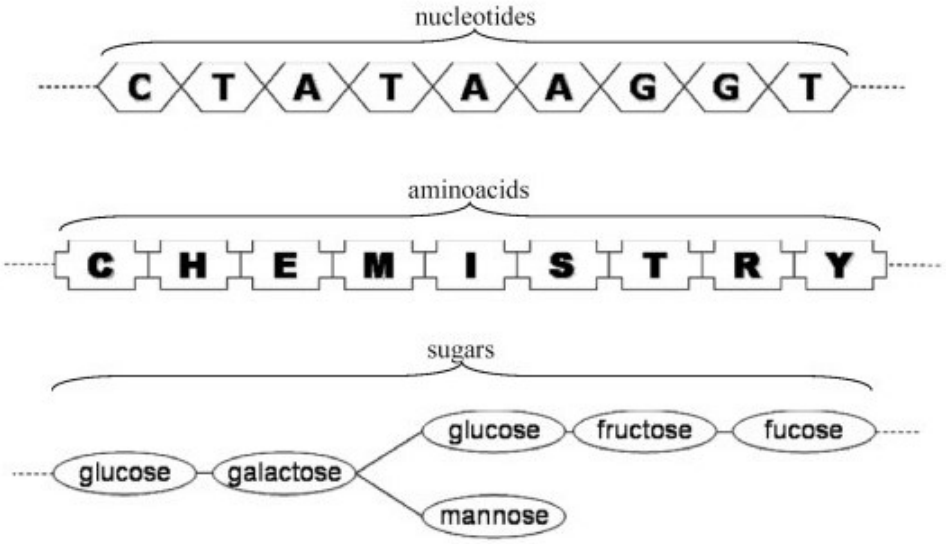

Figure 1: Polymers in biochemistry

2. Amino Acids and Proteins

Amino acids play a central role in the chemistry of living beings as they are involved in several functions: first of all they are the units building proteins, but they are also intermediates of different metabolic pathways. About 250 different amino acids have been isolated from living beings but only 20 out of them are used in proteins. How the impressive range of functions played by proteins can be achieved using this limited set of building blocks is the subject of this section.

2.1. Proteogenic Amino Acids

Proteins of all species, from bacteria to humans are made up by the same set of 20 α–amino acids belonging to the L- series, but other α-amino acids as well as β–amino acids and D-amino acids are present in cells with other functional roles. At a first glance, diversity of shapes and functions observed in polypeptides sounds astonishing as an α–amino acid is quite a simple molecule with a conserved basic structure in which a carboxylic acid group, an amino group and hydrogen are bound to a central carbon atom (Cα). Proline is the exception to the rule in that it has a secondary amino group. The fourth substituent, named the side chain (R), varies among amino acids. In other words, amino acids are molecules identical in their structure but in their side chains. Chemical and functional properties of amino acid monomers are determined by the presence of the amino and carboxylic groups and by those of the side chain. The first two functional groups are ionizable at appropriate pH: at low pH both are fully protonated but as the pH increases first the carboxyl group and then the amino group lose a hydrogen ion. Amino acids with ionizable side chains bear an additional group with a distinctive pK_a. In polypeptide chains all amino and carboxylic groups are involved in peptide bonds with the exception of those belonging to the first and last residue of the chain. The ionization state of the side chains is deeply affected by the polarity of the surrounding residues and their pK_a can greatly vary when compared to that of isolated amino acids. This property is important for protein function, i.e. for the catalytic mechanism of enzymes using acid-basic catalysis.

Amino acids, with the exception of glycine (R=H), have four different substituents arranged around the tetrahedral α-carbon atom and thus can exist in one of two stereoisomers or enantiomers, named L and D according to the convention introduced by Emil Fisher which refers to the configuration of glyceraldehyde stereoisomers. All α–amino acids contained in proteins have L stereochemical configuration.

The chemical nature of side chains is of paramount importance as it has consequence on the folding and 3D structure of protein, for the binding of various cofactors and metal ions, for providing functional and reactive groups.

Side chains are groups common in organic chemistry. A useful, although somehow arbitrary way of classifying amino acids is according to the polarity of their R groups as:

UNESCO – EOLSS SAMPLE CHAPTERS

- hydrophobic side chains enclosing glycine, alanine, valine, leucine, isoleucine, methionine, proline, cysteine and the three aromatic amino acids phenylalanine, tyrosine and tryptophan
- uncharged polar side chains enclosing aspargine, glutamine, serine and threonine;
- charged polar side chains enclosing arginine and lysine which are protonated and thus positively charged at neutral pH, histidine that can be either positively charged or uncharged at neutral pH and finally aspartic acid and glutamic acid negatively charged.

This and similar criteria, find their rationale in the importance of polarity for the folding, 3D structure, stability and function of proteins.

From the point of view of a chemist, side chains can be defined as: hydrocarbons (methyl in alanine, *i*-propyl in valine, butyl in leucine and isoleucine, benzyl in phenylalanine), carboxylic acids (glutamic acid and aspartic acid), amides (asparagine and glutamine), basic nitrogen side chains (lysine and arginine), hydroxyl groups (serine, threonine and tyrosine), sulfur-containing functional groups (methionine and cysteine), cyclic side chains containing nitrogen (hystidine, tryptophan).
Different classifications are not contradictory; they just take into account different physico-chemical properties. In the context of a polypeptide (see Section 2.3), side chains play key roles both in the folding of the polymer to acquire a functional (native) structure and in specific functions. Hydrophobic side chains do not provide functional groups and do not engage in hydrogen bonds but are important in the process of folding and are often buried in the protein core. Hydrophilic side chains make hydrogen bonds to other side chains, to the protein backbone and to various ligands and molecules and are important for stabilizing the secondary and tertiary structure (see Section 2.3.). Functional groups on side chains provide chemical and functional diversity and are involved, among other, in:

- Enzyme active sites: Aspartic and glutamic acid, histidine, cysteine, tyrosine and lysine side chains have pK_a values in the physiological range. This feature allows them to act as general acid and base catalysts. Moreover, the unique property of proteins to arrange several active groups in the proper proximity and relative orientation, provides the basis for concerted acid-base catalysis that is in fact a common catalytic strategy. Covalent catalysis, a mechanism that proceeds through the formation of a catalyst-substrate covalent bond relies on R chains with high nucleophilicity and, at the same time, ability to form good leaving groups. Residues endowed with such features are the hydroxyl group of serine or contain imidazole (histidine), thiol (cysteine), carboxyl (aspartate) functions.
- Sites for the binding of molecules or groups that improve or modify protein functions: Examples are i) specific points for the attachment of sugar chains provided by the amino group of asparagine (N-glycosylation) or the hydroxyl group of serine and threonine (O-glycosylation); ii) sites for the addition of phosphate groups, i.e. on serine, threonine and tyrosine. Phosphorylation is of paramount importance in the regulation of activity and function of several key proteins; iii) attachment of lipid molecules that confer to the target protein the ability to reversibily interact with membranes (see Sections 4.3. and 5.4.).

UNESCO – EOLSS SAMPLE CHAPTERS

Summarizing, in this section the properties of the 20 L-α-amino acids found in proteins have been reviewed and it was shown that the chemical nature of side chains can provide a large variety of functions. Nevertheless, some questions (why α-amino acids? Why L-α-amino acids? Why only these particular 20 L-α-amino acids?) are still unanswered.

α-amino acids seem more suited than β-amino acids to build up polypeptide chains able to spontaneously fold to a given shape. As it will be described in the following, assumption of regular structure depends on sterical and thermodynamic constraints that could not be satisfied by the use of the more flexible β-amino acids. How L-amino acids emerged as protein constituents is a stimulating question, as no obvious biological advantages can be envisaged in the use of one rather than the other stereoisomer. Structural considerations show that protein architectures such as those existing could not be obtained from random mixtures of L and D monomers. However, they might be generated using only D-amino acids. This has been experimentally shown by the chemical synthesis of the protease from the HIV virus using the D counterparts of the amino acids forming the natural sequence. The structure of the synthetic enzyme is the mirror image of the L-enzyme and is catalytically active on substrates containing D-amino acids. In principle, one could hypothesize that selection of L- stereoisomers occurred early during evolution and was consistent with the development of coherent synthetic and metabolic processes.

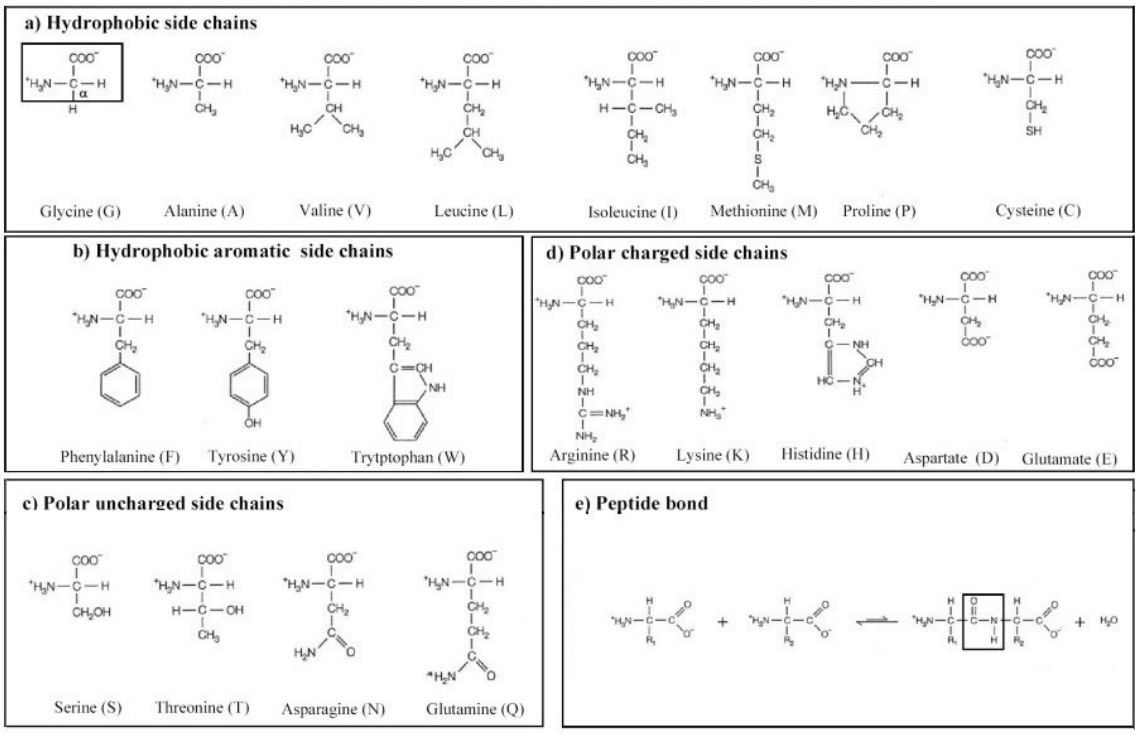

Figure 2: Standard amino acids (a-d). The basic common structure is boxed in glycin. In panel e the peptide bond joining two residues is highlighted

The use in proteins of this small set of amino acids is a consequence of the structure of

the genetic code, in which information for protein sequences is coded as groups of 3 nucleotide bases on the DNA molecule. As the code contains information for the 20 standard amino acids only these will be used in the process of protein biosynthesis (see Section 3.2.1.). However, other amino acids are present in proteins as the result of the post-translational (and therefore not genetically encoded) modification of side chains. Modified residues confer specific properties and are operated by specific enzymes. Thus, for example, in the structural fibrous protein collagen (present in connective tissue, skin, bones, tendons, etc.) proline and lysine side chains are modified by the addition of OH groups that contribute to additional hydrogen bonds thus improving the mechanical stability of this protein. A second carboxylic group can be added to glutamic acid residues in protein involved in blood clotting. This modification increases the negative charge of the protein and favors its association to the sites of damage.

2.2. Non Proteinogenic Amino Acids

Amino acids other than those selected in the course of evolution to constitute proteins play important biochemical roles. Some of them act as chemical messengers within and between cells: γ-aminobutyric acid (GABA) (derived from glutamate) and dopamine (from tyrosine) are neurotransmitters, histamine (from histidine) mediates allergic reactions, thyroxine (from tyrosine) is a thyroid hormone. Other amino acids are intermediates in metabolic processes, for example citrulline, ornithine and homocysteine participate in amino acids metabolism.

D-amino acids are represented in peptides of the bacterial cell wall that are synthesized enzymatically rather than on ribosomes. Interestingly, the presence of non standard amino acids makes such structures resistant to the attack of peptidases (enzymes that hydrolyse peptide bonds of proteins) produced by organisms as a protection toward bacterial aggression. Peptidases have been evolved to degrade proteins composed by standard amino acids and are less active against such unusual peptides. D-amino acids are also present in molecules with antibiotic action obtained from bacteria, such as valinomycin, gramicidin A, actinomycin D

2.3. Amino Acid Polymers: Proteins. Their Structure and Function

UNESCO – EOLSS SAMPLE CHAPTERS

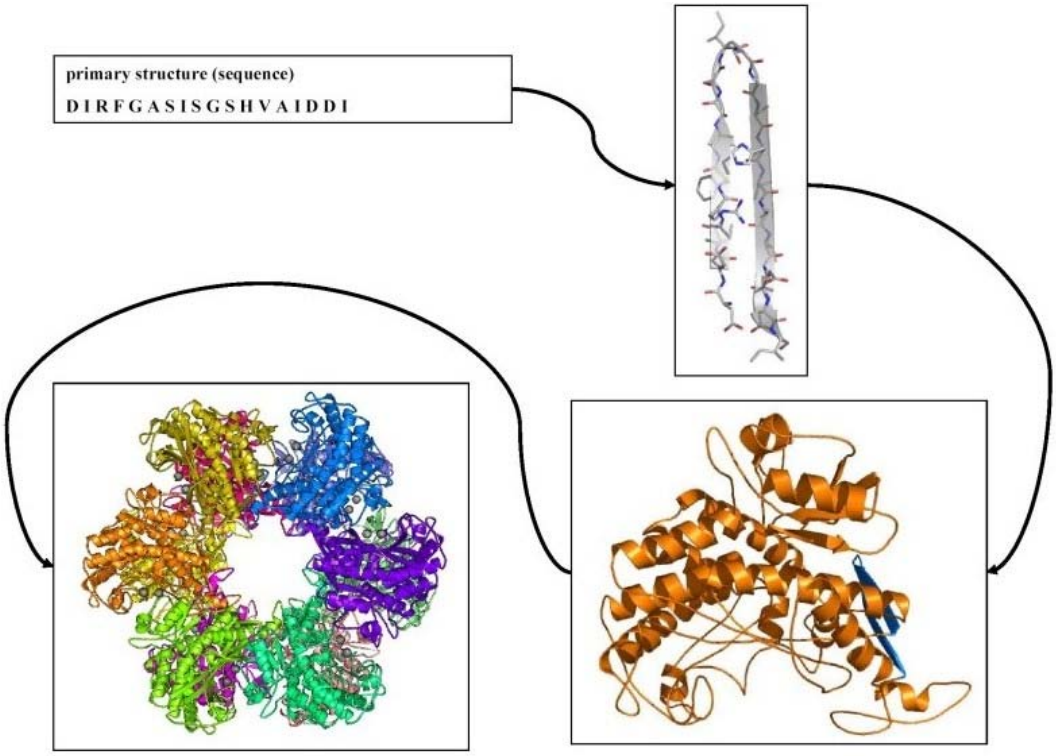

Figure 3: Levels of organization of a polypeptide chain. The protein displayed is glutamine syntethase which consists of 12 identical subunits

Proteins are linear sequences of amino acids linked together by an amide bond or peptide bond, formed between the α-carboxylic group of one amino acid and the α-amino group of the subsequent one, with loss of a water molecule. Having all amino acids the same backbone, a protein is a polymer repetitive in its basic structure (main chain or backbone) from which the side chains of amino acids protrude. It is described by the sequence of the monomers (**primary structure**) running from the amino terminus, where the amino group of the first amino acid of the chain is free to the carboxyl terminus with the last residue with a free carboxyl group. This reflects the order of residues addition during the process of protein synthesis. Protein's properties are completely specified in the succession of amino acids that is dictated by the genetic information in the DNA sequence and in turn determines form and function of the resulting protein. Polypeptide chains achieve their three-dimensional structure and, as a consequence, their function, through a spontaneous process of **folding**. Although information needed to fold is completely specified in the primary sequence, the rules governing this fascinating process and its kinetics are still not completely understood. Protein misfolding causes serious diseases such as Parkinson and Alzheimer diseases.

2.3.1. Proteins Folding and Structure

A major force driving protein folding is the so-called hydrophobic effect, i.e. the tendency to remove hydrophobic groups away from the solvent. In a way, protein folding can be considered as the result of the competition for the establishment of intra-chain or protein-solvent interactions. In water a denatured polypeptide (extended

conformation) composed by both polar and non polar amino acid residues is hydrogen bonded to water molecules through its polar groups, whereas hydrophobic ones are excluded from contacts with the solvent and disrupt the hydrogen bonded structure of water without providing compensating bonds. When a protein folds and achieves a compact three-dimensional structure, most bulky hydrophobic groups are shielded away from the solvent and buried inside the structure (hydrophobic effect). Folding has two energetic advantages: the hydrophobic surface in contact with the polar solvent is minimized and hydrophobic groups are brought close to each other so that they can establish van der Waals interactions. The stability of a folded protein depends on the balance between the energy of the interactions formed or broken in the transition from the unfolded to the folded conformation. Interestingly, the difference in the free energy of a protein in its native or denatured conformation is only about 20-40 kJ/mol. Marginal stability is not deleterious, on the contrary, flexibility is mandatory for protein function, i.e. to allow molecular recognition, such as interactions of enzymes and substrates and antibodies with antigens, and regulation of protein activity through binding to effectors or covalent modification.

In a folded protein, the only covalent linkages are peptide bonds of the main chain. Additional covalent bonds formed between the thiol groups of cysteine residues (disulfide bridges) can be present in proteins secreted to the exterior or enclosed in membranes. Protein structure is indeed stabilized by noncovalent weakly polar interactions such as hydrogen bonds, van der Waals forces, salt bridges, and long-range electrostatic interactions. Their force is in the range of 2-20 kJ/mole according to the nature of the residues involved, their distance and localization in the protein interior or at its surface, as well as on the surrounding protein environment.

Besides interactions with the solvents, the structure of a polypeptide chain is determined by constraints imposed by the features of the peptide linkage and of the side chains of the amino acids represented in the sequence.

The peptide bond exhibits a partial double bond character due to the closeness of the carbonyl carbon-oxygen double bond allowing for the existence of resonance structures. As a consequence, the carbonyl oxygen, the carbonyl carbon and the amide nitrogen participating in this linkage are coplanar and the polypeptide chain results composed by a succession of rather rigid planes. Rotation is only allowed around the N-Cα and Cα-C bonds, provided there is no steric interference among side chains. The angles of these two bonds to the adjacent peptide linkages are called Φ and Ψ torsion angle, respectively. Analysis of the three dimensional structures of proteins shows that the number of possible conformations a polypeptide chain can adopt is restricted to given combinations of torsion angles. When a sufficient number of residues adopt the same conformation, local regular structures, or **secondary structures**, originate. 2D structures common in proteins are right-handed alpha-helix (Φ= -60°, Ψ= -50°), β-sheets composed by β-strand elements (Φ= -120°, Ψ= +120°) and loops of various length and conformation connecting such elements. Residues in an alpha helix are consecutive in the primary sequence, the chain wound around a central axis and it is stabilized by the formation of hydrogen bonds between the carboxyl oxygen of a residue and the amide group of a residue four positions down in the chain. Right-handed helices have therefore 3.6 residues per turn. Residues side chains protrude from the structure

UNESCO – EOLSS SAMPLE CHAPTERS

and interact either with the solvent or with other regions of the protein. β-strands are extended structures and are joined to each other by hydrogen bonds between carboxylic and amide groups of two adjacent chains forming a pleated, twisted β-sheet with the side chains protruding above and below the plane. Secondary structure elements contribute to the stabilization of the protein fold through their extensive networks of hydrogen bonds and provide an elegant way for packing the hydrophilic protein backbone in the hydrophobic core of globular proteins. In fact, the polar groups of the main chain (N-H and carbonyl groups), that in the denatured conformation hydrogen bond to water molecule, in secondary structure elements are linked with each other.

Tertiary structure is obtained when the polypeptide fold in a compact structure stabilized by weak interactions between both polar and non polar groups. In the 3D structure, besides local interactions typical of the 2D elements, long-range interactions are established. The structure is further stabilized by a close packing of the atoms in the protein core. The number of tertiary structures or folds adopted by polypeptides is limited, though large, much smaller of the number of existing proteins. Proteins completely different in sequence and function in fact might adopt the same structural organization. One of the commonest folds is the so-called TIM barrel first described for the glycolytic enzyme triose phosphate isomerase. This consists of regular repetitions of β–α–β–α motives that produce a barrel-like structure with a core formed by the parallel β-sheet which is surrounded by α-helices. This fold occurs in about 10% of all known enzyme structure, without any relation to their function. It is possible that the alpha/beta barrel architecture has been selected and used in different protein families because of its stability. Whether particular folds have emerged because of physico-chemical constraints of the polypeptide chain or they have been just evolved from ancestral folds is not completely clear, although recent experiments of *de novo* design have shown that structures different from those present in today's proteins are possible.

To conclude, it should be recalled that many proteins self-associate into oligomeric assemblies composed by two or more identical subunits or even by different polypeptide chains hold together by the same weak forces involved in maintaining the tertiary structure (**quaternary structure**). Oligomeric proteins can perform different functions at different subunits or can gain new functional and regulatory properties. As an example, two proteins involved in oxygen- binding, myoglobin and hemoglobin can be considered. Myoglobin is monomeric and serves to store oxygen in muscles ready for metabolic requirements. Hemoglobin on the contrary is a tetramer of two different subunits. Its binding to oxygen is cooperative, that means that O_2 binding to one subunit increases affinity of the others. Affinity is modulated by the oxygen concentration and allows this protein to bind O_2 in lungs (where oxygen pressure is high) and to release it to myoglobin in the blood capillaries in the muscle (where oxygen pressure is lower). Allosteric enzymes are another example in which oligomerization allows for regulation. In this case, enzyme activity is regulated by molecules that are not directly related to enzyme substrates and products, termed allosteric effectors. In both oxygen-binding proteins and allosteric enzymes, regulation is mediated by transfer of conformational changes between subunits.

UNESCO – EOLSS SAMPLE CHAPTERS

-
-
-

TO ACCESS ALL THE **33 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Alberts B., Johnson A., Lewis J., Raff M., Roberts K and Walter P. Molecular Biology of the Cell, 4th edition, Garland Science, New York, 2002 [A text book providing deep insight into the molecular events occurring in cells]

Creighton T.E. Proteins. Structures and Molecular properties, 2nd Edition. W. H. Freeman and Company, New York, 1996 [The "classical" reference book to approach the study of proteins]

Dobson C.M., Gerrard J.A. and Pratt A.J. (2001) Foundations of chemical biology. Oxford Chemistry Primers, Oxford Science Publications [This is a very useful booklet to get insight into biochemistry starting from the point of view of a chemist]

Fersht A. R. Structure and mechanism in Protein Science. A guide to Enzyme Catalysis and Protein Folding. W.H. Freeman, New York, 1999 [A book that provides information about the molecular events at the basis of proteins function and folding]

Helenius A. and Aebi M. (2004) Roles of N-linked glycanes in the endoplasmic reticulum. Ann. Rev. Biochem.73: 1019-1049 [This is a recent review article on the biological impact of glycosylation]

Lewin B. Genes VIII. Pearson Education, 2004 [A widely used and always updated textbook of molecular biology]

Petsko G.A. and Ringe D. (2004) Protein Structure and function. New Science Press Ltd, London [This provides an original approach to the comprehension of proteins' function]

Protein misfolding. Nature Insight (2003) Nature 426: 883-909 [A special issue entirely devoted to problems and diseases related to protein misfolding]

Rudd P.M. and Dwek R.A. (1997) Glycosylation: heterogeneity and the 3D structure of proteins. Crit. Rev. Biochem. Mol. Biol. 32: 1-100 [A review article on the effects of post-translational modifications by sugars on protein structure]

Steitz T.A. and Moore P.B. (2003) RNA. The first macromolecular catalyst: the ribosome is a ribozyme. Trends Biochem. Sci. 28: 411-418 [A review article highlighting the complexity of ribosomes and the catalytic role of RNA molecules]

Trombetta E.S. (2003) Glycobiology 13: R77-91 [A review article on the biological role of sugars]

Voet D. and Voet J. G. Biochemistry, 3rd edition. John Wiley & Sons, Inc. New York, 2004 [An updated textbook of biochemistry particularly useful for the study of macromolecules]

Watson J.D., Baker T.A., Bell S. P., Gann A., Levine M. and Losick R. Molecular Biology of the gene, 5th edition Pearson Education Inc., 2004 [A textbook of molecular biology providing information on the regulation of gene expression]

Yonath A (2005) Ribosomal crystallography: Peptide bond formation, chaperone assistance and antibiotics activity. Mol. and Cells 20 (1): 1-16 [This is an updated review on protein synthesis at the molecular level]

UNESCO – EOLSS SAMPLE CHAPTERS

Biographical Sketch

Marina Lotti studied Biology at the University of Milano and carried out her PhD Thesis at the Max Planck Institute for Molecular Genetics in Berlin (Germany) working on ribosomes structure. She moved then to the National Research Council and later on to the State University in Milano where she started to work in the field of heterologous proteins expression and industrial enzymology. She is presently Professor of Biochemistry at the University of Milano-Bicocca where she is the group leader of the Laboratory of Protein Engineering and Enzymology. She is the author of several papers on international journals and book chapters mainly concerning the investigation of structure function relationships in proteins, studied with biochemical, modeling and biophysical approaches. Current scientific interests focus on the modulation of enzyme's properties by protein engineering and on the stability and propensity to aggregation of proteins *in vivo* and *in vitro*.

UNESCO – EOLSS
SAMPLE CHAPTERS