B2 Protein structure and function

Key Notes

Proteins: an overview

Proteins perform a variety of functions, including the transport and storage of other molecules, mechanical support, generation of movement, immune protection, acting as catalysts, involvement in cellular signaling, and transmission of nerve impulses. Each protein has a unique amino acid sequence that is genetically determined.

Peptide bond

A protein is a linear sequence of amino acids linked together by peptide bonds. The peptide bond is a covalent bond between the α-amino group of one amino acid and the α-carboxyl group of another. The backbone conformation of a polypeptide is specified by the rotation angles about the C_α–N bond (*phi*, ϕ) and C_α–C bond (*psi*, ψ) of each of its amino acid residues. The sterically allowed values of ϕ and ψ are visualized in a Ramachandran plot.

Primary structure

The linear sequence of amino acids joined together by peptide bonds is termed the primary structure of the protein. The position of covalent disulfide bonds between cysteine residues is also included in the primary structure.

Secondary structure

Secondary structure in a protein refers to the regular folding of regions of the polypeptide chain. The two most common types of secondary structure are the α-helix and the β-pleated sheet. The α-helix is a cylindrical, rod-like helical arrangement of the amino acids in the polypeptide chain, which is maintained by hydrogen bonds parallel to the helix axis. In a β-pleated sheet, hydrogen bonds form between adjacent sections of polypeptides that are running either in the same direction (parallel) or in the opposite direction (antiparallel).

Tertiary structure

Tertiary structure in a protein refers to the three-dimensional arrangement of all the amino acids in the polypeptide chain. This biologically active, native conformation is maintained by multiple noncovalent bonds.

Quaternary structure

If a protein is made up of more than one polypeptide chain, it is said to have quaternary structure. This refers to the spatial arrangement of the polypeptide subunits and the nature of the interactions between them.

Protein stability

In addition to the peptide bonds between individual amino acid residues, the three-dimensional structure of a protein

	is maintained by a combination of noncovalent interactions (electrostatic forces, van der Waals forces, hydrogen bonds, hydrophobic forces) and covalent interactions (disulfide bonds).
Protein structure determination	The three-dimensional structure of a protein can be determined using complex physical techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryoelectron microscopy.
Protein folding	Proteins spontaneously fold into their native conformation, with the primary structure of the protein dictating its three-dimensional structure. Protein folding is driven primarily by hydrophobic forces and proceeds through an ordered set of pathways. Accessory proteins, including molecular chaperones, assist proteins to fold correctly in the cell. Diseases, such as Alzheimer's, are due to proteins misfolding.
Related topics	(B1) Amino acid structure (B4) Collagen (B3) Myoglobin and hemoglobin (B6) Antibodies

Proteins: an overview

Proteins are the most versatile macromolecules in living organisms and carry out a diverse array of functions. They **transport** and **store** other molecules (e.g. hemoglobin and myoglobin; Section B3), they provide **mechanical support** (e.g. collagen; Section B4), they **generate movement** (e.g. actin and myosin; Section B5), they provide **immune protection** (e.g. antibodies; Section B6), they act as **catalysts** (enzymes; Sections D1–D5), they are involved in **cellular signaling** (Section E5), and they **transmit nerve impulses** (Section E6). Each protein has a **unique amino acid sequence** that is genetically determined (Section H1). Most proteins contain between 50 and 2000 amino acids linked together via **peptide bonds**. With the average molecular weight of an amino acid being about 110 (Table 1 in Section B1), the **molecular weights** of most proteins are between 5500 and 220 000. We can also refer to the mass of a protein, which is expressed in units of **Daltons**, where one Dalton is equal to one atomic mass unit. Thus, a protein with a molecular weight of 25 000 has a mass of 25 000 daltons or 25 kDa.

Peptide bond

Proteins are linear sequences of amino acids linked together by peptide bonds. The peptide bond is a chemical, covalent bond formed between the α-amino group of one amino acid and the α-carboxyl group of another (Figure 1a) (Section B1). Once two amino acids are joined together via a peptide bond to form a dipeptide, there is still a free amino group at one end and a free carboxyl group at the other, each of which can in turn be linked to further amino acids. Thus, long, unbranched chains of amino acids can be linked together by peptide bonds to form oligopeptides (up to 25 amino acid residues) and polypeptides (>25 amino acid residues). Note that the polypeptide still has a free α-amino group and a free α-carboxyl group. Convention has it that peptide chains are written down with the free α-amino group on the left, the free α-carboxyl group on the right and a hyphen between the amino acids to indicate the peptide bonds. Thus, the tripeptide $^{+}H_3N$-serine–leucine–phenylalanine-COO^{-} would be written simply as Ser-Leu-Phe (in three-letter code) or S-L-F (in single-letter code).

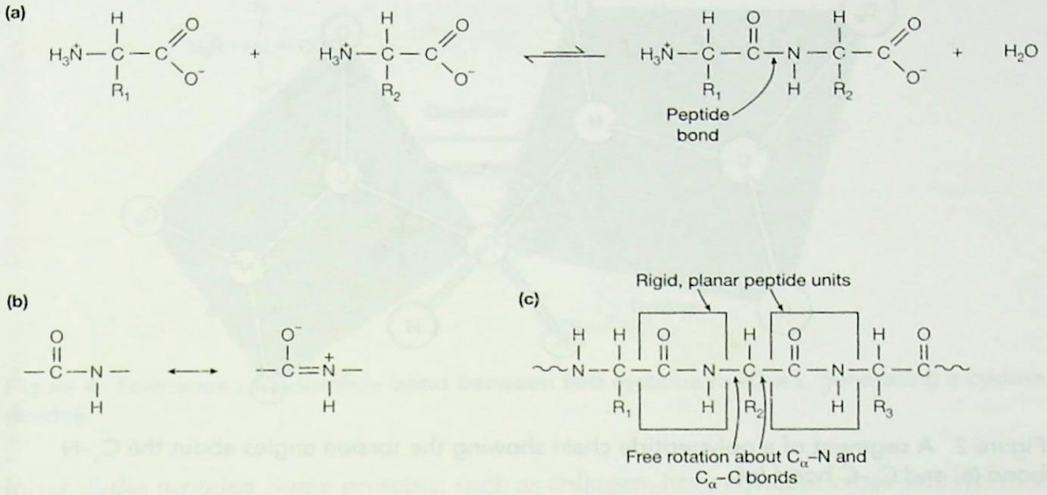

Figure 1. (a) Formation of a peptide bond; (b) resonance structures of the peptide bond; (c) peptide units within a polypeptide.

The peptide bond between the carbon and nitrogen exhibits **partial double-bond character** due to the closeness of the carbonyl carbon–oxygen double bond allowing the **resonance structures** in Figure 1b to exist. Because of this, the C–N bond length is also shorter than normal C–N single bonds. The **peptide unit**, which is made up of the CO–NH atoms is thus relatively rigid and planar, although free rotation can take place about the C_α–N and C_α–C bonds (the bonds either side of the peptide bond), permitting adjacent peptide units to be at different angles (Figure 1c). The hydrogen of the amino group is nearly always on the opposite side (*trans*) of the double bond to the oxygen of the carbonyl group, rather than on the same side (*cis*).

The backbone of a protein is a linked sequence of rigid planar peptide groups. The backbone conformation of a polypeptide is specified by the **rotation angles** or **torsion angles** about the C_α–N bond (*phi*, ϕ) and C_α–C bond (*psi*, ψ) of each of its amino acid residues. When the polypeptide chain is in its planar, fully extended (all-*trans*) conformation the ϕ and ψ angles are both defined as 180°, and increase for a clockwise rotation when viewed from C_α (Figure 2). The **conformational range** of the torsion angles, ϕ and ψ, in a polypeptide backbone is restricted by steric hindrance. The sterically allowed values of ϕ and ψ can be determined by calculating the distances between the atoms of a tripeptide at all values of ϕ and ψ for the central peptide unit. These values are visualized in a steric contour diagram, otherwise known as a conformation map or **Ramachandran plot** (Figure 3). From Figure 3, it can be seen that most areas of the Ramachandran plot (most combinations of ϕ and ψ) are conformationally inaccessible to a polypeptide chain. Only three small regions of the conformation map are physically accessible to a polypeptide chain, and within these regions are the ϕ–ψ values that produce the right-handed α-helix, the parallel and antiparallel β-pleated sheets and the collagen helix (see below and Section B4).

The polypeptide chain folds up to form a specific shape (**conformation**) in the protein. This conformation is the **three-dimensional arrangement** of atoms in the structure and is determined by the amino acid sequence. There are four levels of structure in proteins: **primary**, **secondary**, **tertiary** and, sometimes but not always, **quaternary**.

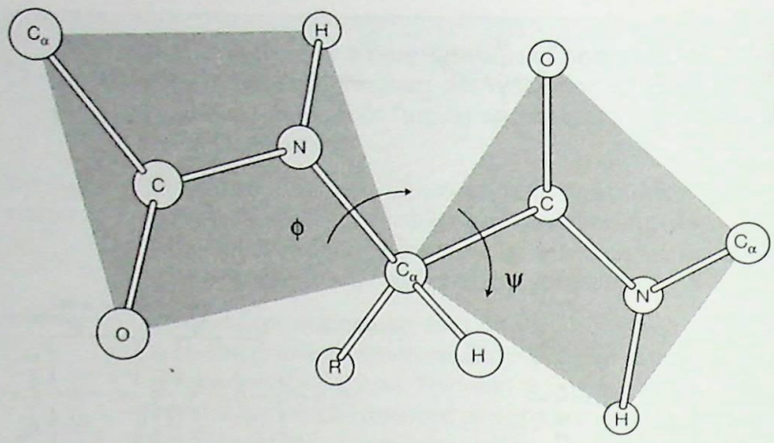

Figure 2. A segment of a polypeptide chain showing the torsion angles about the C_α–N bond (ϕ) and C_α–C bond (ψ).

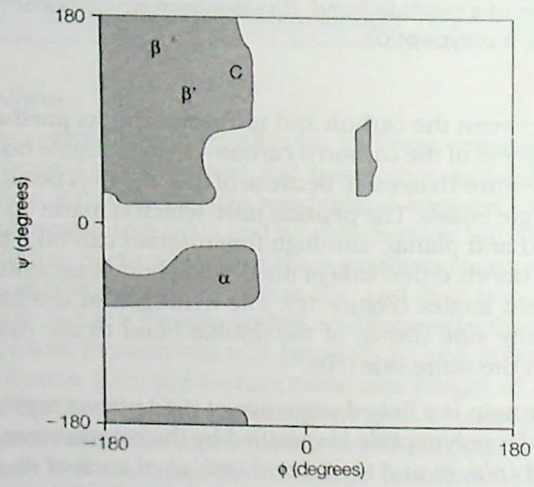

Figure 3. Ramachandran plot showing the allowed angles for poly-L-alanine (gray regions). α, ϕ–ψ values that produce the right-handed α-helix; β, the antiparallel β-pleated sheet; β', the parallel β-pleated sheet; C, the collagen helix.

Primary structure

The primary level of structure in a protein is the **linear sequence of amino acids** as joined together by peptide bonds. This sequence is determined by the sequence of nucleotide bases in the gene encoding the protein (Section H1). Also included under primary structure is the location of any other **covalent bonds**. These are primarily **disulfide bonds** between cysteine residues that are adjacent in space but not in the linear amino acid sequence. These covalent cross-links between separate polypeptide chains or between different parts of the same chain are formed by the oxidation of the SH groups on cysteine residues that are juxtaposed in space (Figure 4). The resulting disulfide is called a **cystine residue**. Disulfide bonds are often present in extracellular proteins, but are rarely found in

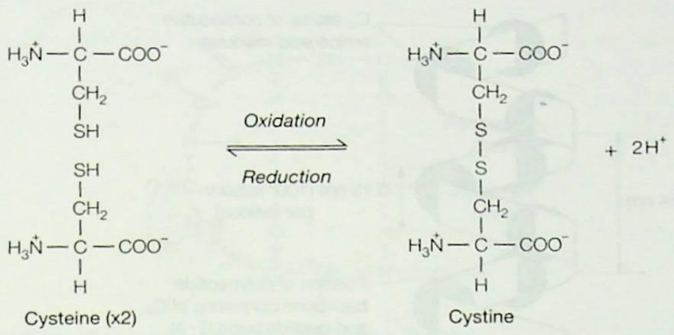

Figure 4. Formation of a disulfide bond between two cysteine residues, generating a cystine residue.

intracellular proteins. Some proteins, such as collagen, have covalent cross-links formed between the side-chains of Lys residues (Section B4).

Secondary structure

The secondary level of structure in a protein is the regular folding of regions of the polypeptide chain. The two most common types of protein fold are the **α-helix** and the **β-pleated sheet**. In the rod-like α-helix, the amino acids arrange themselves in a regular helical conformation (Figure 5a). The carbonyl oxygen of each peptide bond is **hydrogen bonded** to the hydrogen on the amino group of the fourth amino acid away (Figure 5b), with the hydrogen bonds running nearly parallel to the axis of the helix. In an α-helix there are 3.6 amino acids per turn of the helix covering a distance of 0.54 nm, and each amino acid residue represents an advance of 0.15 nm along the axis of the helix (Figure 5a). The side-chains of the amino acids are all positioned along the outside of the cylindrical helix (Figure 5c). Certain amino acids are more often found in α-helices than others. In particular, Pro is rarely found in α-helical regions, as it cannot form the correct pattern of hydrogen bonds due to the lack of a hydrogen atom on its nitrogen atom. For this reason, Pro is often found at the end of an α-helix, where it alters the direction of the polypeptide chain and terminates the helix. Different proteins have a different amount of the polypeptide chain folded up into α-helices. For example, the single polypeptide chain of myoglobin has eight α-helices (Section B3).

In the **β-pleated sheet**, hydrogen bonds form between the peptide bonds either in different polypeptide chains or in different sections of the same polypeptide chain (Figure 6a). The planarity of the peptide bond forces the polypeptide to be pleated with the side-chains of the amino acids protruding above and below the sheet (Figure 6b). Adjacent polypeptide chains in β-pleated sheets can be either **parallel** or **antiparallel** depending on whether they run in the same direction or in opposite directions, respectively (Figure 6c). The polypeptide chain within a β-pleated sheet is fully extended, such that there is a distance of 0.35 nm from one C_α atom to the next. β-Pleated sheets are always slightly curved and, if several polypeptides are involved, the sheet can close up to form a **β-barrel**. Multiple β-pleated sheets provide strength and rigidity in many structural proteins, such as silk fibroin, which consists almost entirely of stacks of antiparallel β-pleated sheets.

In order to fold tightly into the compact shape of a globular protein, the polypeptide chain often reverses direction, making a **β-turn** (hairpin or reverse turn). In these β-turns,

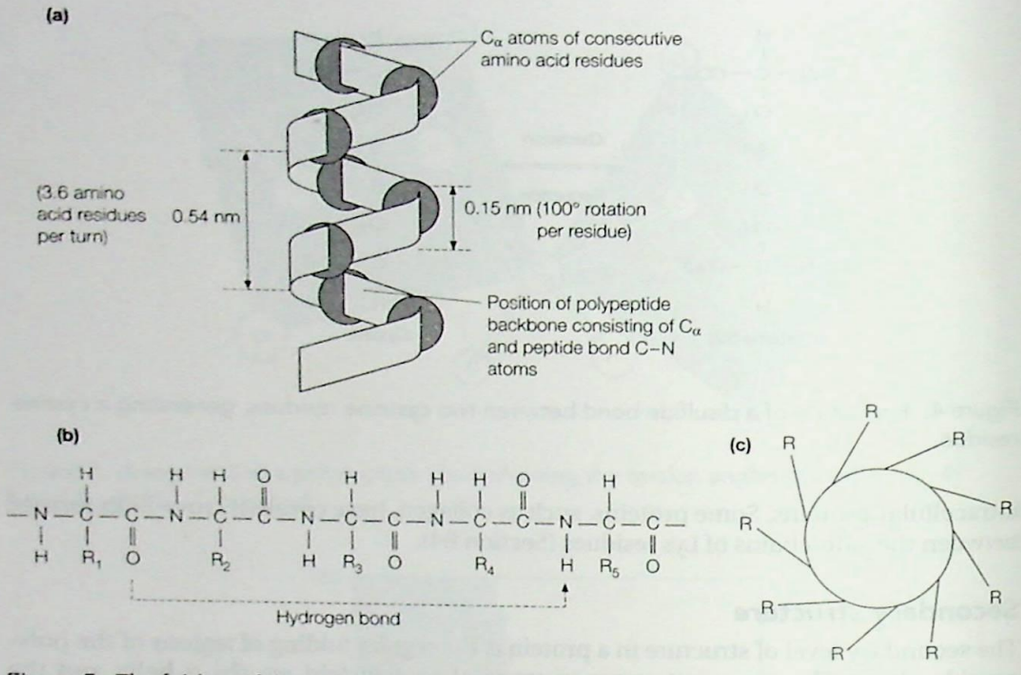

Figure 5. The folding of the polypeptide chain into an α-helix: (a) model of an α-helix with only the C_α atoms along the backbone shown; (b) in the α-helix the CO group of residue *n* is hydrogen bonded to the NH group on residue (*n*+4); (c) cross-sectional view of an α-helix showing the positions of the side-chains (R groups) of the amino acids on the outside of the helix.

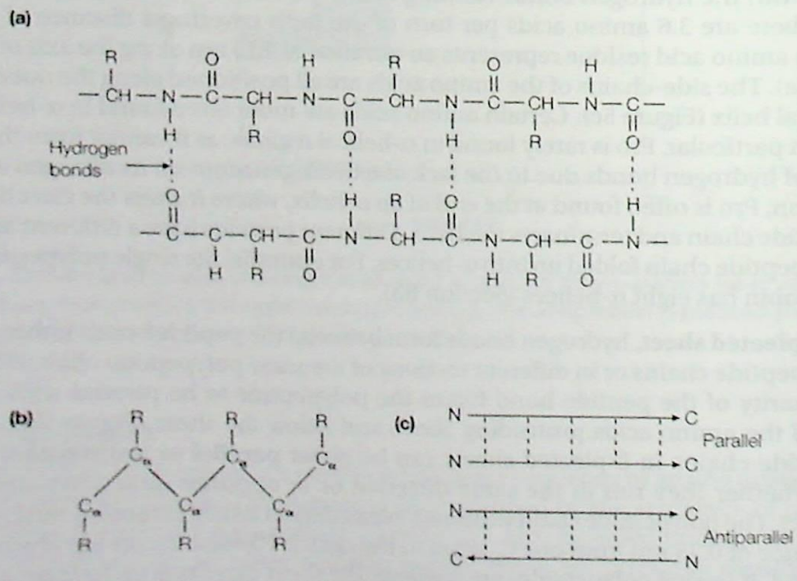

Figure 6. The folding of the polypeptide chain in a β-pleated sheet: (a) hydrogen bonding between two sections of a polypeptide chain forming a β-pleated sheet; (b) a side-view of one of the polypeptide chains in a β-pleated sheet showing the side-chains (R groups) attached to the C_α atoms protruding above and below the sheet; (c) because the polypeptide chain has polarity, either parallel or antiparallel β-pleated sheets can form.

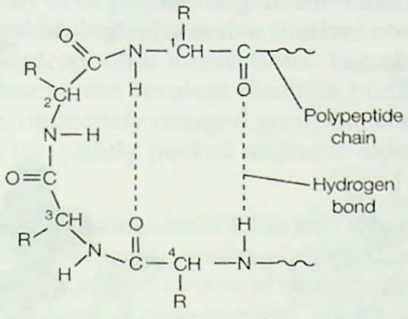

Figure 7. The folding of the polypeptide chain in a β-turn.

the carbonyl oxygen of one amino acid is hydrogen bonded to the hydrogen on the amino group of the fourth amino acid along (Figure 7). β-Turns are often found connecting the ends of antiparallel β-pleated sheets. Regions of the polypeptide chain that are not in a regular secondary structure are said to have a **coil** or **loop conformation**. About half the polypeptide chain of a typical globular protein will be in such a conformation.

Tertiary structure

The third level of structure found in proteins, tertiary structure, refers to the spatial arrangement of amino acids that are far apart in the linear sequence as well as those residues that are adjacent. Again, it is the sequence of amino acids that specifies this final **three-dimensional structure** (Figures 8 and 9). In water-soluble globular proteins such as myoglobin (Section B3), the main driving force behind the folding of the polypeptide chain is the energetic requirement to bury the nonpolar amino acids in the hydrophobic interior away from the surrounding aqueous, hydrophilic medium. The polypeptide chain folds spontaneously so that the majority of its hydrophobic side-chains are buried

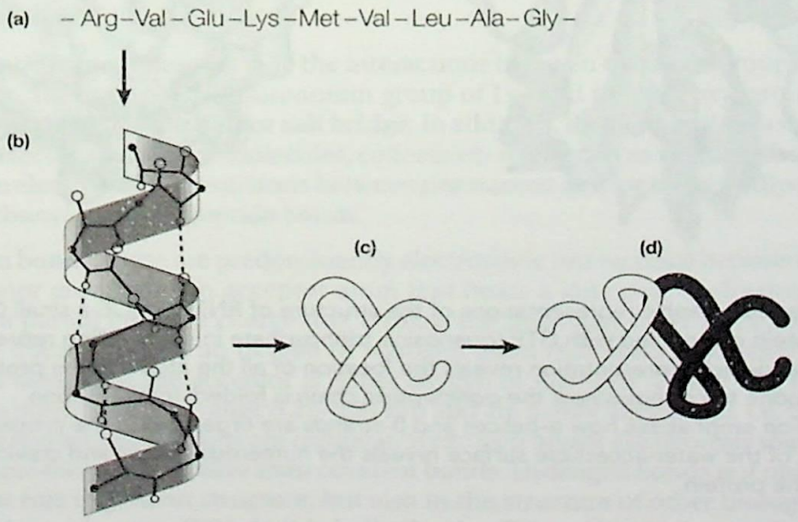

Figure 8. The four levels of structure in proteins: (a) primary structure (amino acid sequence); (b) secondary structure (α-helix); (c) tertiary structure; (d) quaternary structure.

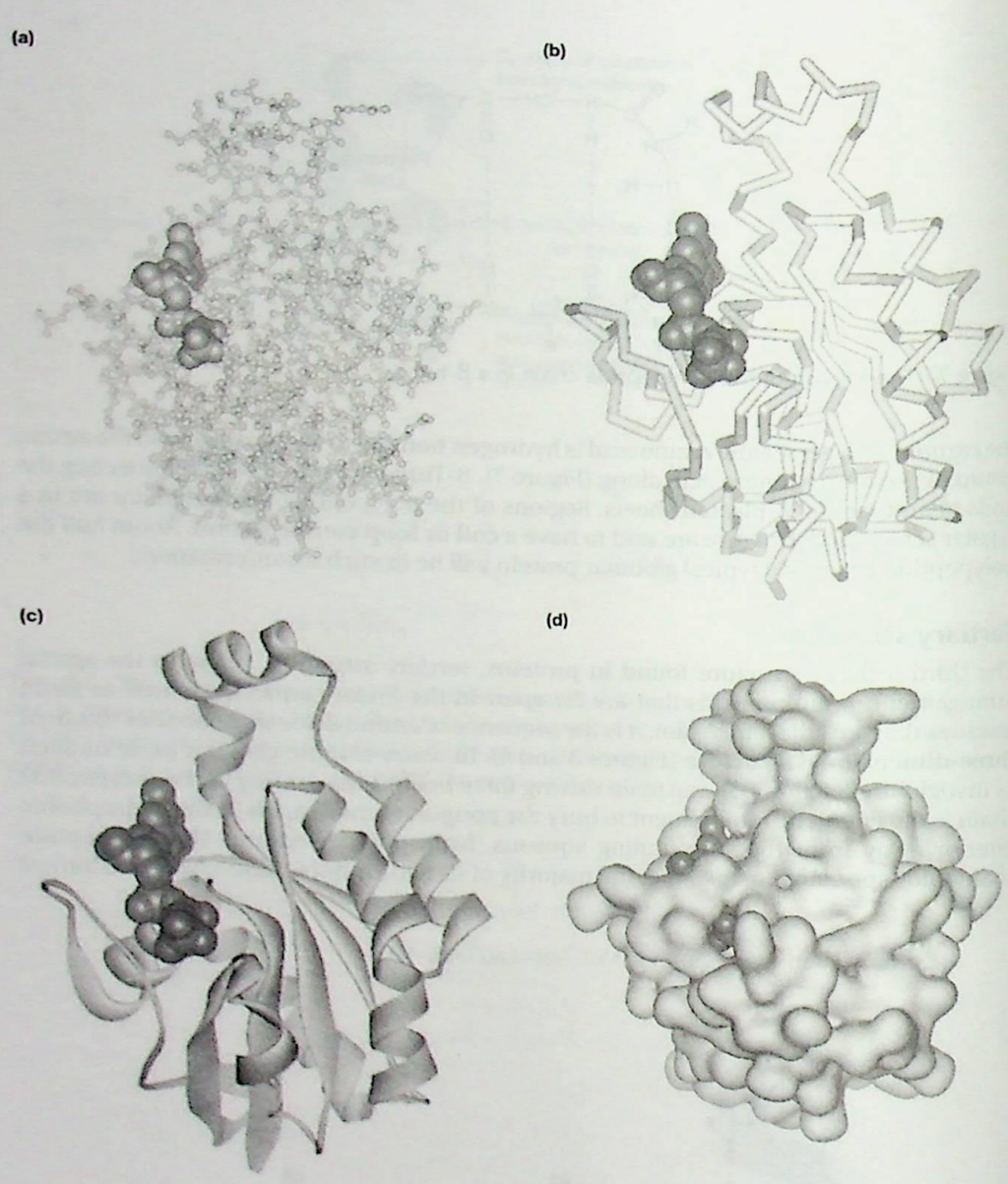

Figure 9. Various graphic representations of the structure of RND3/RHOE a small GTP-binding protein complexed with GTP (guanosine triphosphate in space-filling representation): (a) the ball-and-stick representation reveals the location of all the atoms in the protein; (b) C_α backbone trace shows how the polypeptide chain is folded; (c) the ribbon representation emphasizes how α-helices and β-strands are organized in the protein; (d) a model of the water-accessible surface reveals the numerous bumps and crevices on the surface of the protein.

in the interior, and the majority of its polar, charged side-chains are on the surface. Once folded, the **three-dimensional biologically active (native) conformation** of the protein is maintained not only by hydrophobic interactions, but also by electrostatic forces, hydrogen bonding and, if present, the covalent disulfide bonds. The electrostatic forces include salt bridges between oppositely charged groups and the multiple weak van der Waals interactions between the tightly packed aliphatic side-chains in the interior of the protein.

In many larger proteins, the polypeptide chain folds into two or more compact, globular regions, called **domains**. Most domains consist of 30–400 amino acids in which the polypeptide chain folds into a specific structure with regions of relatively unfolded peptide linking domains together. Consequently, many domains are structurally independent units that have the characteristics of small globular proteins. Somewhat diverse proteins can have domains in common even if their overall tertiary structures are different. Examples of domains in proteins include the **immunoglobulin fold** that is found in antibodies and other proteins of the immune system (Section B6), and the **helix-turn-helix**, **zinc finger** and **basic domains** found in DNA-binding proteins (Section G6).

Quaternary structure

Proteins containing more than one polypeptide chain, such as hemoglobin (Section B3), exhibit a fourth level of protein structure called **quaternary structure** (Figure 8). This level of structure refers to the spatial arrangement of the polypeptide **subunits** and the nature of the interactions between them. These interactions may be covalent links (e.g. disulfide bonds) or noncovalent interactions (electrostatic forces, hydrogen bonding, hydrophobic interactions).

Protein stability

The native three-dimensional conformation of a protein is maintained by a range of noncovalent interactions (electrostatic forces, hydrogen bonds, hydrophobic forces) and covalent interactions (disulfide bonds) in addition to the peptide bonds between individual amino acids.

- **Electrostatic forces**: these include the interactions between two ionic groups of opposite charge, for example the ammonium group of Lys and the carboxyl group of Asp, often referred to as an **ion pair** or **salt bridge**. In addition, the noncovalent associations between electrically neutral molecules, collectively referred to as **van der Waals forces**, arise from electrostatic interactions between permanent and/or induced dipoles, such as the carbonyl group in peptide bonds.
- **Hydrogen bonds**: these are predominantly electrostatic interactions between a weakly acidic donor group and an acceptor atom that bears a lone pair of electrons, which thus has a partial negative charge that attracts the hydrogen atom. In biological systems, the donor group is an oxygen or nitrogen atom that has a covalently attached hydrogen atom, and the acceptor is either oxygen or nitrogen (Figure 10). Hydrogen bonds are normally in the range 0.27–0.31 nm and are highly directional, i.e. the donor, hydrogen and acceptor atoms are collinear. Hydrogen bonds are stronger than van der Waals forces but much weaker than covalent bonds. Hydrogen bonds not only play an important role in protein structure, but also in the structure of other biological macromolecules such as the DNA double helix (Section F1) and lipid bilayers (Section E1). In addition, hydrogen bonds are critical to both the properties of water and to its role as a biochemical solvent.

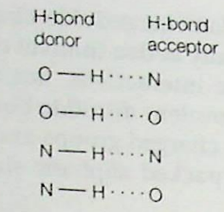

Figure 10. Examples of hydrogen bonds (shown as dotted lines).

- **Hydrophobic forces**: the **hydrophobic effect** is the name given to those forces that cause nonpolar molecules to minimize their contact with water. This is clearly seen with amphipathic molecules such as lipids and detergents that form micelles in aqueous solution (Section E1). Proteins, too, find a conformation in which their nonpolar side-chains are largely out of contact with the aqueous solvent, and thus hydrophobic forces are an important determinant of protein structure, folding and stability. In proteins, the effects of hydrophobic forces are often termed **hydrophobic bonding**, to indicate the specific nature of protein folding under the influence of the hydrophobic effect.

- **Disulfide bonds**: these covalent bonds form between Cys residues that are close together in the final conformation of the protein (Figure 4) and function to stabilize its three-dimensional structure. Disulfide bonds are really only formed in the oxidizing environment of the endoplasmic reticulum (Section A2), and thus are found primarily in extracellular and secreted proteins.

Protein structure determination

The presence of α-helices and β-pleated sheets in proteins can often be predicted from the primary amino acid sequence. However, it is not possible to predict the precise three-dimensional structure of a protein from its amino acid sequence, unless its sequence is very similar to that of a protein whose three-dimensional structure is already known. Sophisticated physical methods and complex analyses of the experimental data are required to determine the conformation of a protein. The three-dimensional structure of a protein can be determined to the atomic level by the techniques of X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryoelectron microscopy.

In **X-ray crystallography** the first requirement is **crystals** of highly purified protein. In the crystal, millions of protein molecules are precisely aligned with one another in a rigid array characteristic of that particular protein. **Beams of X-rays** are then passed through the crystal (Figure 11). The wavelengths of X-rays are 0.1–0.2 nm, short enough to resolve the atoms in the protein crystal. The atoms in the crystal scatter the X-rays, producing a **diffraction pattern** of discrete spots on photographic film. The intensities of the diffraction maxima (the darkness of the spots on the film) are then used mathematically to construct the three-dimensional image of the protein crystal.

Nuclear magnetic resonance (NMR) spectroscopy can be used to determine the three-dimensional structures of small (up to approximately 30 kDa) proteins in **aqueous solution**. It does not require the crystallization of the protein. In this technique, a concentrated protein solution is placed in a **magnetic field** and the effects of different radio frequencies on the spin of different atoms in the protein measured. The behavior of any particular atom is influenced by neighboring atoms in adjacent residues, with closer

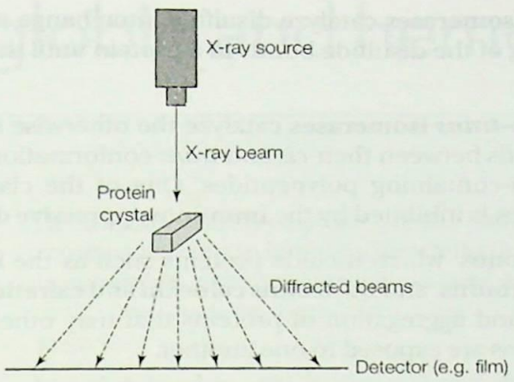

Figure 11. X-Ray crystallography. When a narrow beam of X-rays strikes a crystal, part of it passes straight through and the rest is scattered (diffracted) in various directions. The intensity of the diffracted waves is recorded on photographic film or with a solid-state electronic detector. From the diffraction data, the three-dimensional structure of the protein can be determined.

residues causing more perturbation than distant ones. From the magnitude of the effect, the distances between residues can be calculated and then used to generate the three-dimensional structure of the protein.

Cryoelectron microscopy is often used to determine the three-dimensional structures of proteins, particularly multisubunit proteins, which are difficult to crystallize. In this technique, the protein sample is **rapidly frozen** in liquid helium to preserve its structure. The frozen, hydrated protein is then examined in a cryoelectron microscope using a low dose of **electrons** to prevent radiation-induced damage to the structure. The resulting images are analyzed by complex computer programs and the three-dimensional structure of the protein reconstructed.

Protein folding

Under appropriate physiological conditions, proteins **spontaneously fold** into their native conformation. As there is no need for external templates, this implies that the primary structure of the protein dictates its three-dimensional structure. From experiments with the protein **RNase A**, it has been observed that it is mainly the internal residues of a protein that direct its folding to the native conformation. Alteration of surface residues by mutation is less likely to affect the folding than changes to internal residues. It has also been observed that protein folding is driven primarily by **hydrophobic forces**. Proteins fold into their native conformation through an **ordered set of pathways** rather than by a random exploration of all the possible conformations until the correct one is stumbled upon.

Although proteins can fold *in vitro* (in the laboratory) without the presence of accessory proteins, this process can take minutes to days. *In vivo* (in the cell) this process requires only a few minutes because the cells contain **accessory proteins**, which assist the polypeptides to fold to their native conformation. There are three main classes of protein folding accessory proteins:

- **Protein disulfide isomerases** catalyze disulfide interchange reactions, thereby facil tating the shuffling of the disulfide bonds in a protein until they achieve their corre pairing.
- **Peptidyl prolyl *cis–trans* isomerases** catalyze the otherwise slow interconversion c X–Pro peptide bonds between their *cis* and *trans* conformations, thereby acceleratin the folding of Pro-containing polypeptides. One of the classes of peptidyl prol *cis–trans* isomerases is inhibited by the **immunosuppressive drug** cyclosporin A.
- **Molecular chaperones**, which include proteins such as the **heat shock proteins 7** (hsc70), the **chaperonins**, and the **lectins calnexin** and **calreticulin**. These prevent th improper folding and aggregation of proteins that may otherwise occur as interna hydrophobic regions are exposed to one another.

Some diseases are caused by **protein misfolding**. For example in the fatal neurodegen erative **Alzheimer's disease** the small **amyloid-β peptide** misfolds and aggregates in th brain, killing the surrounding neurons. While in the infectious **prion diseases**, such a Creutzfeldt–Jakob disease (CJD) in humans and bovine spongiform encephalopathy (BS or mad cow disease), the normal cellular form of the **prion protein**, which has mainl α-helical secondary structure, undergoes a conformational transition to the infectiou form of the protein that has predominantly β-sheet structure.