Principles of Protein Structure and Function

1

Juliana Helena Costa Smetana and Gauri Misra

Abstract

The present chapter highlights the structural properties of amino acids and organization of proteins into primary, secondary, tertiary, and quaternary structures. The concept of dihedral angles and Ramachandran plot is given to provide an understanding of the thermodynamically feasible protein conformations. The biophysical concepts of structural hierarchy are explained with the help of suitable examples. A deep understanding of various folds and motifs is provided to clear the theoretical aspects of protein structure and its functional implications including the recently discovered novel fold. Besides thermodynamic parameters of structural organization, various covalent and noncovalent interactions that play an important role in protein structure are explained. The important globin-containing proteins, namely, hemoglobin, myoglobin, cytochromes, and lysozyme, have been discussed with respect to their structural features responsible for their functioning. Further, the structural aspects of integral and peripheral membrane proteins are elaborated.

Keywords

Protein structure • Amino acids • Alpha helix • Beta sheet • Protein folding

J.H.C. Smetana

Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), 13083-970 Campinas, Sao Paulo, Brazil
e-mail: juliana.smetana@lnbio.cnpem.br

G. Misra (✉)
Amity University, Noida, Uttar Pradesh, India
e-mail: kamgauri@gmail.com

1.1   Structure, Classification, and Properties of Amino Acids

Proteins are the result of polymerization of **L-α-amino acids** by **peptide bonding** (Fig. 1.1). This, however, is also true for any polypeptide or small peptide. As an additional information to define what a protein is, one needs to consider that proteins, different from other polymers made of L-α-amino acids, are products of biological evolution. They have a biological function which is defined by their unique three-dimensional structures (Kyte 1991; Branden and Tooze 1999; Petsko and Ringe 2004).

Intrinsically unfolded proteins do not display a rigid structure, yet they still have a biological function which depends on their structural properties. These proteins usually adopt different conformations depending on which interaction partner is associated with them, and when free in solution they explore an ensemble of different conformations.

Amino acids are **zwitterionic** molecules, meaning that they have both acidic (carboxylate) and basic (amino) groups and have neutral net charge. However, in addition to the amino and acid groups, amino acids have a **radical group** (or side chain) covalently linked to a central carbon, the **α-carbon**. If this side chain is not neutral then the resulting molecule will be charged. There are only 20 **proteinogenic** amino acids (the ones found in natural proteins), and each one of them is specifically designated by one or more codons in the **genetic code**.
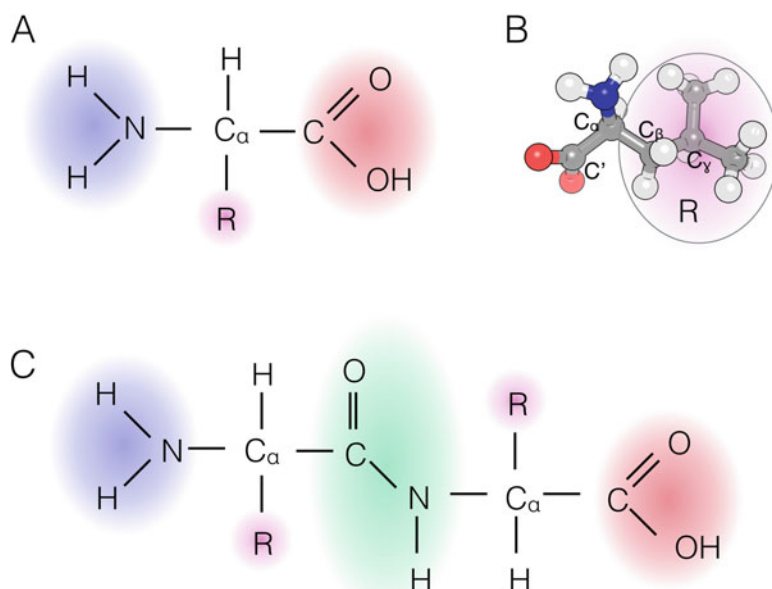

Fig. 1.1 Amino acids and peptide bond (**a**) Illustration of the general structure of alpha-amino acid R: radical group. (**b**) Example of amino acid structure (L-Leucine) in ball and stick representation. (**c**) Illustration of the peptide bond (highlighted in *green*)

The α-carbon is **chiral** and all life on earth has evolved to use only **L-amino acids** in proteins. There are only a few exceptions to this rule, such as peptidoglycans made of D-amino acids which are found in bacterial cell walls. Experiments have shown that artificial, synthetic proteins made of D-amino acids are the exact mirror images of their natural counterparts made of L-amino acids and their substrate specificity reflects this inverted chirality. The reason for the almost exclusive presence of L-amino acids in naturally occurring proteins is not well understood and might have been favored by chance.
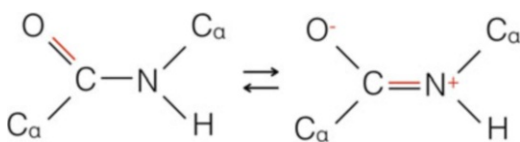
Incorporation of an amino acid in protein results in the loss of both the amino and acidic groups due to peptide bonding; hence it is no longer called an amino acid but "amino acid residue" instead – or, for simplicity, "residue." The peptide bond is catalyzed by the ribosome in living cells and is accompanied by loss of a water molecule. The nature of the peptide bond is of a **partial double bond**, or resonance, implying that the bond is not free to rotate and the four atoms involved in this bond are **coplanar** (Fig. 1.2).

Therefore, the conformational flexibility of the backbone derives from the remaining bonds which are able to rotate around their axes: the N–Cα bond and the Cα–C bond which define, respectively, the **torsion angles phi (φ)** and **psi (ψ)** (Fig. 1.3). These angles are dihedral angles defined three-dimensionally as the angles between planes. There is a limited rotation range for each of these angles in proteins due to steric clashes. A two-dimensional plot of phi and psi angles, the **Ramachandran plot** (Fig. 1.4), (Ramachandran et al. 1963) categorizes the information about possible combinations for these angles in protein structures, highlighting regions that are allowed. Glycine behaves differently because its side chain consists of a hydrogen atom, and its allowed range of phi and psi angles is considerably broader than those of other amino acids. Repetitive patterns of these torsion angles result in secondary structure elements.

Because proteins are not circular, they have extremities which are defined by the presence of unbound amine or carboxyl groups. The amino group defines the amino-terminal or **N-terminal** which is the first to appear during ribosomal translation and usually contains the **initiator methionine**. The carboxyl group defines the carboxy terminus or **C-terminus** of the protein, the last segment to be translated. Therefore, proteins display N- to C-terminal **polarity**.

The side chain atoms are labeled with Greek letters. The atom closer to the α-carbon is β, the next one is labeled γ, and so on. The chemical bonds connecting each of these atoms define dihedral angles in the same way as the phi and psi angles of the main chain. Glycine and alanine are devoid of any degrees of freedom in their side chains; therefore only one **rotamer** exists for each of them. All the remaining amino acids can be found in different conformations or rotamers. The aromatic side

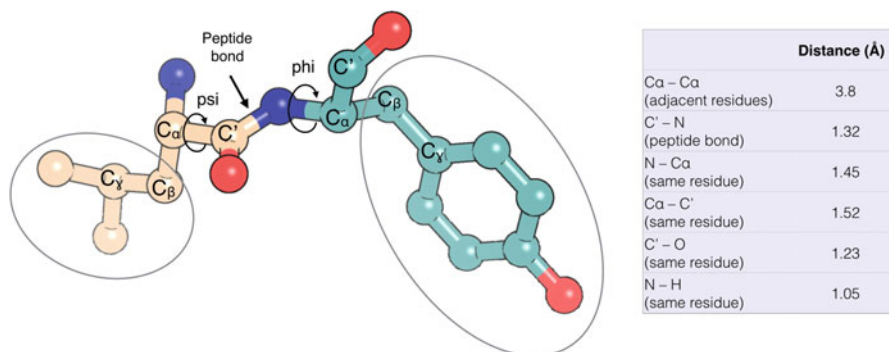**Fig. 1.2** Resonance in the peptide bond

	Distance (Å)
Cα – Cα (adjacent residues)	3.8
C' – N (peptide bond)	1.32
N – Cα (same residue)	1.45
Cα – C' (same residue)	1.52
C' – O (same residue)	1.23
N – H (same residue)	1.05

Fig. 1.3 Peptide bond and torsion angles are highlighted in the structure of a Leu–Phe bond. The table summarizes bond lengths and Cα–Cα distance (Adapted from Withford D)

white regions-sterically disallowed for all amino acids except glycine.

blue regions-allowed regions namely the α-helical and β-sheet conformations.
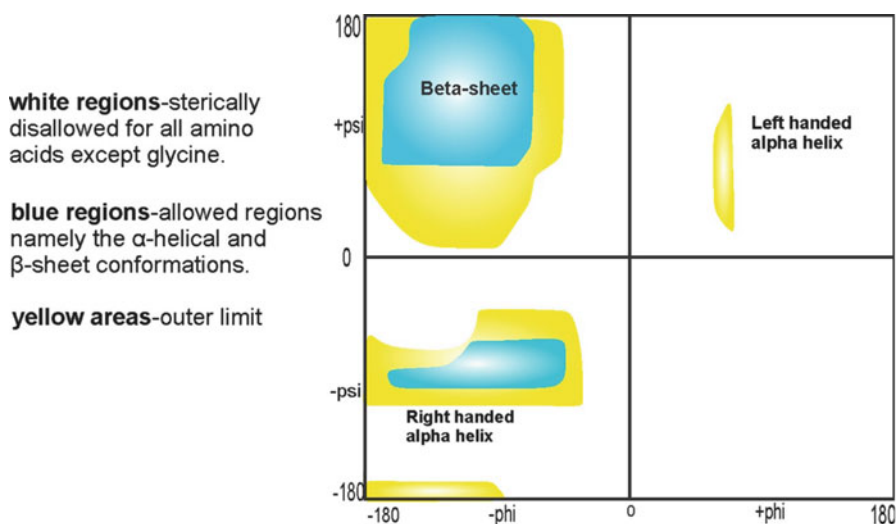
yellow areas-outer limit

Fig. 1.4 Ramachandran plot

chains are restricted by the rigidity of double bonds, the aromatic rings are planar, and the rotational freedom is reduced. For each amino acid, some rotamers are common while others are highly unusual, and this depends on the energetics of each configuration. In general, **staggered conformations** are favored for tetragonal carbons, and *trans* are usually favored over *cis* conformations. The only way to accurately determine the conformation of one given amino acid within a protein is experimentally, usually by X-ray crystallography. The most common softwares for X-ray structure refinement are equipped with rotamer libraries to facilitate the fitting of amino acid side chains to electron density.

The proteinogenic amino acids can be **classified** in different ways according to the nature of their side chains (Fig. 1.5). In a broader way, they can be grouped as

	One letter code	Three letter code	Chemical classification	Functional group in side chain	Elements in side chain (except hydrogen)
Glycine	G	Gly	-	-	-
Alanine	A	Ala	apolar, aliphatic	alkyl (methyl)	C
Isoleucine	I	Ile	apolar, aliphatic	alkyl	C
Leucine	L	Leu	apolar, aliphatic	alkyl	C
Methionine	M	Met	apolar, aliphatic	thioether	C, S
Proline	P	Pro	apolar, aliphatic	imino acid	C
Valine	V	Val	apolar, aliphatic	alkyl	C
Phenylalanine	F	Phe	apolar, aromatic	phenyl	C
Tryptophan	W	Trp	aromatic, amphipatic	indole	C, N
Tyrosine	Y	Tyr	aromatic, amphipatic	phenol	C, O
Aspartate (Aspartic acid)	D	Asp	polar, charged (negative)	carboxylic acid	C, O
Glutamate (Glutamic acid)	E	Glu	polar, charged (negative)	carboxylic acid	C, O
Histidine	H	His	polar, charged (positive)	imidazole	C, N
Lysine	K	Lys	polar, charged (positive)	amine	C, N
Arginine	R	Arg	polar, charged (positive)	guanidine	C, N
Cysteine	C	Cys	polar, uncharged	sulpydryl (thiol)	C, S
Asparagine	N	Asn	polar, uncharged	amide (carboxamide)	C, N, O
Glutamine	Q	Gln	polar, uncharged	amide (carboxamide)	C, N, O
Serine	S	Ser	polar, uncharged	hydroxyl	C, O
Threonine	T	Thr	polar, uncharged	hydroxyl	C, O

Fig. 1.5 Classification of amino acids

hydrophobic, polar, and charged. Glycine is the simplest amino acid which is usually grouped with the hydrophobic amino acids, but due to its unique properties it is often considered to form a separate group. The **polar** amino acids can be further classified into **charged** and **uncharged**. The charged amino acids are the ones whose side chains can be fully ionized in neutral pH. They can be either positively (histidine, lysine, and arginine) or negatively charged (aspartate and glutamate). The side chains of uncharged, polar amino acids display different chemical groups: hydroxyl groups are present in serine and threonine (as well as tyrosine), sulfhydryl (thiol) is found in cysteine, and carboxamide is in asparagine and glutamine. Each chemical group endows them with unique properties: for example, cysteine is the only side chain that can form covalent bonds by means of disulfide bridges, and hydroxyl groups are frequent sites of posttranslational modifications such as phosphorylation. **Hydrophobic**, apolar amino acids can be further classified into **aromatic** and **aliphatic** side chains. Amino acids with aromatic side chains include phenylalanine, which has a benzene group as side chain, tyrosine, and tryptophan, the bulkier amino acid. Together, the aliphatic amino acids leucine, isoleucine, and valine constitute the **branched-chain amino acids** (**BCAA**), defined as aliphatic amino acids with a branch point in their side chains.

1.2   Structural Levels of Proteins

1.2.1   Primary Structure

Linear amino acid sequence in a protein joined together by the peptide bonds defines its primary structure. In essence, the sequence defines the structure and function of the protein, as we will see in the coming sections. It can be represented by the one-letter or three-letter codes for amino acids. For example, both M-A-E-D (one-letter code) and Met-Ala-Glu-Asp (three-letter code) describe the same stretch of residues in a protein, i.e., a stretch of its primary sequence. The primary structure is represented from N-terminus to C-terminus, compatible with the 5′ to 3′ orientation of the DNA and RNA sequences that codify proteins. It can be inferred from the DNA or RNA sequence using the genetic code table. The usual amino acid residue found in the N-terminal extremity of the primary structure in natural proteins (i.e., those expressed in a cell, as opposed to those produced synthetically) is **methionine**, which is specified by the **ATG start codon**. In bacteria, **formyl-methionine** is used as the initiator residue instead of methionine.

1.2.2   Secondary Structure

The repetitive patterns of phi and psi angles of the polypeptide backbone define what is known as **secondary structure**. The classical regular secondary structure elements frequently found in proteins are the **α-helix**, **β-strand**, and the **reverse turns,** in addition to **loop regions** characterized by the lack of a repetitive pattern.

In the Ramachandran plot, each type of secondary structure corresponds to one of the allowed regions in the graphic.

The **α-helix** was predicted by Linus Pauling based on the known chemical structure of polypeptides, including the fact that the peptide bond is planar and the C–N distance within this bond is 1.32Å (Pauling et al. 1951). Using a piece of paper, he drew an extended polypeptide chain and folded it in a way that would maximize the noncovalent interactions. The model, a "hydrogen-bonded helical configuration of the polypeptide chain," was published in PNAS in 1951 along with a second helical model. The structure of myoglobin, the first three-dimensional protein structure to be solved, was composed mostly by α-helices showing the exact geometry predicted by Linus Pauling.

The geometry of the α-helix is defined by the repetitive phi and psi angles found in this structural element, which are, respectively, $-57°$ and $-47°$. These angles are found in the bottom left quadrant of the Ramachandran plot. This combination of angles results in a favorable geometry where all the main-chain N and O atoms are connected by hydrogen bonding. The α-helix has 3.6 residues per turn, and the hydrogen bonds are formed between $C'=O$ of one residue and N–H of another residue, skipping three positions: $C'=O$ of residue 1 pairs with N–H of residue 5, residue 2 with 6, and so on (Fig. 1.6). The α-helices found in proteins are **right-handed** because they are made of L-amino acids.

There are other possible helical elements of secondary structure in proteins which are, however, very rare and energetically unfavorable. The 3_{10}**-helix** is formed when a residue N makes a hydrogen bond with residue N + 3, and the **π-helix** occurs when residue N connects in such a way with residue N + 5. The name of the 3_{10}-helix derives from the fact that it has exactly three residues per turn and ten atoms between the hydrogen bond donor and acceptor. Using the same logic, an α-helix would be a 3.6_{13}-helix, although this nomenclature is not usual. The top view of a 3_{10}-helix is triangular (Fig. 1.7)**.**

The definition of a **π-helix** implies that if a single amino acid insertion occurs in the middle of an α-helix, it could be accommodated by locally adopting the dihedral angles of a π-helix. In fact, these insertions are the proposed evolutionary origin of π-helices, and these are usually found as short elements within α-helices, frequently close to functional sites, rather than as isolated structural elements (Cooley and Karplus 2010).

β-strands, also predicted by Linus Pauling, (Pauling and Corey 1951) are extended structures stabilized by hydrogen bonds between the amide N–H and C=O groups from the main chain (Fig. 1.8). They are usually represented as arrows in schematic representations of protein structures. These groups protrude laterally, making connections with the same kind of groups from an adjacent strand. Two or more β-strands connected in this way form a **β-sheet**. These connections can be either parallel or antiparallel. The amino acid side chains protrude up and down from the β-sheets in an alternating pattern. Typical geometric parameters in β-sheets are the distance of 3.3 Å between consecutive residues and the phi and psi angles which are $-130°$ and $125°$, respectively. They are also characterized by a pronounced **right-handed twist** originating in steric effects attributed to the
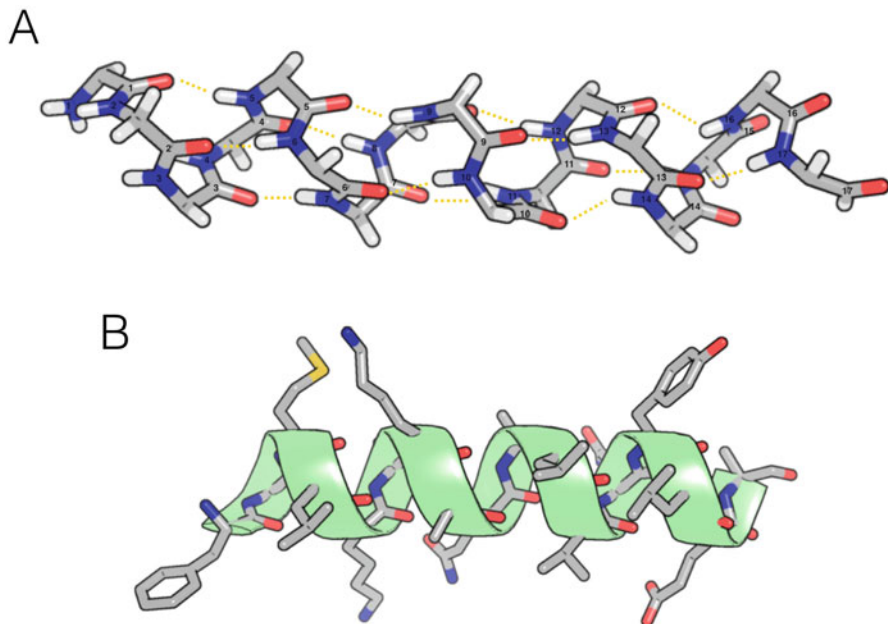
Fig. 1.6 Secondary structure: alpha helix. (**a**) Representation of an alpha helix shown here in stick representation (upper panel). The amino acid side chains were omitted. Hydrogen bonds are represented as dashed lines, and the amino acid residues are numbered. (**b**) Illustration of an alpha helix combining sticks and cartoon representation. Carbon, *gray*; Nitrogen, *blue*; Oxygen, *red*; Hydrogen, *white*; Sulfur, *yellow*

presence of L-amino acids; this twist allows them to form saddlelike structures and closed, circular structures.

Both α-helices and β-sheets can be **amphipathic** or **hydrophobic.** Amphipathic helices and sheets are found on the surfaces of proteins, allowing one of the sides to engage in hydrophobic interactions with other structural elements inside the protein's core, while another side is polar and exposed to solvent. β-sheets formed exclusively by hydrophobic amino acids are often found in the interior of proteins, particularly when they form parallel sheets. Hydrophobic antiparallel β-sheets as well as hydrophobic α-helices are typically found as transmembrane structures, in the form of β-barrels and single-span helices, respectively.

Loop regions are stretches of irregular structure (the angles phi and psi do not follow regular patterns) connecting helices, sheets, and other regular elements, and they are usually localized at the surface of proteins. These regions are frequent sites of insertions and deletions as their irregular structures are more amenable to accommodate changes than regular secondary structure elements. Loops are also functionally important elements in active sites of enzymes, antigen-binding sites of antibodies, and in protein–protein and protein–small molecule interfaces. Charged and polar residues are common in loops, and both their side chains and main chains ($C'=O$ and NH groups) are involved in hydrogen bonding with solvent or ligands.
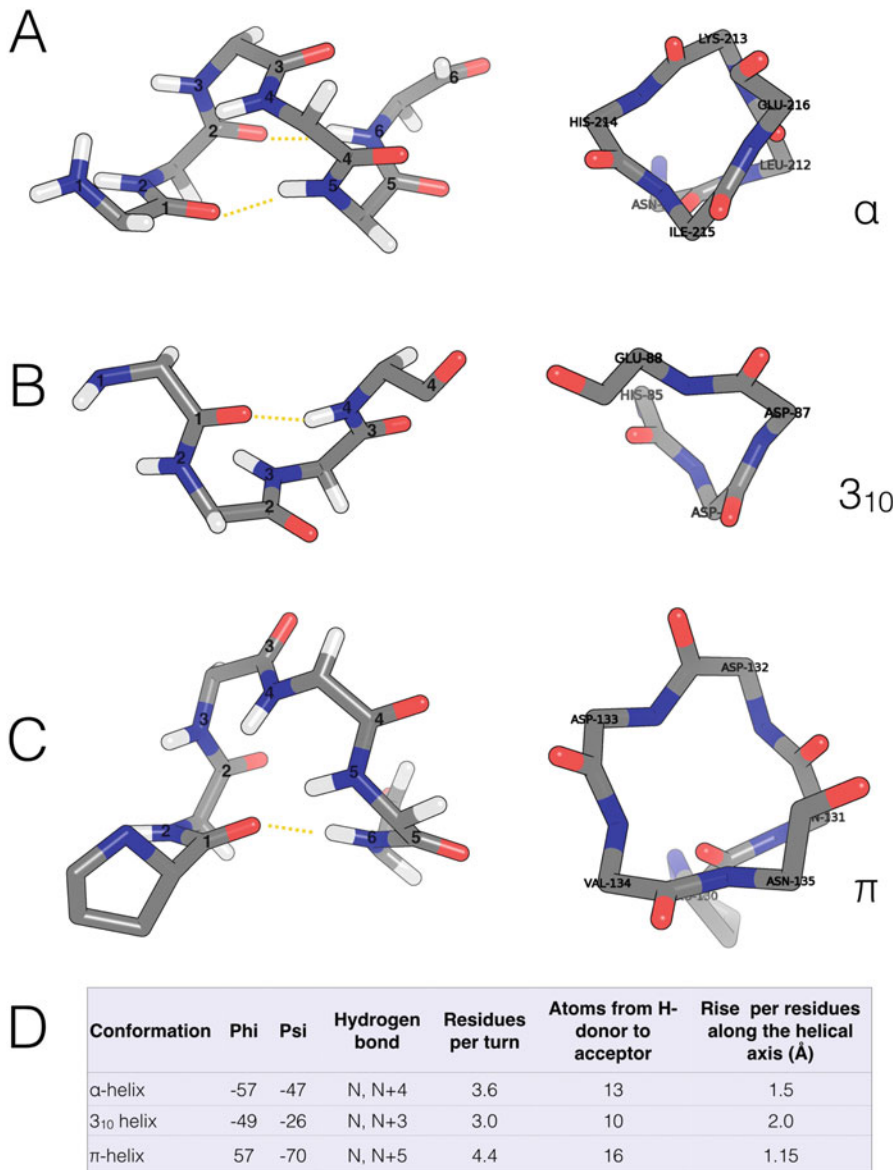
Fig. 1.7 Secondary structure: helical elements. Stretches of amino acid residues adopting α-helical (**a**), 3_{10} (**b**), or π-helical (**c**) configurations were extracted from the structure of fumarase C (PDB: 1FUO). The left panels show side views and the numbering of residues is relative. Side chains are omitted for clarity, except for proline. The *right panels* show upper views where the amino acid residues are labeled according to the original PDB file. Side chains (except proline) and hydrogen atoms are omitted. Carbon, *gray*; Nitrogen, *blue*; Oxygen, *red*; Hydrogen, *white*. Figure adapted from Weaver, 2000. (**d**) Summary of the properties of the α-helix, 3_{10}-helix, and π-helix (Adapted from Petsko and Ringe)

Table (D):

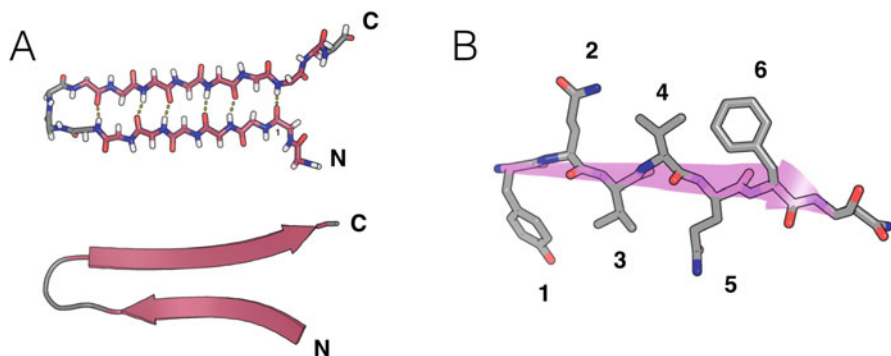| Conformation	Phi	Psi	Hydrogen bond	Residues per turn	Atoms from H-donor to acceptor	Rise per residues along the helical axis (Å)
α-helix	-57	-47	N, N+4	3.6	13	1.5
3_{10} helix	-49	-26	N, N+3	3.0	10	2.0
π-helix	57	-70	N, N+5	4.4	16	1.15

Fig. 1.8 Secondary structure: antiparallel beta-sheet. (**a**) A stretch of amino acids in β-sheet connected by a β-hairpin structure/type I turn, shown here in stick representation (*upper panel*). The amino acid side chains were omitted (they would be protruding above and below the plane of the figure). The same peptide is shown below in cartoon representation. (**b**) In beta-sheets, the amino acid side chains protrude up and down in an alternating fashion

The length of loops in proteins is variable, ranging from two to around 20 residues in the majority of proteins, although there is no upper limit for their size. Long loops account for a protein's mobility, as their irregular structure allows them to adopt variable conformations, sometimes in a regulated fashion responding to stimuli. Short loops of only two to four residues are more rigid. **Hairpin loops** are short loops connecting adjacent antiparallel β-strands. **Turns** are also considered to be a special type of short loop with defined characteristics. The terms hairpin loop, beta turn, and reverse turn are often used interchangeably.

The constraints for polypeptide chain reversal by turns were established by Venkatachalam in 1968 (Venkatachalam et al. 1968). **Reverse turns** are stretches of four amino acid residues in which the peptide chain undergoes a nearly 180° reversal, forming a hydrogen bond between the C=O group of residue N and the NH group of residue N+3. Considering the possible conformations of residues in turns, Venkatachalam identified three types of reverse turns based on their allowed dihedral angles. In **type I turns**, φ_2, ψ_2, φ_3, and ψ_3 are −60°, −30°, −90°, and 0°, respectively. In **type II turns**, these angles are −60°, 120°, 80°, and 0° (Fig. 1.9). **Type III turns** display the angles −60°, −30°, −60°, and −30°, which are actually identical to the 3_{10}-helix. If all the signs of these dihedral angles are reversed, the turns are classified as types I′ II′, and III′, respectively, which are also allowed conformations. There is a tolerance of about 15° for each of these angles (Crawford et al. 1963).

Some amino acids are more likely to be found in α-helices, while others in β-sheets, turns, or loops. Long side chains, such as those of Leu, Met, Gln, are frequently found in α-helices, while branched-chain amino acids (Val, Ile) and bulky side chains like Trp, Tyr, and Phe are usually found in β-sheets. Proline and glycine have a strong preference for turns because of their unique conformations. Proline is favored at the second position in both type I and type II turns, while it is
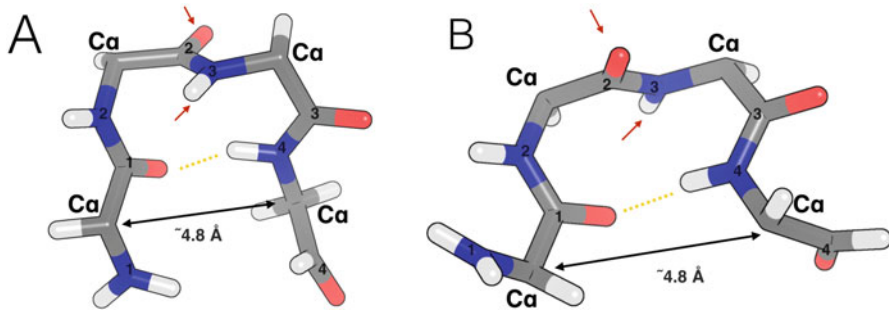
Fig. 1.9 Secondary structure: reverse turns. Examples of residues adopting the indicated conformations were taken from the lysozyme structure (PDB: 193L) and carboxypeptidase A structure (PDB: 5CPA) according to their classification by Crawford et al. (1972). (**a**) Type I turn (residues 41–44 of carboxypeptidase A; sequence: Ser-Tyr-Glu-Gly). (**b**) Type II turn (residues 17–20 of lysozyme, sequence: Leu-Asp-Asn-Tyr). Residues are numbered from one to four according to their positions in the turn. The double-headed arrows indicate the distance between Cα carbons of the first and last residues. The red arrows indicate the major differences between type I and type II turns, located in the relative positions of C=O (second residue) and NH groups (third residue). In type I turns, the oxygen atom is projecting to the back, and the hydrogen is projecting to the front in the orientation shown in the figure. In type II turns, these positions are inverted. Side chains are omitted for clarity. Carbon, *gray*; Nitrogen, *blue*; Oxygen, *red*; Hydrogen, *white*

not allowed at the third position in type I turns. However, as a rule, no amino acid is forbidden in any type of secondary structure – even proline can participate in α-helices where it causes a distortion of the hydrogen bond pattern that sometimes can be accommodated with a small kink.

Secondary structure can be inferred with high accuracy from primary structure using bioinformatics tools. Experimentally, the global secondary structure content can be determined spectroscopically using circular dichroism or infrared spectroscopy.

1.2.3 Tertiary Structure

A **fold** is characterized both by the relative position of secondary structure elements and by the arrangement of their connections. **Topology** refers to the interconnections among secondary structure elements. The usual topological representation of proteins depicts α-helices as cylinders and β-sheets as arrows in a flat projection of secondary structure elements and their connections. Alternatively, flat representation of protein topology can be depicted as if the observer is looking at the secondary structure elements from top to bottom. In this case, α-helices are represented as circles and β-strands as triangles which point up or down to indicate the direction of the strand.

Examples of different folds showing similar structural elements but different topologies are the roadblock and longin folds (Levine et al. 2013) (Fig. 1.10). These
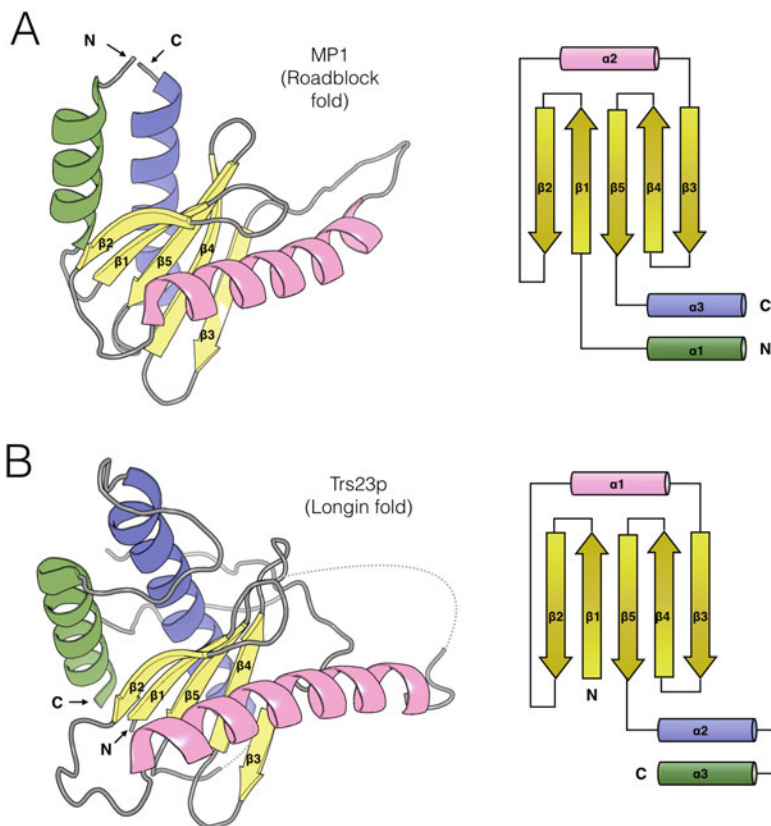
Fig. 1.10 Different topology results in different fold. (**a**) MP1 adaptor protein, a roadblock fold protein (PDB 1VET, chain B). (**b**) Yeast protein Trs23p, a longin fold (PDB 3CUE)

folds are present in adaptor proteins involved in vesicle trafficking, motility, and GTPase regulation. The relative three-dimensional positions of the secondary structure elements in roadblock and longin folds are very similar: both are α/β domains of three α-helices organized around a core of five antiparallel β-strands. However, one of these α-helices is positioned N-terminally in the roadblock fold, while it is located C-terminally in the longin fold. This is likely the result of an ancient event of circular permutation from a common ancestor fold.

The number of protein folds in nature is large but finite. As more and more protein structures become available, it becomes increasingly rare to find novel folds. Several efforts have been made to categorize and explore the wealth of structural information currently available. The major resources dedicated to categorize protein folds and structures are the online **databases CATH** and **SCOP**. CATH is an acronym for Class, Architecture, Topology/fold and Homologous superfamily. According to the latest census available in the CATH database, there are 2737 superfamilies. In SCOP (Murzin et al. 1995), domains are defined

as structurally and functionally independent evolutionary units that can either fold in a single-domain protein on their own or recombine with others to form part of a multidomain protein. Figure 1.11 illustrates an example of a recently described protein fold. The human TIPRL structure reported by Scorsato et al. (2016) displays two layers of antiparallel β-sheet surrounded by α-helices and the rare 3_{10}-helices as well as a considerable amount of random loops in a completely novel architecture. This structure is likely the representative of the conserved Tip41/TIPRL family fold, belonging to regulatory proteins that modulate the activity of type 2A serine/ threonine phosphatases.

Protein domains are classified in families according to the types of secondary structure present in them. α-domains are made exclusively of α-helices and β-domains exclusively of β-sheets, and there are two types of mixed structures: α/β domains and α + β domains. All of these also have elements of random structure (loops) connecting the helices and sheets. Additionally, some protein domains are stabilized by disulfide bridges and coordination of metal ions rather than secondary structure elements; these are known as crosslinked domains.

Multidomain proteins are, by definition, constituted by more than one domain in a single, continuous polypeptide chain. They evolve by gene duplication, divergence, and fusion. Each domain in a protein can fold and function independently, while still suffering influences from the other domains. The relative independence of protein domains is illustrated by the fact that isolated domains from multidomain proteins can be expressed in heterologous systems, characterized biochemically and functionally, and even crystallized. In fact, this is often the only way to get high resolution information on some large multidomain proteins which are frequently too flexible to be crystallized.

α-domains are fundamentally organized upon the relatively few possible ways to pack α-helices against each other. In soluble globular proteins, these helices are usually amphipathic, meaning that they have a polar side facing the solvent and a hydrophobic side facing another α-helix. Most of the helix–helix packing found in globular α-domain proteins can be explained by the "ridges in grooves" model (Fig. 1.12). The side chains in the surface of α-helices form ridges and grooves which interdigitate, forming connections that stabilize the interaction. There are two ways to trace imaginary parallel lines connecting the side chains in α-helices to define the ridges. If we pick the side chains which are four residues apart, these would form an angle of $25°$ relative to the direction of the helix. Instead, if we choose to pick side chains separated by three residues, the imaginary lines form an angle of $-45°$. If we try to pack two helices according to the $25°$ angles so that the ridges defined by these lines will interdigitate with the grooves in the other helix, we would have to turn one of the helices so that the angle between them is $25 + 25 = 50°$. Another possible scheme is combining the $25°$ lines in one helix with the $-45°$ lines in another helix, which would result in an angle of $25° - 45° = -20°$. In fact, $50°$ and $-20°$ are two of the most frequent angles found in helix–helix packing. The $-20°$ angle is found in the **four-helix bundle**, a common structural motif characterized by four helices packed in a bundle which can be either parallel or antiparallel (Fig. 1.13). Some of the simplest α-domain proteins, such as the human
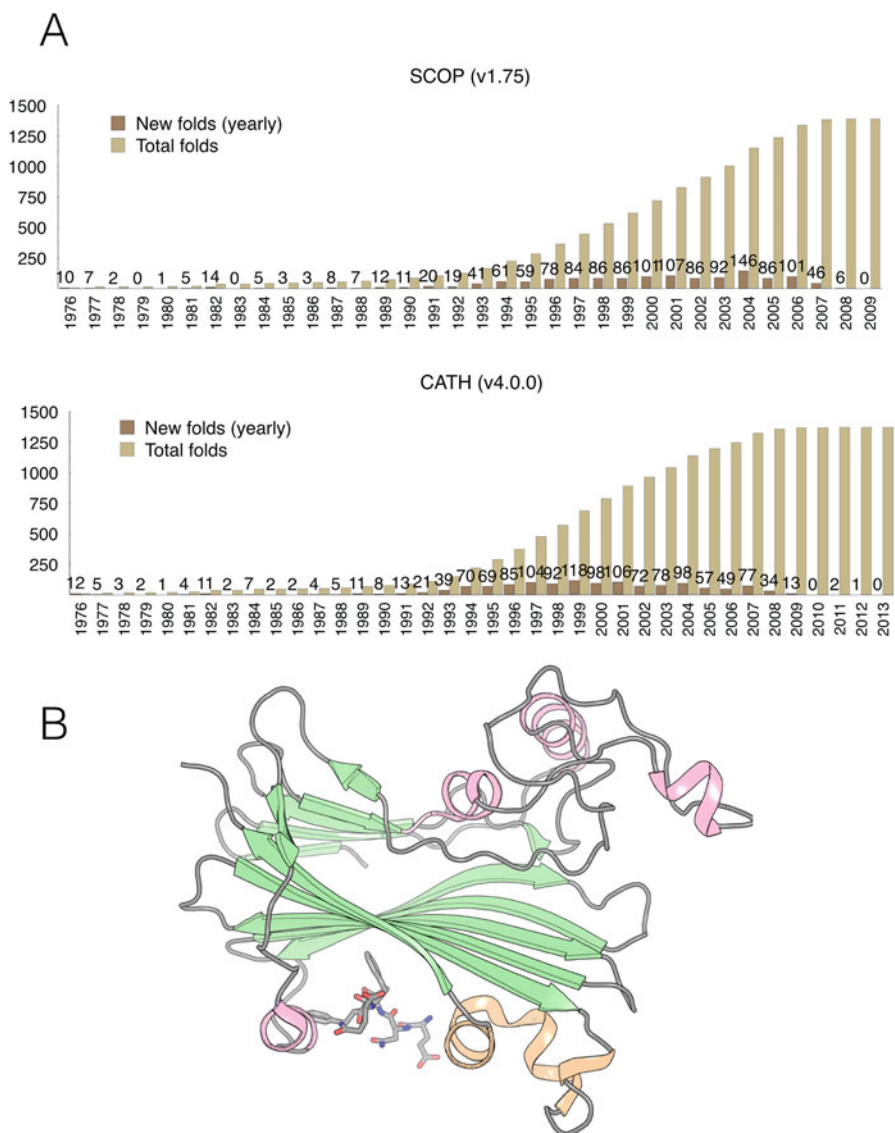
Fig. 1.11 The scarcity of novel folds. (**a**) The number of novel folds deposited on the PDB each year according to CATH (lower panel) and SCOP (*upper panel*). *Red lines* indicate the number of structures, and *blue lines* indicate the number of novel folds per year. (**b**) A recently described fold of a human protein: the TIPRL/hTip41 protein, a regulator of type 2A phosphatases (PDB 5D9G, released in 2016). The fold has a peptide-binding cleft which accommodates the extreme C-terminus of PP2Ac. The peptide is displayed as sticks
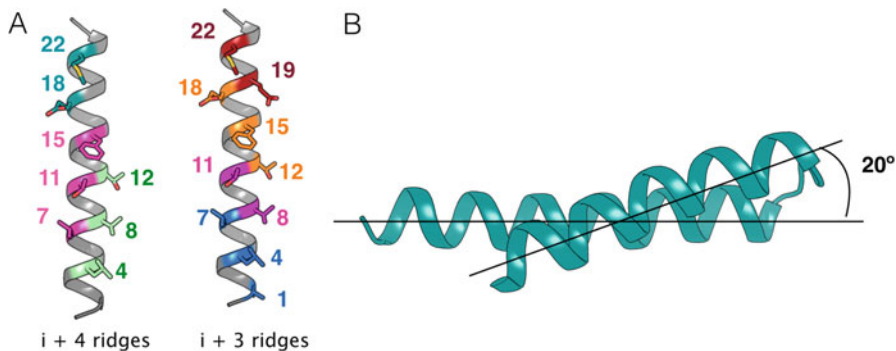
Fig. 1.12 The ridges-in-grooves model of helix packing. (**a**) The two ways to ridges (amino acid side chains) on the surface of alpha helices: the $i + 4$ model, forming parallel lines of 25°, and the $1 \neq 3$ model, forming parallel lines of −45°. Each ridge is formed by residues of the same color. (**b**) The $i + 4$ and $i + 3$ ridges/lines packing results in an angle of 20° between the packed helices
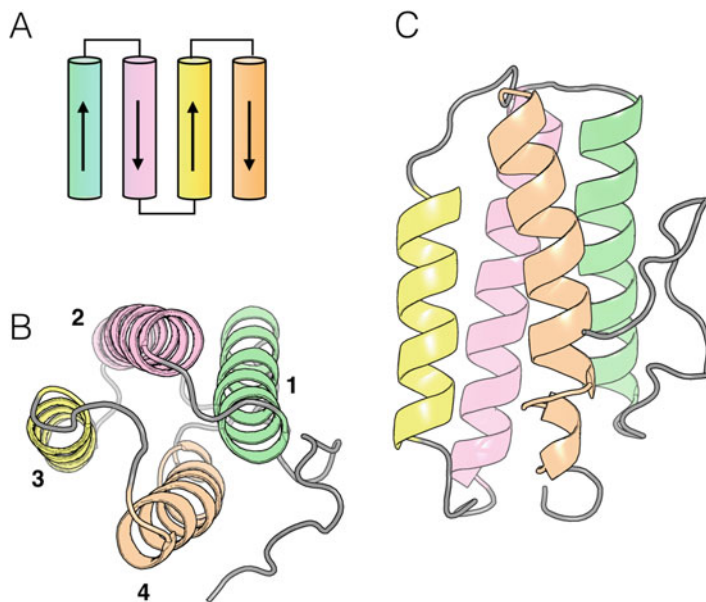

Fig. 1.13 α-domains: the 4-helix bundle of myohemerythrin (PDB 2MHR). (**a**) The up-and-down topology of the 4-helix bundle in myohemerythrin. (**b**) Top view of the 4-helix bundle. The helices are numbered from N- to C-terminus. (**c**) Side view of the 4-helix bundle

growth hormone, are formed only by the four-helix bundle. Another common α fold is the **globin fold**, in which the 50° packing angle is found between some of its eight helices.

In **β-domains**, the β-strands are arranged in antiparallel β-sheets because there are no helices to make the kind of connections observed in parallel β-sheets. There
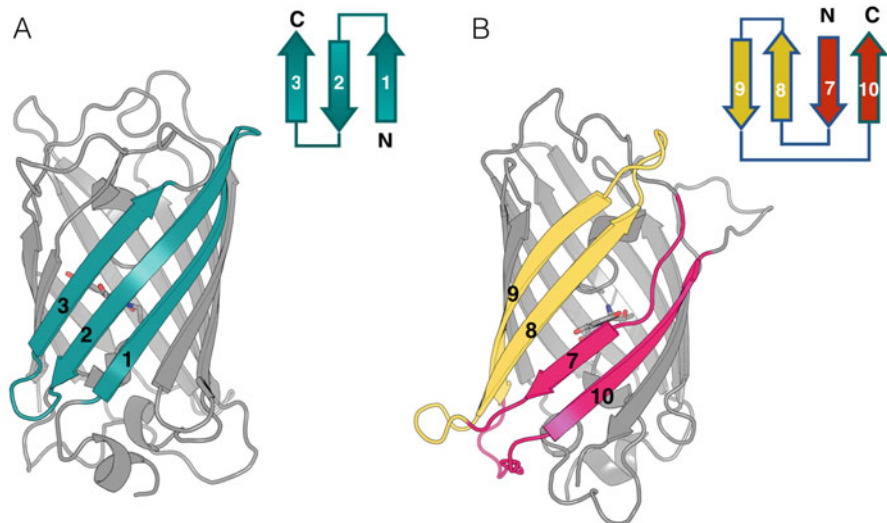
Fig. 1.14 Illustration of beta motifs in the structure of the green fluorescent protein (GFP), PDB: 1EMA. The structure is shown in cartoon representation, except for the chromophore, which is shown as sticks. The *top-right panels* illustrate the topologies of the highlighted motifs. (**a**) Strands 1–3 in the structure of GFP form an antiparallel up-and-down motif, as well as strands 4–6 (not shown). (**b**) Strands 7–10 in the structure of GFP form a Greek key motif

are two major types of connectivity in β-domains: up-and-down and Greek key motifs, both of which can be found in the two major types of β fold: the β-barrel and β-sandwich.

In the **up-and-down** motif, the β-strands go up and down as the name suggests, forming sequential antiparallel interactions. The **Greek key**, a structural motif of antiparallel β-sheets named after a common decorative element found in ancient Greek vases, is slightly more intricate than its up-and-down counterpart. To understand how a Greek key is organized, imagine a segment of four β-strands of similar size numbered from 1 to 4 (from N- to C-terminus). Then fold it in half between strands 2 and 3 forming a hairpin structure, so that strands 2 and 3 are hydrogen bonded in an antiparallel fashion, as well as strands 1 and 4. Finally, fold it again in half so that strands 3 and 4 also engage in antiparallel interactions. A single antiparallel β-domain can be formed by up-and-down motifs, Greek keys, or combinations of these motifs, such as in the green fluorescent protein (Fig. 1.14).

A **β-barrel** is a continuous antiparallel β-sheet where the first and last strands are hydrogen bonded to each other, closing the structure in the shape of a barrel. This kind of structure fulfills most if not all of the hydrogen bonding potential of the main chain within itself. This makes it especially suited for transmembrane proteins because no polar groups are left without hydrogen bonding, which would be unfavorable in the hydrophobic interior of membranes (considering that the side chains are all hydrophobic in the face exposed to lipids). In fact, transmembrane domains made of β-sheets often form barrels, which make good pores because they

can have a polar surface in the inner side of the barrel and a hydrophobic surface in the outside. β-barrels can also be found in soluble proteins, one such example is the green fluorescent protein (GFP) described above.

A **β-sandwich** is characterized by two β-sheets packed against each other, reminiscent of two slices of bread. These sheets usually display a polar surface exposed to solvent and a hydrophobic surface facing the other sheet, forming the hydrophobic core of the domain. In each of these interacting sheets, the strands located in the extremities have exposed polar groups in their main chains which can engage in hydrogen bonding with solvent or with side chains. The immunoglobulin fold is a typical example of β-sandwich fold.

α/β domains usually display variations on the theme of β-α-β structural motif. The β-sheets are either parallel or mixed. When they are parallel, the α-helix functions as a connectivity element between two strands. The connectivity of β-α-β structural motif can be either left-handed or right-handed, but due to the right-handed twist of the β-sheet, it is found to be right-handed in more than 95% of protein structures. α/β domains display two types of fold: the α/β barrel and the α/β twist.

The **α/β barrel** is formed by tandem repeats of the α-β-α motif forming a closed barrel-like structure. A minimum of four repeats is able to form this kind of structure, but the presence of eight strands in the barrel confers highest stability. The core of the barrel is formed by the parallel β-strands organized as a closed β-sheet, and the outer side is occupied by the connecting amphipathic α-helices which have a hydrophobic side packed against the sheet and an outer polar side facing the solvent. Although the α/β barrel might seem to have an inner hole when represented as cartoon, its core is usually filled with hydrophobic side chains from the β-strands. The eight-stranded version of this fold is called the **TIM barrel** because it is found in the enzyme triosephosphate isomerase (TIM). It is the most common fold, occurring in 10% of enzyme structures (Fig. 1.15). These enzymes, despite sharing a common fold, display a broad range of enzymatic activities.

Another common fold in α/β proteins is the **α/β twist**, also known as the **nucleotide-binding fold** because it is present in many nucleotide-binding proteins. It differs from the α/β barrel mostly because of its open structure. The core is formed by an open parallel β-sheet twisted in a saddlelike shape. A classical example of this category is Rossmann fold (Fig. 1.16) that consists of basic **β-α-β repeat** unit connected by loops. This is one of the most common folds present in most of the enzymes involved in nucleotide binding.

α + β domains are characterized by the presence of α-helices and β-sheets, but here these structural elements are found as segregated entities instead of the alternating pattern of helices and sheets found in α/β domains. This allows greater structural variability in these domains, because there are no special organizing principles underlying their structures. Winged helix domain is an example of α + β domain structure found in DNA-binding proteins. The domain is formed by an association of three alpha helices with three beta-sheets that interacts with the major groove of DNA. The two small winglike structures can interact with different regions on DNA. Several transcription factors such as LexA, arginine, or biotin
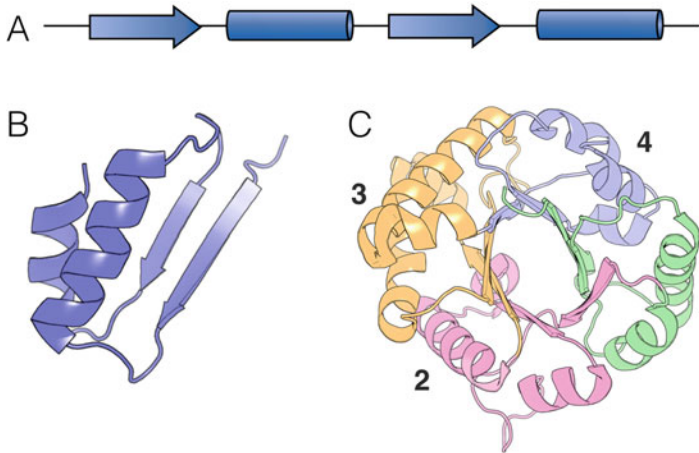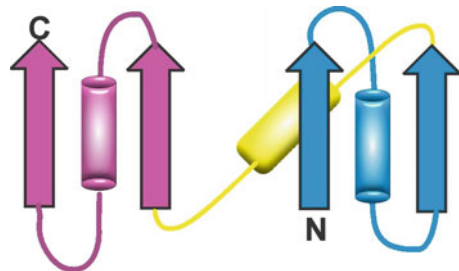
Fig. 1.15 α/β structures: the TIM barrel. (**a**) Schematic representation of a β-α-β-α motif. A TIM barrel fold is made of four tandem repeats of this motif. (**b**) Structure of the β-α-β-α motif of a TIM barrel from yeast. (**c**) The TIM barrel fold (PDB 1YPI). The motifs are numbered from N- to C-terminus

**Fig. 1.16** Rossmann fold

repressor proteins contain winged helix domain where it plays an important role in protein–protein interactions (Fig. 1.17)**.**

1.2.4   Quaternary Structure

The arrangement of proteins in oligomeric supramolecular assemblies defines what we know as **quaternary structure**. These can be **homooligomers** (made of identical subunits) or **heterooligomers** (made of distinct subunits which can be structurally related or completely unrelated). There are specific names to reflect the number of subunits in the oligomeric assembly: dimer, trimer, tetramer, pentamer, and hexamer, indicating the presence of two, three, four, five, or six subunits, respectively. Oligomeric assemblies of proteins are also called **protein complexes**.

Fig. 1.17 Winged helix cartoon representation (**a**) and topology (**b**)
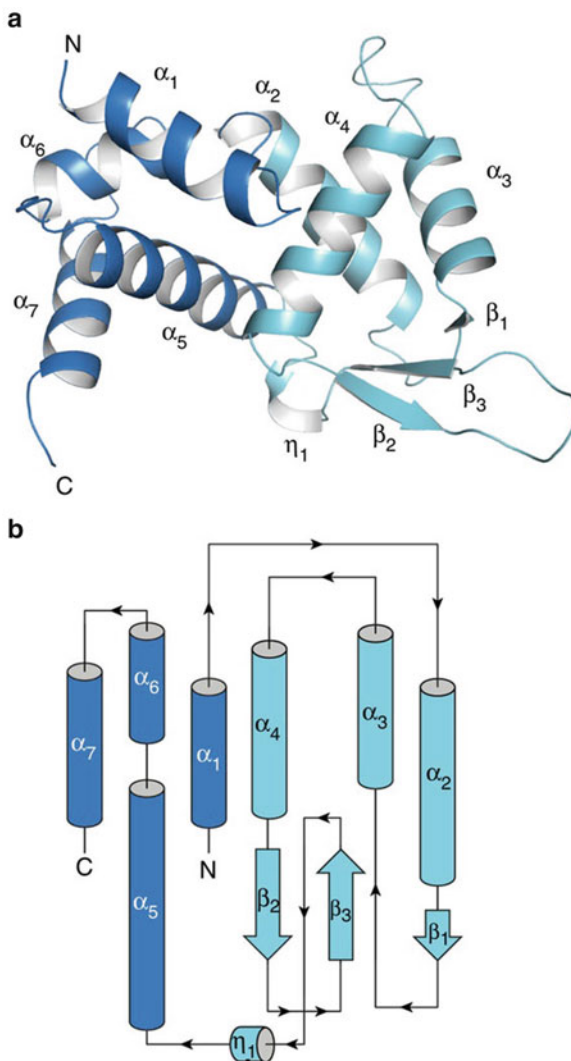
Many heterooligomers are made of structurally related subunits that sometimes do not display significant sequence similarity. Their evolutionary origin involves gene duplication and divergence. However, contrasting with the evolution of multidomain proteins, in the case of oligomers there is no gene fusion so the peptide chains remain as separate entities.

The forces that hold oligomers together are the same types of noncovalent forces that stabilize the tertiary structure. The subunits display complementary surfaces allowing them to engage in noncovalent interactions which stabilize the interfaces.

Fig. 1.18 Quaternary structure: example of a pseudosymmetric heterodimer, the MP1-p14 adaptor complex (PDB: 1VET). The subunits MP1 and p14 are shown separately (*above*) and then in the complex (*below*). Each one of them has a roadblock fold
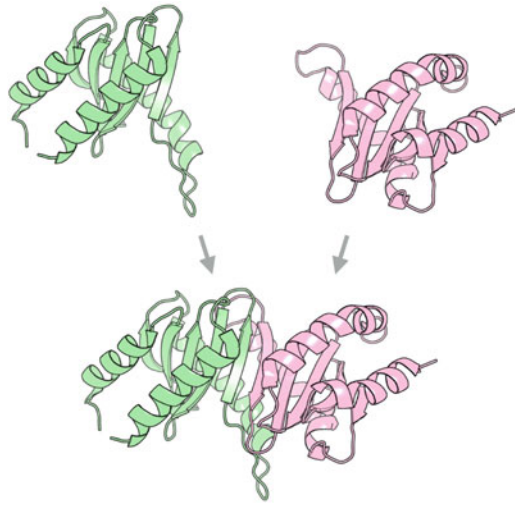
In extracellular and secreted proteins, disulfide bridges can help to stabilize the interfaces of oligomers.

The relative strength of binding between subunits of a complex is variable, and there are several ways to measure it. Whenever the high resolution structure of the complex is available, it is possible to measure the interface area with high accuracy, which is a strong predictor of the biological relevance of the interface. Some protein complexes have high dissociation constants and are found in equilibrium under normal temperature and pH, while others are virtually indissociable reflecting a small dissociation constant and tight packing of subunits.

Homooligomers have a strong tendency to form **symmetric structures**. Homodimers are always symmetrical and are able to form parallel or antiparallel arrangements (also known as head-to-tail dimers). Homotrimers usually have a central threefold rotation axis. Tetramers often form tetragonal structures, but they can also be arranged as a planar assembly with a central fourfold symmetry axis. Planar assemblies with central symmetry axis of the same order of the number of subunits are also frequently found in pentamers, hexamers, and heptamers. Another frequent way to assemble hexamer is a trimer of dimer with central three fold axis (Misra et al. 2009). In the case of heterooligomers of structurally related subunits, these symmetric arrangements have their **pseudosymmetric** counterparts. Hemoglobin, for example, is a pseudosymmetric tetramer made of two α and two β subunits which are nearly identical. Another example of pseudosymmetric quaternary assembly is the MP1–p14 adaptor complex, a dimer of two roadblock folds (Fig. 1.18) (Kurzbauer et al. 2004). Even protein complexes made of unrelated subunits, which are usually asymmetrical, can form symmetric assemblies if the entire unit repeats itself at least once around a symmetry axis.

1.2.5   Forces Stabilizing Protein Structures

Protein folding is the process by which an extended polypeptide chain adopts its native three-dimensional conformation. Although it is believed that a protein's primary sequence contains all the information necessary to define its fold, predicting a protein's structure ab initio with high accuracy is still a major unsolved problem in biophysics (Murphy 2001).

A landmark experiment performed by Christian B. Anfinsen (1973) is considered to be the fundamental demonstration that protein folding can occur spontaneously and that a protein can adopt its native conformation in the absence of any information other than its own primary sequence. In this experiment, he used the bovine pancreatic ribonuclease A (RNAse A) enzyme. The enzyme was denatured in the presence of 8 M urea (a chaotropic agent) and β-mercaptoethanol (a reducing agent to break the S–S bonds). Upon removal of the denaturing agents and addition of oxidizing agents to promote S–S bonds, the protein was able to fully recover its enzymatic activity. RNAse A has eight cysteine residues which form four S–S bonds in defined positions, which characterize the native state of this enzyme. Theoretically, these eight S–H groups could be reorganized in 105 different S–S pairs upon a cycle of reduction and oxidation; nevertheless, they go on to reconstitute exactly the same pairs which were originally present.

Thermodynamically, the simplest way to describe the folding process is to consider the existence of discrete native (N) and unfolded (U) states for a given protein which are in equilibrium. The folded state is energetically favored (displays a lower Gibbs energy) under appropriate conditions such as temperature, ionic strength, and pH, and this lower energetic state drives the folding reaction. The thermodynamic equilibrium hypothesis requires that the unfolding reaction is reversible, although, in practice, many proteins do not show this kind of behavior: instead, they aggregate or precipitate upon denaturation which renders the process irreversible.

By assuming the thermodynamic equilibrium between states, we can define an equilibrium constant for the unfolding reaction $\mathbf{K = [U]/[N]}$ and then use this constant to calculate the Gibbs free energy involved in the transition from folded to unfolded states as $\mathbf{dG° = -RT \ln K}$, where R is the universal gas constant and T is the absolute temperature in Kelvin. The Gibbs free energy has both **enthalpic (H)** and **entropic (S)** contributions described by the equation $\mathbf{dG° = dH° - TdS°}$.

In thermodynamics, **entropy (S)** is defined as the natural logarithm of the number of microscopic configurations of a system, multiplied by the Boltzmann constant (kB). According to this definition, the number of states in a system is directly proportional to its entropy. Intuitively, entropy can be understood as a measure of disorder. A system displaying an elevated degree of freedom is more entropic (disordered) than a comparable system in an ordered state. The **second law of thermodynamics** states that the entropy of an isolated system never decreases – it either increases over time or stays unchanged. In other words, spontaneous processes always involve a gain of entropy. This fundamental law underlies every process in the Universe and, not surprisingly, applies to biophysics as well.

A folded protein, even when some flexibility is allowed, is an ensemble of a relatively small number of configurations compared to an unfolded polypeptide chain, which samples large areas of conformational space. This means that, in terms of entropy, the folding of a protein is a largely unfavorable process because it forces the protein to restrict its conformational freedom, related to the rotation of main chain angles phi (φ) and psi (ψ) and side chains around χ angles. Obviously, it does not violate the second law of thermodynamics because the protein is immersed in solvent and the system "protein + solvent" always increases its entropy. However, there's an entropic cost which needs to be counteracted by other forces to allow a protein to fold.

The major drivers of protein folding, which counteract the negative effect of the loss of configurational entropy, are the **hydrophobic effect** and **hydrogen bonding**.

The noncovalent interactions which act to stabilize the structures of proteins can be explained in terms of electronegativity and polarity. **Electronegativity** is the tendency of an atom to attract electron density toward its nucleus. It depends both on the atomic number and the distance between the valence atoms and the atomic nucleus. In proteins, the most electronegative atom is oxygen, followed by nitrogen – their electronegativity values in the Pauling scale are 3.5 and 3.0, respectively. Hydrogen and carbon are more electropositive: their electronegative values are low (2.5 and 2.1, respectively) meaning that they are more likely to donate electron density to other atoms. For comparison, the most electronegative atom in the periodic table is fluorine (value = 4.0 in the Pauling scale). When a hydrogen atom is covalently bound to an atom with high electronegativity, this atom will affect the electrons around it generating a partial positive charge in the hydrogen atom. This effect, which applies to any covalent bond between atoms of different electronegativities, is the origin of **polarity**. The N–H and O–H groups are polar because the electron cloud is polarized toward the nitrogen or oxygen atom, which bears a partial negative charge, while the hydrogen atom displays a partial positive charge resulting in a **dipole**. In comparison, C–H groups are **apolar** because the electronegativities of C and H are both low, resulting in a more even distribution of the electron cloud.

Hydrogen bonds always involve a shared hydrogen atom between an atom to which it is covalently bonded (the donor) and another non-hydrogen atom which is negatively polarized (the acceptor). Both donor and acceptor are electronegative atoms, usually O and N in proteins. Hydrogen bonds allow the donor and acceptor atoms to be closer than they would be in the absence of such interaction. Whenever two non-hydrogen atoms are found less than 3.5 Å apart in a protein structure, it is assumed that they are involved in hydrogen bonding. The typical distance involved in hydrogen bonds is 3.0 Å, and the free energy ranges from 2 kJ/mole in water up to 21 kJ/mole if the donor or acceptor is ionized – in this case, it would be called a **salt bridge**, and the typical distance decreases to about 2.8 Å. A typical example of salt bridges in proteins involves the side chain of a positively charged residue such as lysine or arginine interacting with the side chain of negatively charged glutamate or aspartate.

Van der Waals interactions are weak electrostatic interactions involving induced dipoles in the fluctuating electron clouds of atoms or groups of atoms. They typically occur at a distance range of about 3.5 Å and drop rapidly with distance (r) following a $1/r^6$ dependence. These interactions are common among methyl groups in aliphatic, hydrophobic side chains such as leucine, isoleucine, and valine which are highly polarizable. The free energy involved in a Van der Waals interaction is small (only ~4 kJ/mole), but the net effect of all these interactions in a single protein can be quite large, often reaching hundreds of kJ/mole.

The **hydrophobic effect** relates to the energetics involved in the solvation of hydrophobic side chains. The hydrophobic effect has both entropic and enthalpic contributions. When hydrophobic molecules are in contact with water, the water molecules are oriented around the hydrophobic groups in a way that results in a decrease of entropy compared to the state of water molecules in the absence of any hydrophobic groups. This "cage" of ordered water molecules around exposed hydrophobic groups can be experimentally observed in some high-resolution protein structures. Therefore, the system will spontaneously rearrange so that the hydrophobic molecules are grouped together reducing the area of interface between hydrophobic and polar molecules, so that the water molecules will be able to adopt a more disordered state, increasing the entropy of the system. In a protein, this translates as a tendency for the hydrophobic side chains to occupy the interior of globular proteins, while the polar and charged side chains are found in the solvent-exposed areas.

Energetically, protein stability results from a balance of both stabilizing and destabilizing forces. The difference between them is usually small and is strongly influenced by environmental factors such as temperature pH and ionic strength; this is the major reason why most proteins have narrow ranges of stability and are easily denatured.

1.3   Conformation of Globular Proteins

Globular proteins are formed by organizing their polar groups at protein's surface, and nonpolar groups are directed toward the center of the protein. The globular proteins containing polar side chains exhibit strong interactions toward other polar groups of atoms within the protein molecule as well as toward polar molecules in the protein's surroundings. Similarly, nonpolar side chains have attraction toward other nonpolar side chains within the protein. These proteins have high aqueous solubility that helps them perform diverse biological functions such as enzymatic catalysis, antibodies, DNA replication, and repair.

1.3.1   Hemoglobin

It is the most characterized protein present in red blood cells in humans. The structure was first determined by Max Perutz in 1959 (Perutz et al. 1960). The

Fig. 1.19 Hemoglobin structure
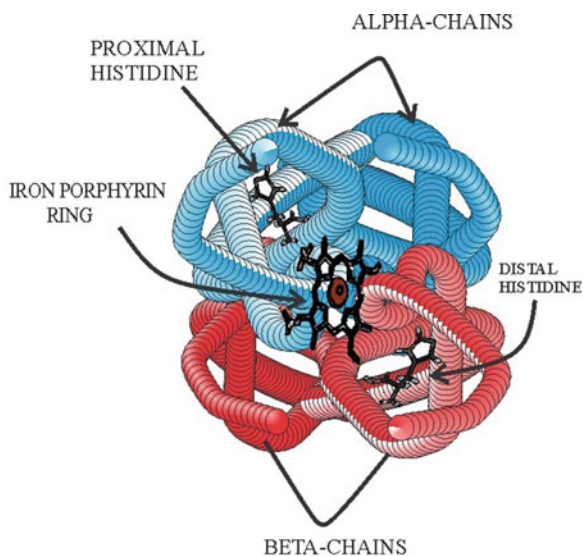
quaternary structure of hemoglobin consists of four polypeptide chains – two identical α-chains consisting of 141 amino acid residues and two identical β-chains consisting of 146 amino acid residues contributing toward a molar mass of 64,500. Hemoglobin is one of the members of the globin superfamily. The globin fold consisting of eight alpha helices is an example of all alpha fold. This fold dictates the functional properties of globin superfamily members including hemoglobin and myoglobin. In hemoglobin, this fold undergoes heterodimerization to form a tetramer with a centrally located iron-protoporphyrin IX ring coordinated to a heme prosthetic group (Fig. 1.19). The iron is in the physiological ferrous state coordinated to the four pyrrole nitrogen atoms in one plane, to an imidazole nitrogen atom of His 8 in the "F" helix, and to a gas atom opposite to this His residue. Each molecule of gas binds to these four ferrous ions in the globin chain respectively accounting for the transport of O_2, CO, and NO. Carbon dioxide, however, binds to amino-terminal of hemoglobin in place of iron atoms forming a weak carbamino complex and transported in blood. The oxygen binding to hemoglobin is represented by a sigmoidal curve (Fig. 1.20) defined by the following equation:

$$Y_{O_2} = \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n}$$

The sigmoidal curve indicates the presence of allosteric cooperativity in oxygen binding where one molecule of oxygen binding to one heme subunit induces conformational changes in other subunits thus promoting the binding of three more molecules of oxygen to the other three subunits of hemoglobin. The deoxy
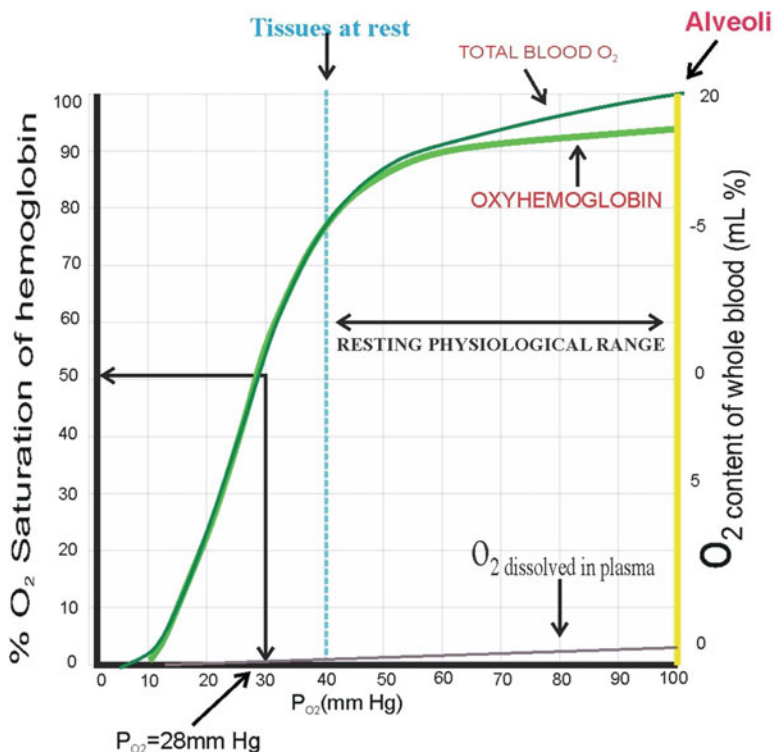
Fig. 1.20 Hemoglobin saturation curve

hemoglobin has heme present in the subunits in T (tensed) form which is the low affinity state for O_2 binding. Once the O_2 molecule binds to any of the subunits, there is a conversion of T to R (relaxed) form having high affinity for O_2 molecule.

There are significant tertiary and quaternary structure changes responsible for this conversion. $Fe + 2$ is pulled in to the plane of the porphyrin ring by 0.039 nm when oxygen binds with the help of His 8 in the "helixF." One alpha–beta pair moves relative to the other by $15°$ upon oxygen binding resulting in the breaking of ionic bonds in alpha chains and in beta chains in the R-state decreasing pK's of sidechains that releases protons. Release of oxygen in tissues is further regulated by the allosteric binding of 2,3-bisphosphoglycerate to the beta chain of hemoglobin that has more affinity for deoxy hemoglobin. The Hill coefficient for the oxygen binding to hemoglobin is 2.8–3, emphasizing the binding cooperativity. The functioning of hemoglobin is further described by Bohr's effect which postulates that when hemoglobin binds O_2 there is a release of protons.

$$Hb\,(O_2)n\,Hx + O_2 == Hb(O_2)n + x[H+].$$
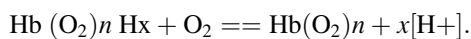
When the oxygenated blood reaches tissues, there is a release of oxygen because there is low pH in cells due to high concentration of protons. The increased respiratory activity inside the cells results in the increase in the concentration of CO_2. The enzyme carbonic anhydrase present in red blood cells converts CO_2 to carbonic acid which further dissociates to form bicarbonate ions and protons, thus resulting in the release of oxygen from hemoglobin.

There are different types of hemoglobin based on the type of chains comprising the hemoglobin molecule. In erythrocytes of normal human adults, hemoglobin A ($\alpha2\beta2$) is majorly present (97%) in blood, hemoglobinA2 with $\alpha2$ $\delta2$ (2%) and hemoglobin F or fetal hemoglobin with $\alpha2\gamma2$ chains (1%). Hemoglobin binds oxygen in lungs at high PO_2 of100 mm Hg, which is the high affinity state for oxygen and releases it to metabolically active cells at a lower PO_2 of 30 mm Hg that signifies the low affinity state for oxygen in tissues. Besides, its main role of oxygen transport hemoglobin also binds carbon dioxide (CO_2), carbon monoxide (CO), and nitric oxide (NO).

1.3.2 Myoglobin

Myoglobin is the heme containing protein found in vertebrate skeletal and cardiac muscle. This is the first protein crystal structure determined using X-ray crystallography by John Kendrew and colleagues in 1958. It has a single polypeptide chain consisting of 153 amino acid residues that binds and stores oxygen in skeletal muscles from blood and is the source of continuous oxygen supply to active muscle groups during respiration. A high degree of conservation is observed in the tertiary and ligand binding regions of myoglobin and other globin specifically the hydrophobic heme pocket; the distal pocket containing distal H64 and proximal H93 along with four other pockets spanning from Xe1 to Xe4 is conserved across species (Fig. 1.21). Presence of the porphyrin ring in the plane of the heme pocket is characteristic of the globin proteins. Noncovalent interactions involving hydrophobic interactions with nonpolar amino acids and electrostatic interactions between the heme propionic sides with polar amino acids near the cavity including H97, R45, and S92 stabilize the structure of the myoglobin. The backbone structure of myoglobin is stable, whereas the side chains exhibit conformational dynamism required for binding oxygen and other ligands. The proximal histidine H93 is covalently bound to the iron at the center of the heme, while the distal histidine H64 is known for stabilizing the oxygen–heme bond. The distal pocket is the gap adjacent to the heme iron, and the distal histidine serves as a gate allowing the heme to bind oxygen and other ligands. The variations observed in globin structure are responsible for the differential binding of oxygen. It has high oxygen affinity which reaches saturation at low PO_2 (2 mm Hg). The partial pressure of oxygen in capillaries is 30 mm Hg where myoglobin easily binds oxygen beyond saturation. However, it releases oxygen in metabolically active cells where $PO_2 < 2$ mm Hg. The oxygen binding curve of myoglobin is hyperbolic (Fig. 1.22) defined by the following equation:
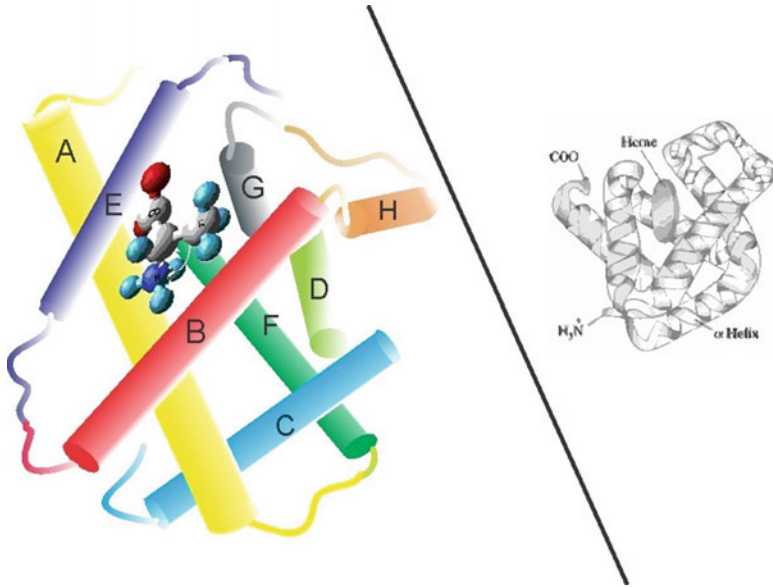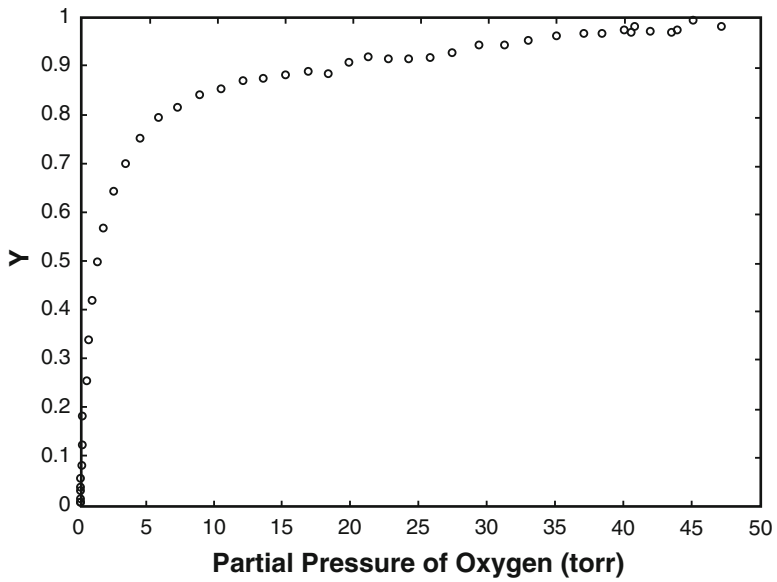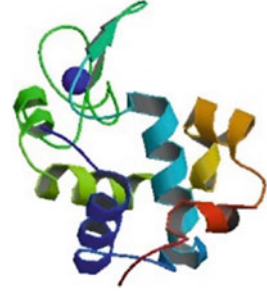
Fig. 1.21 Myoglobin structure

Fig. 1.22 Myoglobin saturation curve (Barlow et al. 1992)

Fig. 1.23 Lysozyme
structure

$$Y_{O_2} = \frac{(pO_2)}{p_{50} + (pO_2)}$$

Scientific research has provided insights related to other roles of myoglobin. It regulates the intramuscular bioavailability of nitric oxide (NO), thereby influencing mitochondrial respiration. The oxygen availability in the intracellular milieu is responsible for the transitioning of myoglobin from NO scavenger (at high PO_2) to a NO producer as PO_2 decreases.

1.3.3 Lysozyme (1,4-β-N-Acetylmuramidase)

This is the widely studied protein to understand the effect of various factors on protein structure. It consists of a single polypeptide chain with 129 amino acids (Fig. 1.23). It folds into a compact three-dimensional structure with a long cleft present on the surface under physiological conditions. It is present across species ranging from plants, animals, and humans. The mucous lining of the nasal cavity, tear ducts, kidney tissue, milk, leukocytes, saliva, etc. are some of the places where lysozyme is expressed. The main function of this enzyme is the hydrolysis of the bond linking N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) resulting in the increase of bacterial cell wall permeability. This has important implications in the prevention of various bacterial infections.

1.3.4 Cytochromes

Cytochromes belong to a class of hemoproteins involved in the transport of electron or proton transport by reversibly changing the valency of iron atom present in the heme group. There are different types of cytochromes, namely, a, b, c, and d, depending on their spectrochemical characteristics. The existence of different variants of these basic types of cytochromes has been reported in bacteria, green plants, and algae (e.g., cyt f is a variant of cyt c). The mitochondrial system of

Fig. 1.24 Cytochrome structure
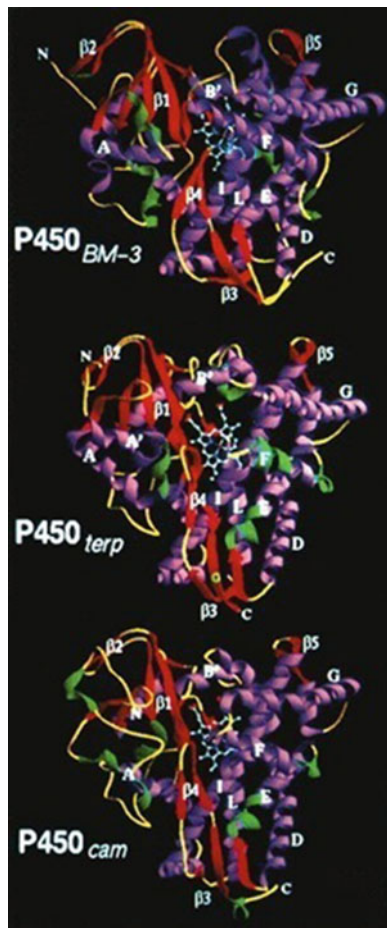
cytochromes provides electron transport through cytochrome c oxidase to molecular oxygen as the terminal electron acceptor (respiration). Cytochromes P450 belonging to the superfamily of heme thiolate enzymes have been extensively characterized. Despite a significant difference in the primary sequence of various P450 enzymes, the functional tertiary fold is conserved. This P450 fold contains 13 α-helices (A, B, B′, and C-L) and five β-sheets (31–135) (Fig. 1.24). They are involved in a variety of functions including biosynthesis of steroids, fats, bile acids, monooxygenation of hydrophobic substrates, hydroxylation of bicyclic monoterpene camphor, drug and xenobiotic metabolism, etc. They have been classified as class I and class II P450s based on their sequence similarity and the nature of electron donor which is iron–sulfur proteins in case of class I and FAD/FMN-containing reductase proteins in class II enzymes.
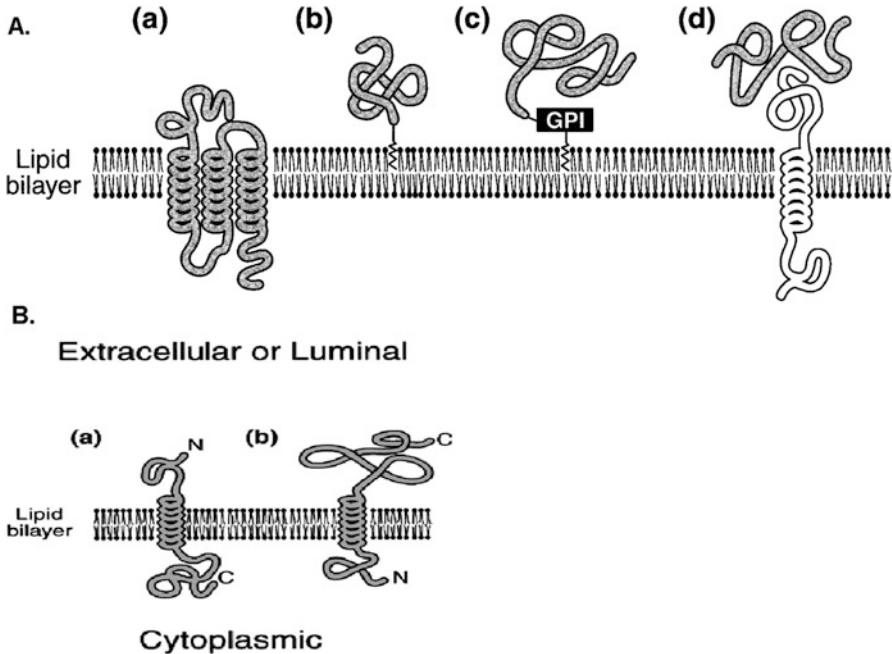
Fig. 1.25 Integral and peripheral membrane proteins (Chou et al. 1999)

1.4 Membrane Proteins

All the eukaryotic cells and their constituent organelles are enclosed within the plasma membrane. The structural organization of this membrane is a lipid bilayer interspersed with membrane proteins. These proteins have key functional roles spanning from cell signaling to ion transport. Membrane proteins undergo extensive oligomerization to enhance their stability and genetic efficiency, e.g., cytb6f. These membrane proteins are classified as integral or peripheral proteins depending upon the strength of their association with the plasma membrane (Fig. 1.25).

1.4.1 Integral Proteins

These proteins pass the lipid bilayer either once (single pass proteins) or multiple times (multiple pass proteins). The single pass proteins are classified as type I proteins having extracellular (or luminal) amino terminus and cytosolic carboxy terminus. The multiple pass proteins are also referred to as type II with luminal carboxy terminus and cytosolic amino terminus. Thus, they are capable of functioning on both sides of the lipid bilayer, e.g., ion channels (Chou et al. 1999). The

integral proteins are enriched in alpha helical content as their presence in lipid bilayer facilitates the intrachain hydrogen bonding characteristic of alpha helix. These proteins are majorly composed of hydrophobic residues. Many transmembrane motifs such as four-helix bundle and beta barrel are the predominant super secondary structures present in these proteins. The classical example of all alpha helical membrane protein is bacteriorhodopsin with serpentine receptor (seven helices spanning the lipid bilayer). Another protein, namely, porin, is an example of all beta strand structure. Most of the channel proteins are made up of antiparallel beta strands forming the beta-sheet where the surface-exposed residues are mostly hydrophobic, and the hydrophilic residues are present inside the channel where water molecules are present (Rees et al. 1989).

There is also a class of proteins termed as anchored proteins which are covalently linked to the bilayer by either fatty acid/ lipid chains having prenyl groups (lipid anchored) or by glycosylphosphatidylinositol (GPI anchored). There are sequence-specific features of such anchored proteins having isoprenylation at carboxy terminus with CAAX consensus motif, myristoylation at the amino terminus with conserved GXXXS/T motif, or palmitoylation at the same terminus at specific cysteine residue. However, GPI-anchored proteins have a GPI residue attached at the carboxy terminus with a hydrophobic tail (Chou et al. 1999).

1.4.2   Peripheral Proteins

These proteins are attached through noncovalent interactions either to the hydrophobic head of the lipid or to the domains of integral proteins exposed to the surface. In the former case they are referred to as amphitropic proteins. They do not have any significant consensus sequences and can be disrupted using detergents or change in pH unlike the integral proteins. Many proteins such as synucleins, several annexins, enzymes, namely, lipoxygenases, etc., are included in this category.

References

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230
Barlow CH, Kelly KA, Kelly JJ (1992) Simultaneous determination of hemoglobin and myoglobin oxygen binding curves by spectral curve fitting. Appl Spectrosc 46(5):758–763
Branden C, Tooze J (1999) Introduction to protein structure, 2nd edn. Garland Publishing Group, New York
Chou K-C, Elrod DW (1999) Prediction of membrane protein types and subcellular locations. Proteins 34:137–153. doi:10.1002/(SICI)1097-0134(19990101)34:1<137::AID-PROT11>3.0.CO;2-O

Cooley RB, Arp DJ, Karplus PA (2010) Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. J Mol Biol 404:232–246

Crawford JL, Lipscomb WN, Schellman CG (1963) The reverse turn as a polypeptide conformation in globular proteins. Proc Natl Acad Sci U S A 70:538–542

Kurzbauer R et al (2004) Crystal structure of the p14/MP1 scaffolding complex: how a twin couple attaches mitogen-activated protein kinase signaling to late endosomes. Proc Natl Acad Sci U S A 101:10984–10989

Kyte J (1991) Structure in protein chemistry. Garland Publishing Group, New York

Levine TP et al (2013) Discovery of new longin and roadblock domains that form platforms for small GTPases in ragulator and TRAPP-II. Small GTPases 4:62–69

Misra G, Aggarwal A, Dube D, Zaman MS, Singh Y, Ramachandran R (2009) Crystal structure of the Bacillus anthracis nucleoside diphosphate kinase and its characterization reveals an enzyme adapted to perform under stress conditions. Proteins Struct Funct Bioinform 76:496–506

Murphy KP (2001) Chapter 1: Stabilization of protein structure in protein structure, stability, and folding. In Methods in molecular biology, vol 168. Humana Press

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Pauling L, Corey RB (1951) The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci U S A 37:251–256

Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37:205–211

Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å. Resolution, obtained by X-Ray analysis. Nature 185:416–422

Petsko GA, Ringe D (2004) Protein structure and function. New Science Press

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99

Rees DC, DeAntonio L, Eisenberg D (1989) Hydrophobic organization of membrane proteins. Science 245(4917):510–513

Scorsato V et al (2016) Crystal structure of the human Tip41 orthologue, TIPRL, reveals a novel fold and a binding site for the PP2Ac C-terminus. Sci Rep 6:30813

Venkatachalam CM (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 6:1425–1436