

IMPACT 21 ...ON BIOLOGY: Biological macromolecules

A protein is a polypeptide composed of linked α -amino acids, $\text{NH}_2\text{CHR}\text{COOH}$, where R is one of about 20 groups. For a protein to function correctly, it needs to have the correct conformation. For example, an enzyme has its greatest catalytic efficiency only when it is in a specific conformation. The amino acid sequence of a protein contains the necessary information to create the active conformation of the protein as it is formed. However, the prediction of the observed conformation from the primary structure, the so-called *protein folding problem*, is extraordinarily difficult and is still the focus of much research. The other class of biological macromolecules to be considered are the nucleic acids, which are key components of the mechanism of storage and transfer of genetic information in biological cells. Deoxyribonucleic acid (DNA) contains the instructions for protein synthesis, which is carried out by different forms of ribonucleic acid (RNA).

(a) Proteins

The origin of the secondary structures of proteins is found in the rules formulated by Linus Pauling and Robert Corey in 1951, that seek to identify the principal contributions to the lowering of energy of the molecule by focusing on the role of hydrogen bonds and the peptide link, $-\text{CONH}-$. The latter can act both as a donor of the H atom (the NH part of the link) and as an acceptor (the CO part). The **Corey–Pauling rules** are as follows (Fig. 1):

1. The four atoms of the peptide link lie in a nearly rigid plane.

The planarity of the link is due to delocalization of π electrons over the O, C, and N atoms and the maintenance of maximum overlap of their p orbitals.

2. The N, H, and O atoms of a hydrogen bond lie in a straight line (with displacements of H tolerated up to not more than 30° from the N–O direction).

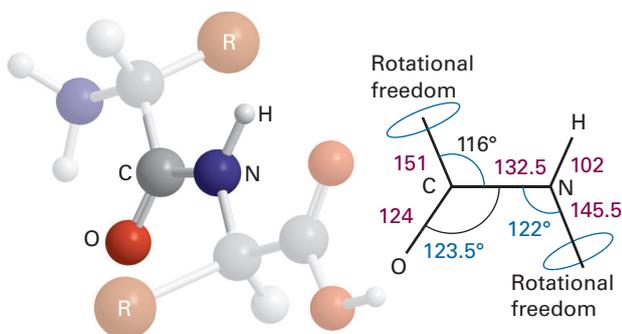


Figure 1 The dimensions that characterise the peptide link. The C–NH–CO–C atoms define a plane (the C–N bond has partial double bond character), but there is rotational freedom around the C–CO and N–C bonds.

3. All NH and CO groups of the peptide links take part in hydrogen bonding.

The rules are satisfied by two structures. One, in which hydrogen bonding between peptide links leads to a helical structure, is a *helix*, which can be arranged as either a right- or a left-handed screw. The other, in which hydrogen bonding between peptide links leads to a planar structure, is a *sheet*; this form is the secondary structure of the protein fibroin, the constituent of silk.

Because the planar peptide link is almost rigid, the geometry of a polypeptide chain can be specified by the two angles that two neighbouring planar peptide links make to each other. Figure 2 shows the two angles ϕ and ψ commonly used to specify this relative orientation. The sign convention is that a positive angle means that the front atom must be rotated clockwise to bring it into an eclipsed position relative to the rear atom. For an all-*trans* form of the chain, all ϕ and ψ are 180° . A helix is obtained when all the ϕ are equal and when all the ψ are equal. For a right-handed helix, an α -helix (Fig. 3), all $\phi \approx -57^\circ$ and all

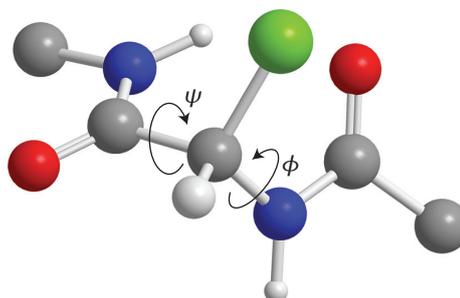


Figure 2 The definition of the torsional angles ψ and ϕ between two peptide units. In this case (an α -L-polypeptide) the chain has been drawn in its all trans form, with $\psi = \phi = 180^\circ$.

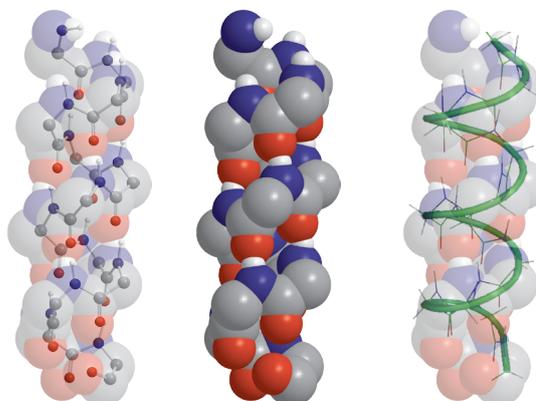


Figure 3 The polypeptide α -helix, with poly-L-glycine as an example. There are 3.6 residues per turn, and a translation along the helix of 150 pm per residue, giving a pitch of 540 pm. The diameter (ignoring side chains) is about 600 pm.

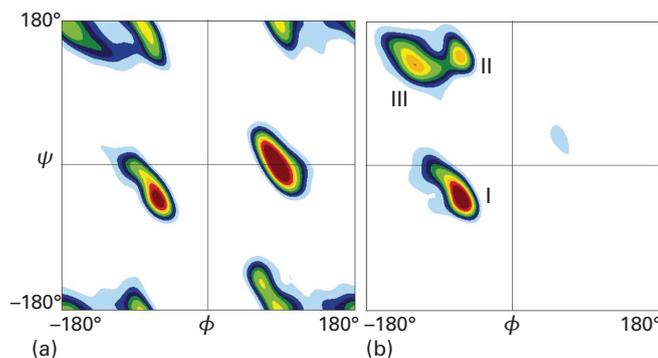


Figure 4 Contour plots of potential energy against the angles ψ and ϕ , known as a Ramachandran plot, for (a) a glycyl residue of a polypeptide chain and (b) an alanyl residue. The glycyl plot is symmetrical, but that for alanyl is unsymmetrical and the global minimum corresponds to an α -helix. (T. Hovmoller, et al., *Acta Cryst.* **D58**, 768 (2002).)

$\psi \approx -47^\circ$. For a left-handed helix, both angles are positive. The torsional contribution to the total potential energy is

$$V_{\text{torsion}} = A(1 + \cos 3\phi) + B(1 + \cos 3\psi) \quad (1)$$

in which A and B are constants of the order of 1 kJ mol^{-1} . Because only two angles are needed to specify the conformation of a helix, and they range from -180° to $+180^\circ$, the torsional potential energy of the entire molecule can be represented on a **Ramachandran plot**, a contour diagram in which one axis represents ϕ and the other represents ψ .

Figure 4 shows the Ramachandran plots for the helical form of polypeptide chains formed from the nonchiral amino acid glycine ($R = H$) and the chiral amino acid L-alanine ($R = \text{CH}_3$). The glycine map is symmetrical, with troughs of equal depth, shown as the red regions in Fig. 4a. In contrast, the map for L-alanine is unsymmetrical, and there is a pronounced minimum (labelled I) at angles consistent with a right-handed helix. This result is consistent with the observation that polypeptides of the naturally occurring L-amino acids tend to form right-handed helices.

A β -sheet (also called the *β -pleated sheet*) is formed by hydrogen bonding between two extended polypeptide chains (large absolute values of the torsion angles ϕ and ψ). In an **anti-parallel β -sheet** (Fig. 5a), $\phi \approx 139^\circ$, $\psi \approx 113^\circ$, and the N-H \cdots O atoms of the hydrogen bonds form a straight line. This arrangement is a consequence of the antiparallel arrangement of the chains: every N-H bond on one chain is aligned with a C=O bond from another chain. Antiparallel β -sheets are very common in proteins. In a **parallel β -sheet** (Fig. 5b), $\phi \approx 119^\circ$, $\psi \approx 113^\circ$, and the N-H \cdots O atoms of the hydrogen bonds are not perfectly aligned. This arrangement is a result of the parallel arrangement of the chains: each N-H bond on one chain is aligned with a N-H bond of another chain and, as a result, each C=O bond of one chain is aligned with a C=O bond of another chain. These structures are not common in proteins.

Covalent and non-covalent interactions may cause polypeptide chains with well-defined secondary struc-

tures to fold into tertiary structures. Although the rules that govern protein folding are still being elucidated, a few general conclusions may be drawn from X-ray diffraction studies of water-soluble natural proteins and synthetic polypeptides. In an aqueous environment, the chains fold in such a way as to place nonpolar R groups in the interior (which is often not very accessible to solvent) and charged R groups on the surface (in direct contact with the polar solvent). Other factors that promote the folding of proteins include covalent disulfide ($-\text{S}-\text{S}-$) links, Coulombic interactions between charged groups (which depend on their degree of protonation and therefore on the pH), van der Waals interactions, and hydrophobic interactions (Topics 14B and 14E). The clustering of nonpolar, hydrophobic, amino acid residues into the interior of a protein is driven primarily by hydrophobic interactions.

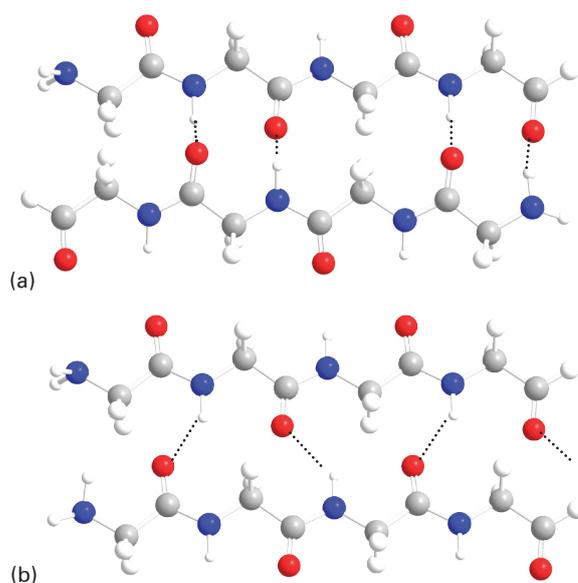


Figure 5 The two types of β -sheets: (a) antiparallel, (b) parallel. The two strands are in a better alignment to form linking hydrogen bonds in the antiparallel arrangement than in the parallel arrangement.

(b) Nucleic acids

Both DNA and RNA are *polynucleotides* (1), in which base-sugar-phosphate units are linked by phosphodiester bonds. In RNA the sugar is β -D-ribose and in DNA it is β -D-2'-deoxyribose (as shown in 1). The most common bases are adenine (A, 2), cytosine (C, 3), guanine (G, 4), thymine (T, found in DNA only, 5), and uracil (U, found in RNA only, 6). At physiological pH, each phosphate group of the chain carries a negative charge and the bases are deprotonated and neutral. This charge distribution leads to two important properties. One is that the polynucleotide chain is a **polyelectrolyte**, a macromolecule with many different charged sites, with a large and negative overall surface charge. The second is that the bases can interact by hydrogen bonding, as shown for A-T (7) and C-G base pairs (8). The secondary and tertiary structures of DNA and RNA arise primarily from the pattern of this hydrogen bonding between bases of one or more chains.

In DNA, two polynucleotide chains wind around each other to form a double helix (Fig. 6). The chains are held together by links involving A-T and C-G base pairs

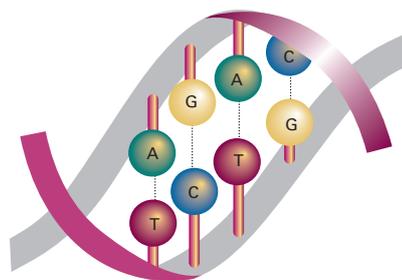


Figure 6 DNA double helix, in which two polynucleotide chains are linked together by hydrogen bonds between adenine (A) and thymine (T) and between cytosine (C) and guanine (G).

that lie parallel to each other and perpendicular to the axis of the helix. The structure is stabilized further by interactions between the planar π systems of the bases. In B-DNA, the most common form of DNA found in biological cells, the helix is right-handed with a diameter of 2.0 nm and a pitch of 3.4 nm.

