Krishna Mohan Poluri
Khushboo Gulati
Sharanya Sarkar

Protein-Protein Interactions

Principles and Techniques: Volume I

Springer

Protein-Protein Interactions

Krishna Mohan Poluri • Khushboo Gulati •
Sharanya Sarkar

Protein-Protein Interactions

Principles and Techniques: Volume I

Springer

Krishna Mohan Poluri
Department of Biotechnology
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand, India

Khushboo Gulati
Department of Biotechnology
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand, India

Sharanya Sarkar
Department of Biotechnology
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand, India

ISBN 978-981-16-1593-1 ISBN 978-981-16-1594-8 (eBook)
https://doi.org/10.1007/978-981-16-1594-8

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Dedicated to all our teachers, mentors, and parents for their invaluable guidance and unconditional support throughout our research careers

Preface

Mother Nature has bestowed a variety of functions to the cell by sculpting distinct variety strings of proteins in which the amino acids were systematically positioned as beads of a chain. Proteins perform their functions either alone or in conjunction with several biomolecules. Most of the cellular processes are conducted via complex interactions among proteins. These interactions are well known as protein–protein interactions (PPIs). To obtain deep insights into such functional processes of cell and modulate them in a variety of pathophysiological conditions, it is essential to understand the basic concepts underlying PPIs, which opened an emerging era of research that focuses on PPIs.

PPIs are highly prominent in terms of their magnitude and diversity. It has been established that proteins interact with multiple protein partners and these differentially interacting proteins together form an interaction network. This is an outcome of evolvement of high throughput experimental techniques that are able to decode the criteria followed by the proteins during the course of their interactions with other protein partners. This phenomenon has also stimulated computational biologists to formulate novel computational approaches for the prediction of PPIs, tools to predict missing connections in PPI network, to identify distinctive roles of each PPI network component, and databases to store and organize the ample amount of PPI data generated. Hence, studying PPIs is a multidisciplinary endeavor. To gain better understanding of PPIs and for the development of drugs targeting specific PPIs, it is essential to analyze basic physicochemical properties of PPIs, their evolutionary profiles, thermodynamic and kinetic parameters associated with them, and additionally the experimental and computational means to unravel them.

Keeping in mind the contemporary requirements for studying PPIs, the current book is designed. This book comprises six chapters, each of them discusses a unique perspective about this briskly evolving field of protein–protein interactions (PPI). Chapter 1 presents an overview of the fundamental concepts pertaining the protein's structure–function–folding paradigm and also provides an introduction to the evolution of proteins and glimpse of the PPI field. Chapter 2 presents a comprehensive knowledge about the classification of PPIs, delineates their functional significance for the sustenance of life, and discusses all the parameters employed for analyzing a specific PPI. Chapter 3 provides an outline about the basic principles of energetics and majorly aims at describing the thermodynamic and kinetic parameters associated with PPIs. Several mutagenesis related case studies have also been discussed to

provide glimpse of high affinity protein–protein interactions. Chapter 4 sheds light on the protein complexes that are formed as a resultant of interactions among the proteins. This chapter describes the mechanism of formation of protein complexes, their structures, and applications. Evolutionary concepts involved in the formation of these protein complexes in the living cell are also outlined. Chapter 5 focuses on the wide spectrum of experimental techniques employed to decipher interactions among the proteins. Such techniques not only aid in discovering novel protein–protein interactions but also assist in validating the hypothesized interactions. This chapter also offers one to understand the basic principle of each technique along with its strengths and limitations. Chapter 6 is devoted to discuss computational approaches currently being employed for the prediction of PPIs and their analysis. It also provides a comprehensive view of databases designed to embrace differential aspects related to a variety of PPIs.

This book is intended to provide thorough understanding of the egressing field of protein–protein interactions. The book is designed to be self-contained as it offers an outline about the fundamental concepts related to PPIs, an in-depth knowledge of various experimental and computational approaches for the prediction and development of PPI libraries. Hence, this book is of great interest for variety of researches, students, and academicians who are seeking to gain comprehensive understanding and an up-to-date knowledge about PPIs.

Roorkee, India

Krishna Mohan Poluri
Khushboo Gulati
Sharanya Sarkar

Acknowledgements

KMP greatly acknowledges the support of Grants CRG/2018/001329 from Science and Engineering Research Board (SERB), DBT-IYBA fellowship-BT/07/IYBA/2013-1 from Department of Biotechnology, and GKC-01/2016-17/212/NMCG-Research from NMCG-MoWR, Government of India (GoI). The authors are also immensely thankful to Ms. Nupur Nagar and Ms. Shweta Tripathi of Department of Biotechnology, IIT Roorkee, for their sincere assistance in preparation of figures and/or proof reading the chapters contained in this book.

Contents

About the Authors

Krishna Mohan Poluri is currently working as an Associate Professor at the Biotechnology Department, IIT-Roorkee. He completed his PhD at Tata Institute of Fundamental Research, TIFR-Mumbai and Postdoc at Rutgers University and the University of Texas Medical Branch (UTMB-Texas). His areas of expertise are protein biochemistry and biophysics, bio-molecular interactions, structural biology, glycoimmunology, bionanotechnology, and algal biotechnology. He has published more than 110 articles and editorials in various respected international journals, books, and book chapters. He is also a guest editor/editorial board member for several international research journals. He won numerous awards and fellowships for his research work including; young scientist awards from the Indian Science Congress Association (ISCA) in 2009, National Academy of Sciences India (NASI) in 2014, and the DBT Innovative Young Biotechnologist Award (IYBA) in 2013.

Khushboo Gulati holds a PhD in Protein Biophysics from the Department of Biotechnology, Indian Institute of Technology (IIT), Roorkee. She completed her bachelor's in Bioinformatics and master's in Biotechnology at Punjab Panjab University, India. Her research interests are in the area of protein–protein interactions, protein engineering, molecular biophysics, and structural biology. She has authored 18 peer-reviewed research articles in SCI indexed international journals, 6 book chapters, and a book on protein engineering techniques. She is a recipient of the ICMR young scientist award.

Sharanya Sarkar is currently working as a Research Associate at the Mechanistic Biological Chemistry Laboratory, Department of Biotechnology, Indian Institute of Technology (IIT), Roorkee. She completed her master's degree in Biotechnology from IIT Roorkee in 2019 and her bachelor's in Microbiology from St. Xavier's College, Calcutta, in 2017.

Structural and Functional Properties of Proteins

1

Abstract

Proteins are one of the most fundamental biomolecules present in living organisms. They are indispensable to the sustenance of these organisms owing to the multitude of functions that they perform. They serve as structural constituents of cells and are involved in cell signaling, transport of biological components, cell adhesion, eliciting immune responses, catalyzing chemical reactions, etc. In order to get better insights into these functional aspects, it is quintessential to understand not only the physicochemical features of proteins, but also their structural organization (starting from primary sequences and ending in quaternary structures), and the various interactions that stabilize the higher-order structures. Many biological functions in the cell are mediated when a protein interacts with one or two other proteins in a phenomenon known as protein–protein interaction (PPI). The chapter concludes by providing an overview of PPIs including their functional importance.

Keywords

Transcription · Translation · Ramachandran map · Structural organization · Energy landscape · Biomolecular engineering · Protein folding · Non-covalent interactions

1.1 Introduction

Proteins are the ultimate players in the biochemical factory of cell. They are the work horses of cell which actively participates in almost all the cellular events. Proteins are highly interesting molecules as they are the most generous macromolecules in cell.

K. M. Poluri et al., *Protein-Protein Interactions*,
https://doi.org/10.1007/978-981-16-1594-8_1

Moreover, they are also highly diverse in contrast to any other macromolecules. Distinct variety of proteins can range from small or large, fibrillar or globular, monomers or oligomers, hydrophilic or hydrophobic, fixed conformation or alternative conformations, existing alone or associated with other macromolecules including carbohydrates and/or ligands, etc. Such diversity in proteins is the direct outcome of differential arrangements of amino acids. Since amino acids are the building blocks of proteins, they act as the beads of the protein string. Further, cellular events such as post-translational modification and gene splicing also add complexity to the diversity of these entities. Such a variation results in different structures of proteins which ultimately make them amicable to perform differential functions (Goodsell 1991).

1.2 Functions of Proteins

Proteins play numerous biological roles in cell including maintaining cell shape, structural organization, act as catalysts, molecular sensors, help in cell interactions, cell movement, provide defense mechanism to cell, etc. (Fig. 1.1). Different proteins exhibiting variety of functional roles are discussed in the following paragraphs.

Enzymes are the proteins that exclusively catalyze the biochemical processes in a given living organisms. Reaction is catalyzed when there is a binding of substrate in the protein's active site. In case of large substrates, there is more than one binding site in the enzyme. Each binding site binds to the different part of the substrate. However, catalysis occurs in one of those sites. For many reactions, such as for transferring a chemical group to the substrate, transferring electron, or substrate polarization, etc., enzymes require small molecules or metallic elements. When these molecules are permanently attached to the enzymes, these groups are known as prosthetic groups. In contrast, when attached temporarily, they are known as co-enzymes. Enzymes are broadly classified into six different categories, namely: oxidoreductases, hydrolases, transferases, lyases (syntheses), isomerases, and ligases (synthetases) (Silverman 2000). Each of these enzymes carries out specific chemical reaction.

Oxidoreductases aid in catalyzing redox reactions that mediate transfer of electrons from one molecule to the other. Such reactions are mostly involved in extracting out energy from the food stuff or in utilization of energy in the synthesis of molecular complexes. *Transferases* are involved in transferring a chemical moiety from one molecule to the other molecule. For instance, the transfer involves variety of chemical groups such as carbonyl, acyl, amino, glycosyl, methyl, sulfate, phosphate, or selenium containing groups. *Hydrolases* use water as an attacking group to cleave covalent bond in the substrate. Several types of bonds are attacked such as ester, amide, glycoside, and C-C bond. These enzymes exhibit low specificity toward the substrate. For instance, phosphatases remove the phosphate group from the substrate. *Lyases* catalyze the formation or cleavage of covalent bonds in the substrate without utilizing water. When the reaction involves the formation of bond the enzyme is known as synthase. *Isomerases* catalyze the movement of the

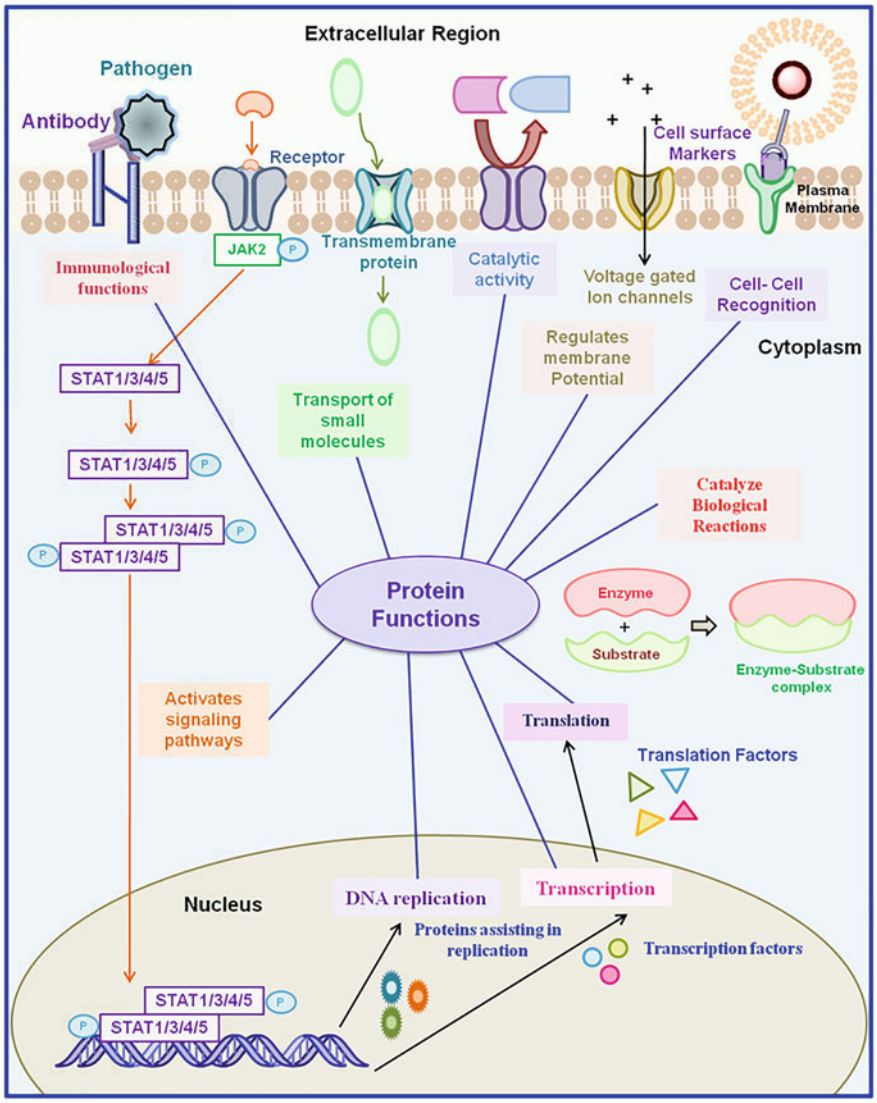

Fig. 1.1 Schematic showing the different functional roles of proteins

chemical moieties from one position to the other in the substrate. Isomerases are termed in accordance with the specific reaction they carry out. For example, epimerase, phosphomutase, and racemase belong to isomerase group, however they carry out specific reactions. *Ligases* catalyze the covalent binding of two substrate molecules for the formation of a single product. Such enzymes have traditional naming, for example, synthase, carboxylase, synthetase. DNA ligase is involved in

ligating the two DNA strands via catalyzing the formation of phosphodiester bond (Moss 2020).

Membrane proteins are broadly divided into two categories, namely: integral (intrinsic) or peripheral (extrinsic) membrane proteins. Most of the bio-membranes encompass both types of membrane proteins. Intrinsic proteins, also termed as integral membrane proteins, anchor via two or more fragments across the phospholipid bilayer of the membranes. These proteins comprise hydrophobic amino acids, whose side chains are involved in interaction with the fatty acyl groups in the membrane. Most of the proteins completely span across the phospholipid bilayer and also exhibit domains flanking on both extracellular and cytosolic sides of the membrane. They are known as *transmembrane proteins*. In contrast, some of the integral membrane proteins are attached through one membrane via covalently bound fatty acids. Peripheral membrane proteins are also termed as extrinsic proteins; they do not span the membrane and are also not involved in interaction with the hydrophobic core of the phospholipid bilayer. They are present on the surface of the membrane. They are either directly interacting with the lipid's polar head groups or indirectly interacting with integral membrane proteins. Membrane proteins carry out various diversified roles. Some of them maintain cell shape, while others are involved in cell adhesion and cell-to-cell communication. Membrane proteins also assist in triggering the signaling pathways once they receive the signals (Tan et al. 2008).

Antibodies are also known as immunoglobulins. They are Y-shaped proteins secreted mostly by plasma cells that neutralize the pathogens. They are divided into five major classes on the basis of types of the heavy chains in their structures. They are namely IgG, IgM, IgA, IgD, IgE. All of them exhibit differential immunological functions (Schroeder and Cavacini 2010).

Variety of other types of proteins such as translation machinery proteins, transcription factors, polymerases, helicases, etc. are also present in the cell which participates in differential cellular processes such as vesicular trafficking, transcription, replication, translation, etc. Such flexibility and diversity in the functions of proteins is due to their structural features that in turn depend on their amino acid sequence, oligomerization, dynamic features, and folding (Voet and Voet 2011).

1.3 A Journey from Genes to Proteins

DNA (de-oxyribonucleic acid) is a long linear polymer of nucleotides that carries information of the cell and can be transferred from one generation to the other. Nucleotides are made of three components, namely: deoxy-ribose sugar, nitrogenous bases, and phosphate group. They have been classified into four different types on the basis of their constituting bases. These nitrogenous bases are of four different types, namely: adenine (A), cytosine (C), guanine (G), and thymine (T). Nucleic acid chain is formed by the sequential arrangement of these nitrogenous bases which encode the genetic information. This genetic information is coded to RNA and that process of formation of RNA is known as *transcription*. RNA (ribose nucleic acid)

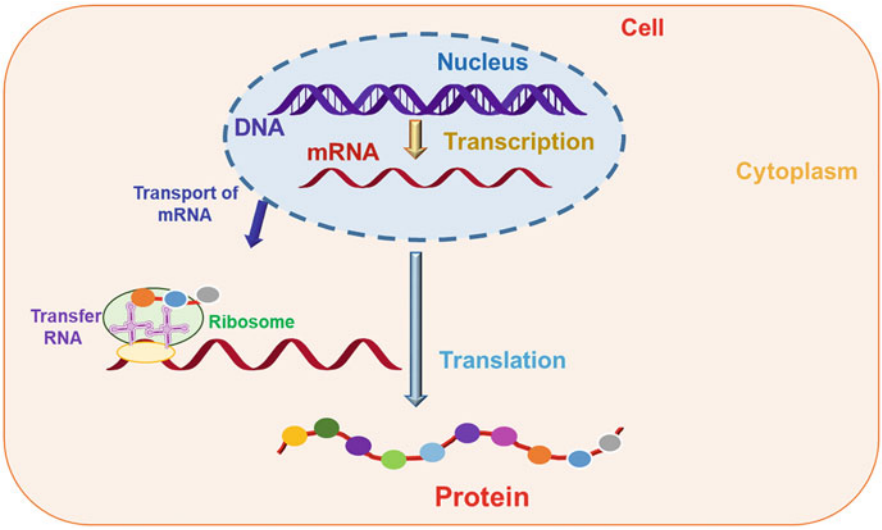

Fig. 1.2 Illustration depicting the flow of genetic information form DNA to RNA to protein termed as "central dogma" of molecular biology

uses codons to carry this information and ultimately decodes this message in the process of *translation*, whereby the RNA codes for distinct amino acids that results in the formation of the polypeptide chain of protein. The whole process of transfer of information from DNA to RNA and RNA to proteins is termed as "central dogma" (Fig. 1.2). There are 64 possible codons which code for 20 distinct amino acids. This difference in number of codons and amino acids results in degeneracy in the genetic code. This degeneracy implies that more than one codon is encoding for an amino acid. Among 64 possible codons, three codons including UAA, UAG, UGA do not code for any of the amino acid and are known as stop codons. These codons are added at the end to terminate the growth of polypeptide chain during the translation process.

Translation involves three distinct forms of RNA, namely: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). mRNA contains the information that directs the amino acid sequence of a polypeptide chain. The association of amino acid into a polypeptide chain is performed in rRNA. tRNA aids in transferring amino acids from cytosol to the developing chain of polypeptide in accordance with mRNA's triplet code. The process of adding amino acids continues till the stop codon has appeared in mRNA. Finally, the ribosome releases the complete polypeptide chain for further processing (Crick 1970).

1.4 Amino Acids: Beads in the String of Proteins

Proteins are made of combination of one or more polypeptide chains. A polypeptide chain is a polymer of amino acid residues which are linked with each other through the formation of peptide bond which is also known as CO-NH linkage. Such a linkage is made as a result of removal of water elements from α-amino moiety of one of the amino acids and α-carboxyl group of the other amino acid residue. There are around 20 naturally occurring amino acids that are frequently present in proteins. Asparagine was the first amino acid that was discovered in 1806. The last was the threonine which was discovered in 1938. Different polypeptides differ in the sequence and composition of their amino acids. These polypeptides or proteins can be hydrolyzed into their constituting amino acids (Nelson and Cox 2013).

1.4.1 Structural Characteristics of Amino Acids

All the 20 amino acids are also termed as α-amino acids since each of them exhibits a carboxyl moiety and an amino moiety that are bonded to the single α-carbon atom. Amino acids vary because of their R group or side chain groups (Fig. 1.3). R groups differ in their structure, charge, size, and also affect the amino acid's solubility in water. In addition to 20 amino acids, some other amino acids that are either modified after the formation of protein have also been observed or they are present in the organism but are not a part of proteins. All the standard 20 amino acids are assigned with a 3-letter abbreviation and a single code letter symbols (Table 1.1 and Fig. 1.3).

All the amino acids except glycine have four distinct groups, namely: carboxyl group, amino group, hydrogen atom, and R group. Glycine in place of R group contains hydrogen atom. Hence, the α-carbon in rest of 19 amino acid residues is a chiral center which makes them optically active. The four different groups can be organized in two different manners that are non-superimposable mirror images of each other. These two forms are class of stereoisomers (D and L forms of amino acids) known as enantiomers. All the amino acid residues that are observed in proteins are in L form. D form amino acids are present in minute quantities mostly in small peptides, such as those found in cell wall of bacteria and some of the antibiotics that are based on peptides.

1.4.2 Classification of Amino Acids

Amino acids are divided into three different groups on the basis of the nature of their R groups. They include: (a) non-polar, (b) uncharged polar, and (c) charged polar amino acids.

1.4.2.1 Non-Polar Amino Acids

Nine amino acid residues encompass non-polar R groups. Six amino acid residues, namely: alanine, glycine, methionine, isoleucine, leucine, proline, valine, exhibit

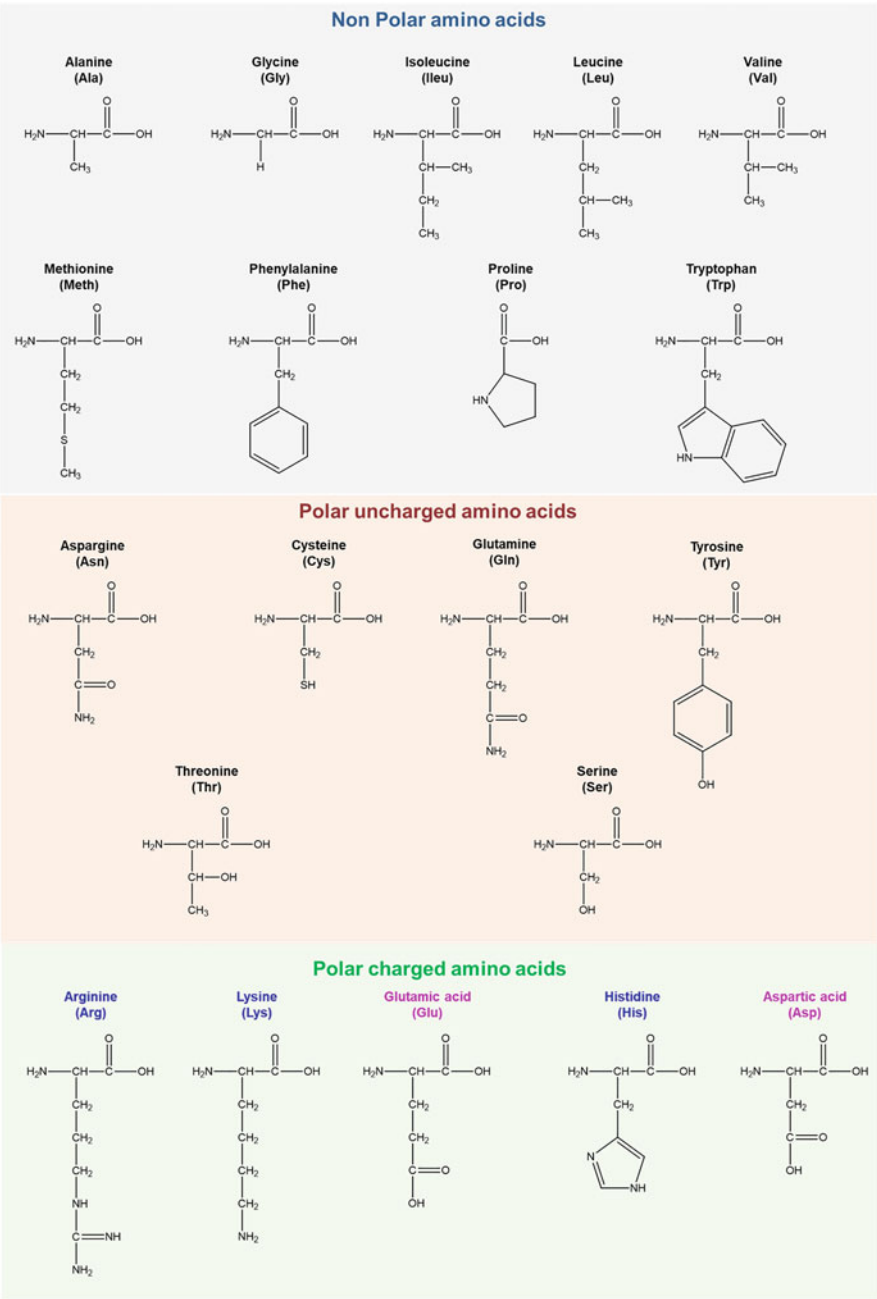

Fig. 1.3 Schematic depicting the structures of 20 standard amino acids, and their three-letter codes are mentioned in parenthesis

Table 1.1 Demonstrating the distinctive physiochemical characteristics of 20 standard amino acids which are categorized on the basis of their charge. Single letter codes of each amino acid are also shown in parenthesis

Nature	Amino acid	Codons	Molecular weight	Hydropathy	pKa	P(α)	P(β)
Non-polar	Alanine (A)	GCA, GCG, GCU, GCG	71.09	1.8	–	1.45	0.97
	Phenylalanine (F)	UUC, UUU	147.18	2.8	–	1.12	1.28
	Glycine (G)	GGA, GGC, GGG, GGU	57.05	−0.4	–	0.53	0.81
	Isoleucine (I)	AUA, AUC, AUU	113.16	4.5	–	1.00	1.60
	Leucine (L)	UUA,UUG,CUA, CUC, CUG,CUU	113.16	3.8	–	1.34	1.22
	Methionine (M)	AUG	131.19	1.9	–	1.2	1.67
	Proline (P)	CCA, CCC, CCG, CCU	97.12	−1.6	–	0.59	0.62
	Valine (V)	GUA,GUC, GUG, GUU	99.14	4.2	–	1.14	1.65
	Tryptophan (W)	UGG	186.21	−0.9	–	1.14	2.29
Polar (charged)	Aspartic acid (D)	GAC, GAU	115.09	−3.5	3.9	0.98	0.80
	Glutamic acid (E)	GAA,gag	129.12	−3.5	4.07	1.53	0.26
	Histidine (H)	CAC, CAU	128.17	−3.2	6.04	1.24	0.71
	Lysine (K)	AAA, AAG	128.17	−3.9	10.79	1.07	0.74
	Arginine (R)	Aga,AGG,CGA,CGC,CGG,CGU	156.19	−4.5	12.48	0.79	0.9
Polar (uncharged)	Cysteine (C)	UGC,UGU	103.15	2.5	8.35	0.77	1.30
	Asparagine (N)	AAC, AAU	114.11	−3.5	–	0.73	0.65
	Glutamine (Q)	CAA,cag	128.14	−3.5	–	1.17	1.23
	Serine (S)	AGC, AGU, UCA,UCC,UCG,UCU	87.08	−0.8	16	0.79	0.72
	Threonine (T)	Aca,ACC,ACG,ACU	101.11	−0.7	16	0.82	1.2
	Tyrosine (Y)	UAC, UAU	163.18	−1.3	–	0.61	1.29

aliphatic non-polar side chains. The side chains of alanine, isoleucine, leucine, and valine are involved in stabilizing the proteins via hydrophobic interactions. Glycine with simple hydrogen as R group does not really make any contribution to the hydrophobic shell of the protein. Methionine contains thio-ether group in its side chain. Proline contains pyrrolidine ring which contributes to the reduction in the structural flexibility of the protein. Phenylalanine, tryptophan, tyrosine are aromatic amino acids with relatively non-polar side chains. Tryptophan and tyrosine are relatively more polar than phenylalanine, owing to the tryptophan's nitrogen-containing indole ring and hydroxyl group of tyrosine. Tryptophan, tyrosine, and to some extent phenylalanine absorb UV light, this feature is utilized by many researchers to characterize the protein's structure-stability-functional aspects.

1.4.2.2 Uncharged Polar Amino Acids

The amino acids, namely: asparagine, glutamine, threonine, serine, and cysteine, fall under this category. These amino acids show more water solubility than the non-polar amino acids. This is due to the presence of groups that can involve in hydrogen bonding with water. Hydroxyl groups of threonine and serine, amide groups of glutamine and asparagine, and sulfhydryl group of cysteine contribute to the polarity of these amino acids. Cysteine gets readily oxidized to form cystine by forming disulfide bonds. Cysteine is highly hydrophobic and is mainly involved in joining different parts of the proteins or for connecting two different polypeptide chains by a covalent disulfide linkage.

1.4.2.3 Charged Polar Amino Acids

Charged amino acids are highly hydrophilic in nature. Out of 20, there are five amino acids that are charged. Out of five, there are three residues with positive charge, namely: histidine, lysine, and arginine. Histidine with its imidazole moiety, lysine with its aliphatic side chain containing amino group at ε position, and arginine with guanidino group are highly positively charged at pH 7.0. Remaining two amino acids, namely: glutamic acid and aspartic acid with side chain carboxyl groups, are the negatively charged amino acids that remain negative at pH 7.0.

1.4.2.4 Uncommon Amino Acids

Beside 20 amino acids, many other unusual or uncommon amino acids have also been observed. Some of these amino acid residues are the modifications of the common amino acids. For instance, 5-hydroxylysine and 4-hydroxyproline are derived from lysine and proline, respectively. 4-hydroxyproline is present in cell walls of protein. However, both the 4-hydroxyproline and 5-hydroxylysine are present in collagen, which is an abundant fibrous protein found in connective tissues. Carboxyglutamate is present in prothrombin and also in other calcium-binding proteins. Desmosine is observed in elastin. Selenocysteine is introduced in the protein during the synthesis of proteins rather than by the post-translational modifications. In this amino acid sulfur in cysteine is replaced by selenium 6-N-methyllysine which is present in contractile muscle protein known as myosin. Many other uncommon amino acids with variety of functions are present in cell but they

are not present as the part of proteins. For example, citrulline and ornithine are important intermediates during the arginine biosynthesis and urea cycle.

1.4.3 Acidic and Basic Properties of Amino Acids

Amino acids dissolved in water exit as dipolar ions or zwitter ions in the solution. This implies that they can act as both acid and base (proton donor as well as proton acceptor). Such dual nature entities are amphoteric in nature and are known as ampholytes. pk1 denotes the α-carboxylic group values which lie in the range of 2.2 indicating that it stays in the form of carboxylate ion when the pH is above 3.5. In contrast, pk2 refers α-amino group values which lie in the range of 9.4 implying that it remains as ammonium ions when the pH is below 8.0. pkr denotes the values for different side chains of amino acids. pKa values of all the 20 amino acids are summarized in Table 1.1.

1.4.4 Hydrophobicity Profiles of Amino Acids

Hydrophobicity is a measure of affinity of interaction between water/polar solvents and side chain of amino acids. Different nature of side chains of proteins attributes their differential interaction with polar solvents. Hydropathy is the measure of exact relative values of hydrophobicity or hydrophilicity of amino acids. All the non-polar amino acid residues with lower affinities for water are hydrophobic in nature and exhibit higher positive values in the hydrophobicity scale as compared to other polar charged/uncharged amino acids. Hydropathy scores of all the amino acid residues are briefly described in Table 1.1 (Kyte and Doolittle 1982).

1.5 Structural Hierarchy of Proteins

Proteins are the large complex macromolecular entities. They are made of different types of polypeptides with distinct amino acid sequences which are folded in a fashion to give biologically and functionally active proteins (Edison 2001). Hence, the structures of these complex entities have been defined at four different levels, namely: primary, secondary, tertiary, and quaternary structures (Poluri and Gulati 2017f). Primary structures define the amino acid sequence. Secondary structure denotes the arrangement of amino acids in a repetitive structural fashion. Tertiary structure implies the folding of polypeptide chains in three dimensions. Quaternary structures imply the arrangement of polypeptide chains in space.

1.5.1 Primary Structure of Proteins

Primary structure of protein is defined as the linear organization of amino acid residues that are linked to each other through the formation of covalent peptide bonds (Fig. 1.4). These peptide bonds are rigid and planar owing to the resonance or partial sharing of pair of electrons between amide nitrogen and carbonyl oxygen. These peptide bonds are devoid of free rotation due to the partial double bond character. Peptide bonds mostly exist in "trans" conformation. Polypeptide backbone is allowed to rotate about two bonds, namely: Cα-N and Cα-C. The angles about these bonds Cα-N and Cα-C are known as Φ (phi) and Ψ (psi) torsion angles. These angles dictate the conformational state of a polypeptide chain. When the conformation of polypeptide chain is fully extended, the values for torsion angles are 180°. G. N. Ramachandran pioneered a two-dimensional graph termed as Ramachandran plot which defines the energetically favorable or sterically allowed values of Ψ and Φ that a polypeptide can adapt (Fig. 1.5). 75% of the region in graph denotes the inaccessible conformations of the polypeptide. Ψ and Φ values for all the secondary structures in proteins lie in the conceded regions of the Ramachandran plot. Some residues that are involved in protein folding compensate for steric clashes, hence their value lie in the forbidden regions of Ramachandran plot. Ψ and Φ values for all the secondary structure are represented in Table 1.2.

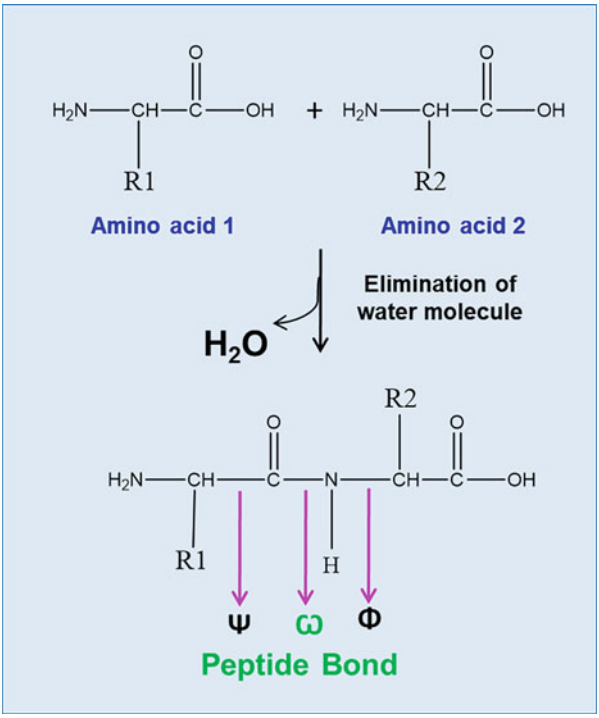

Fig. 1.4 Schematic showing the covalent peptide bond formation between the two amino acid residues; the backbone torsion angles (Ψ and Φ) are also shown

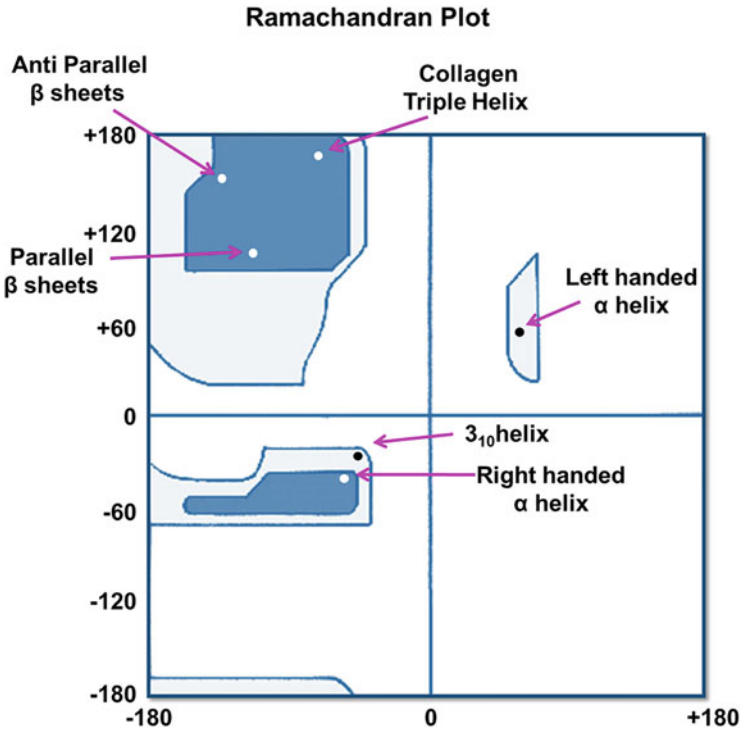

Fig. 1.5 Ramachandran plot showing sterically allowed values of Ψ and Φ. Conformational regions for differential secondary structures are also marked

Table 1.2 List of Ψ and Φ values for different secondary structure elements

Secondary structure	Φ (in degrees)	Ψ(degrees)
2.27 –ribbon	−78	59
Antiparallel β-pleated sheet	−139	135
Left-handed α helix	57	47
Left-handed polyglycine	−79	150
Polyproline type I helix	−75	160
Polyproline type II helix	−75	150
Parallel β-pleated sheet	−119	113
Right-handed α helix	−57	−47
Right-handed 310 helix	−49	−26
Right-handed π helix	−57	−70
Type I β-bend (residue2)	−60	−30
Type I β-bend (residue3)	−90	0
Type II β-bend (residue2)	−60	120
Type II β-bend (residue3	90	0

1.5.2 Secondary Structure of Proteins

Secondary structure represents the next degree of organization of structure in proteins. It basically refers to the local conformation attained by various parts of protein. It is a regular folding pattern followed by the polypeptide backbone. Existence of these structures was first predicted by Pauling and Corey in 1951. Different types of secondary structures were observed among the proteins; however, few of them including α-helices and β-sheets are highly stable and commonly present in proteins. Different secondary structures are described in the following paragraphs.

α-helices—polypeptides acquire helical structure when they twist by the equal quantities around each of its Cα atom. *α-h*elices are made as a resultant of formation of hydrogen bonds between amide groups and carbonyl groups found in close vicinity in the sequence. Hence, the polypeptides with helical structures are fully involved in hydrogen bonding except the amide group of the initial (first) residue and the carbonyl group of the amino acid present at the end of the polypeptide chain. This type of exhaustive hydrogen bonding cylinders the protein structures in which the middle axis is represented by the fully hydrogen bonded backbone and all the R groups or side chains are protruding outwards from the polypeptide for interacting with differential regions of other polypeptide chains or to some other different protein (Fig. 1.6a).

In the right-handed α-helices, the repeating unit is the single turn of helix with pitch (which is defined as the distance that the helix rises along its axis per turn) of 5.4 Å, torsion angles $\Psi = -47^\circ$ and $\Phi = -57^\circ$ and around 3.4 amino acid residues per helical turn (n). Hydrogen bonds are formed between amide (N-H) group of n^{th} amino acid residue and carboxyl (C=O) group of n-fourth amino acid residue. The common notation for helices is n_m, where n displays the number of residues per

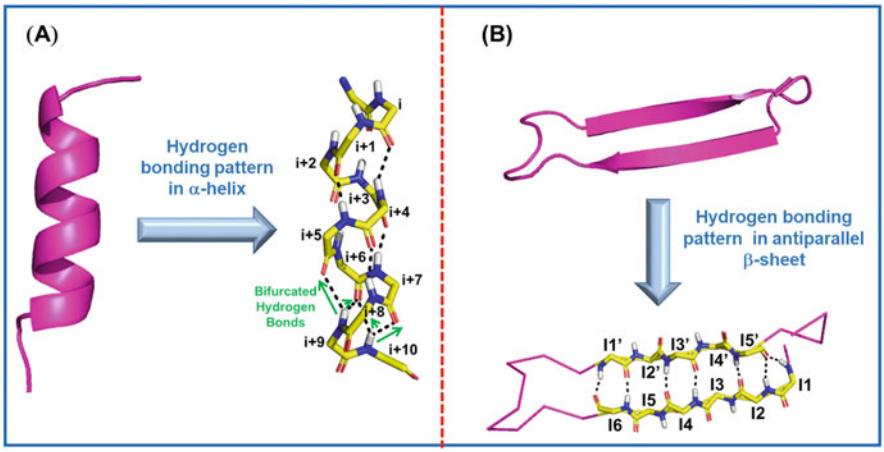

Fig. 1.6 Hydrogen bonding patterns followed by amino acids in (**a**) α-helix and (**b**) antiparallel β-strands

helical turn and m denotes the number of atoms that are indulged in hydrogen bonding, respectively. Differential forms of helices are observed in proteins such as α-helix, 4.4_{16} or π-helix, 2.2_7 ribbon, and 3_{10}-helix. These helices have been categorized on the basis of differential torsion angles, pitch, and number of residues per helical turn (n) (Table 1.2). α-helices are one of the most commonly observed helical structures in globular and fibrous proteins in contrast to other helices such as 4.4_{16} and 3_{10}-helix, which are found only in small segments of proteins or in turns. In general, ¼ of the amino acids acquire helical structure in all the polypeptides, the exact fraction varies from protein to protein. The preference for formation of helices in the protein structures is majorly due to their stability and the involvement of hydrogen bonding as a major stabilizing force. L-amino acids, that occur naturally, can form both the right-handed and the left-handed α-helices. However, left-handed α-helices are never observed in proteins. Formation of α-helices and their stability are highly dependent on sequence and identity of amino acids in a polypeptide. For instance, the presence of a block of acidic amino acids like Glu in a segment of protein will not be able to form α-helix at pH -7.0, as the negative charge of carboxyl groups of Glu residues will make them repel from each other. Similarly, presence of block of Lys/Arg residues will inhibit the formation of α-helix due to repulsion of positively charged NH groups of Lys/Arg. Proline because of its ring structure introduces a kink in the α-helix. Hence, Pro is rarely observed in helices. Gly is observed infrequently in α-helices owing to its conformational flexibility.

β-*sheets*—these are the second type of repetitive structures that are acquired by the proteins. β-sheet structure was confirmed using X-ray analysis. It is a more extended conformation of polypeptide chains with a zigzag structure. This structure resembles pleats when polypeptide chains are arranged side by side. β-sheets are made due to formation of hydrogen bonds between the neighboring segments of polypeptide chains. Segments forming β-sheets can be the segments that belong to the same polypeptide chain that can be present nearby or far from each other. The other possibility is that they can be the segments of different polypeptide chain. The amino acid side chains point outwards from the zigzag structure in the opposite directions (Fig. 1.6b).

β-sheets are of two types, namely: parallel or antiparallel depending on the orientation of adjacent polypeptide chains. Parallel β-sheets are formed when both the adjacent polypeptides have same amino to carboxyl orientations. Antiparallel β-sheets are formed as a result when adjacent polypeptides run in opposite direction. Structures for both parallel and antiparallel are quite similar; however, their repeat distances and hydrogen bonding patterns are different. Repeat distance for parallel and antiparallel β-sheets is 6.5 Å and 7.0 Å, respectively. C=O and NH groups from the two polypeptides that are facing each other form hydrogen bonds (Fig. 1.6b). Other NH and C=O groups that are at the ends can be involved in hydrogen bonding with water molecules that are present on the surface or these groups can also be involved in interaction with side chains in α-helices to pack against them. In case of β-barrel formation, NH and C=O groups from the adjacent polypeptides curve around one another. β-sheets that are antiparallel are more common among the proteins in comparison with parallel β-sheets, as antiparallel β-sheets are more stable

than parallel β-sheets. This is due to the presence of distorted patterns of hydrogen bonding in parallel β-sheets. β-sheet strands are either connected by β-turns or hairpin turns or by complex helical turns. Knotted topologies have never been observed in β-sheets. β-sheets exhibit a structure that is fully extended and have phi and psi values ($\Phi = -119^\circ$ and $\Psi = 113^\circ$ for parallel β-sheets and $\Phi = -139^\circ$ and $\Psi = 135^\circ$ for antiparallel β-sheets) that belong to sterically permitted regions of Ramachandran plot. β-sheets are present on the surface of the protein owing to the presence of adjacent hydrophobic and hydrophilic amino acids in the adjoining β-strands of the β-sheets.

In addition to these regular secondary structures, globular proteins also exhibit some irregular and non-repetitive structures. These structures are known as loops/turns/coil conformations. Loops and turns allow the polypeptide chain to make a turn in reverse direction and also aid in connecting other regular secondary structures. β-turns are most commonly found in linking the segments of antiparallel β-sheets. β-turn involves four residues in which carbonyl group of the first amino acid is involved in forming hydrogen bond with the amino group of the fourth amino acid. The primary two residues are not involved in formation of any inter-residue hydrogen bonding. Proline and glycine are most commonly found in β-turns because of the small size of glycine and cyclic structure of proline. Turns with glycine and proline are tight turns. β-turns are usually observed on the surface of the protein in which the peptide groups of central amino acid residues can involve in hydrogen bonding with water. Three-residue turn known as γ-turn in which hydrogen bond forms between the first and third amino acid residue is less common among the proteins.

Amino acids do have the preferences for the secondary structures to adapt. Some of them, for instance, proline and glycine prefer to be present in turns than in regular secondary structures including β-sheets or α-helices. This is due to the several constraints in secondary structures. In 1974, Chou and Fasman performed an analysis on several protein crystal structures and determined the preferential scores of all the 20 standard amino acids for differential secondary structures including α-helices, β-sheets, and reverse turns (Fasman 1990). Based on which they categorized residues as strong helix breaker or former and strong β-sheet breaker or former. Table 1.1 enlists the conformational parameter scores for each of the standard amino acid.

1.5.3 Tertiary Structure of Proteins

Tertiary structure is defined as the spatial arrangement or three-dimensional arrangements of all the atoms in the protein with an aim to form a compactly folded conformation. In secondary structures, the spatial arrangement of adjacent amino acids has been defined. In contrast, spatial arrangement of long range amino acids will be defined in tertiary structure. Amino acids that are present far from one another in polypeptide sequence or the residues present in differential secondary structure may interact in a folded protein structure. Residues may interact via weaker interactions or by covalent bonds such as disulfide bonds. Proteins acquire natively

folded structures rapidly; however, in some cases, proteins remain in their partially folded or natively unfolded state.

In globular proteins, the side chains of non-polar residues like Ile, Leu, Met, Phe, and Val occur in the protein's hydrophobic interior. These non-polar side chains can interact to each other and form a cluster in the water repelling interior of the protein and thereby resulting in a hydrophobic core in the interior of a protein. In contrast, the side chains of charged polar residues like Asp, Arg, Glu, His, and Lys reside on the surface of globular proteins from where these residues are accessible to the polar solvents or aqueous environment. In case of large polypeptides, the tertiary structural arrangement is in the form of domains. Different segments of polypeptide form different globular domains which can either directly interact with one another or via a connecting polypeptide. Such domains can have different functions and conformations which can alter by the interaction with each other by means of cooperativity and allostery. In most of the cases, the interactions between the domains are highly flexible, as the interface forms the binding site for both biomolecules and small molecules (Voet and Voet 2011).

1.5.4 Quaternary Structure of Proteins

Numerous proteins comprise multiple polypeptide chains/subunits and their association performs variety of functions. Multi-subunit proteins are involved in regulatory functions. Small molecule binding may affect the interaction between the subunits of the protein, thereby greatly affects their activity. In many cases, different subunits perform differential but related functions. Large proteins serve as the site of various multi-step complex reactions (Petsko and Ringe 2004).

Multi-subunit proteins are known as "multimers." They are made of either two subunits or several hundreds of subunits. If a protein is composed of only few subunits, then it is known as oligomer. Protein is said to be asymmetric, if it is composed of non-identical subunits. In multimers, mostly the subunits are identical or they constitute repeating groups of non-identical subunit arranged in symmetric fashion. Hemoglobin is first oligomeric protein whose 3D structure was defined by employing X-ray crystallography in 1959. It is composed of four polypeptide chains along with four heme prosthetic groups with iron atoms in ferrous state. These polypeptide chains are arranged symmetrically.

Oligomers can range from dimers, trimers, tetramers, etc. Identical subunits interact in symmetric manner due to the presence of complementary binding surfaces. In contrast, non-identical subunits interacting asymmetrically will have different binding surfaces. For instance, a subunit with X binding surface will interact with the subunit having X' binding surface. The asymmetric subunit is also referred to as the protomer. Subunits that have two complementing interacting surfaces result in formation of more complex from of protein. For example, monomers interact to form dimers, dimers can interact to form tetramer (Fig. 1.7a–c). Further, interaction of subunits can also lead to hexamer and octamer (Fig. 1.7d, e) and other higher order structures like decamers and dodecamers. Some of the