1

The Building Blocks

It is impossible to pack a complete biochemistry course into a single introductory chapter. Some of the basic properties of the structure of simple biological macromolecules, lipids and micro organisms are covered. The aim is to give a basic grounding in the rich variety of molecules that life presents, and some respect for the extreme complexity of the chemistry of biological molecules that operates in a wide range of cellular processes.

1.1 PROTEINS

Polymers consist of a large number of sub-units (monomers) connected together with covalent bonds. A protein is a special type of polymer. In a protein there are up to twenty different amino acids (Figure 1.1) that can function as monomers, and all the monomers are connected together with identical peptide linkages (C–N bonds, Figure 1.2). The twenty amino acids can be placed in different families dependent on the chemistry of their different side groups. Five of the amino acids form a group with lipophilic (fat-liking) side-chains: glycine, alanine, valine, leucine, and isoleucine. Proline is a unique circular amino acid that is given its own separate classification. There are three amino acids with aromatic side-chains: phenylalanine, tryptophan, and tyrosine. Sulfur is in the side-chains of two amino acids: cysteine and methionine. Two amino acids have hydroxyl (neutral) groups that make them water loving: serine and threonine. Three amino acids have very polar positive side-chains: lysine, arginine and histidine. Two amino acids form a family with acidic

Applied Biophysics: A Molecular Approach for Physical Scientists Tom A. Waigh

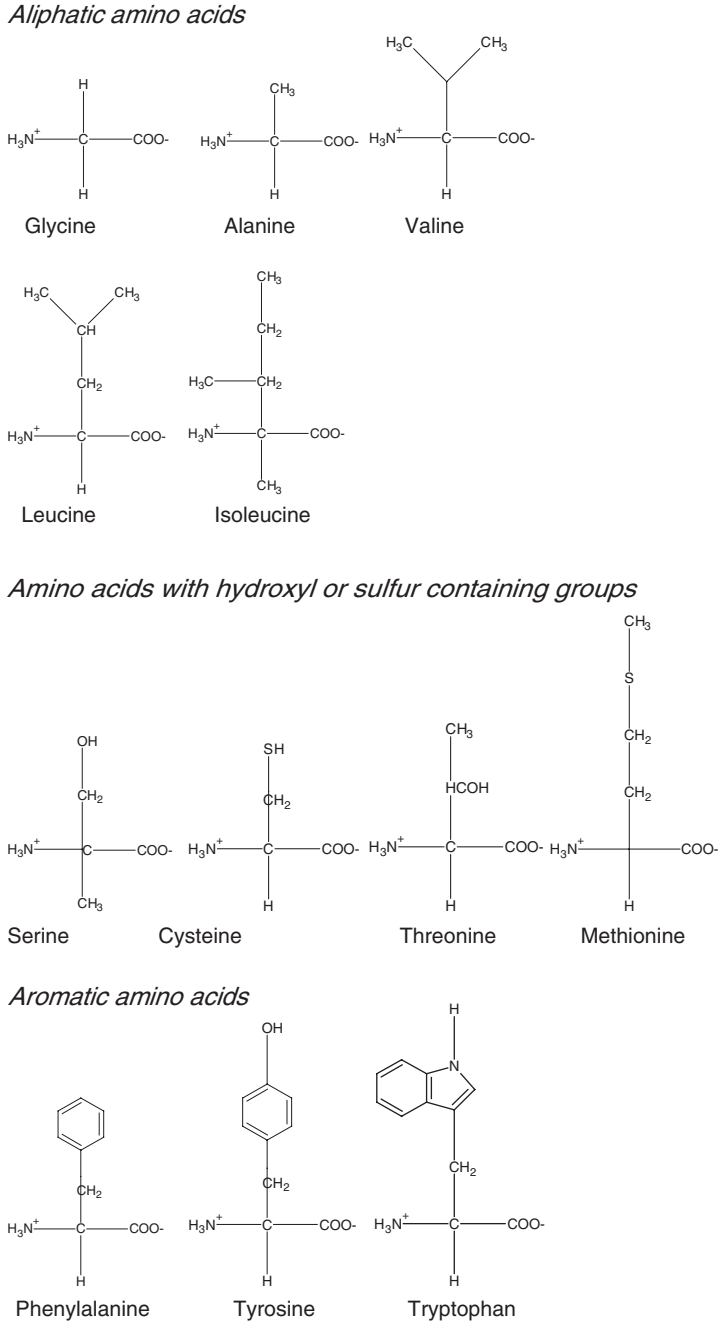
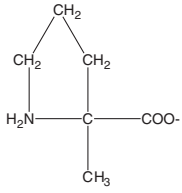


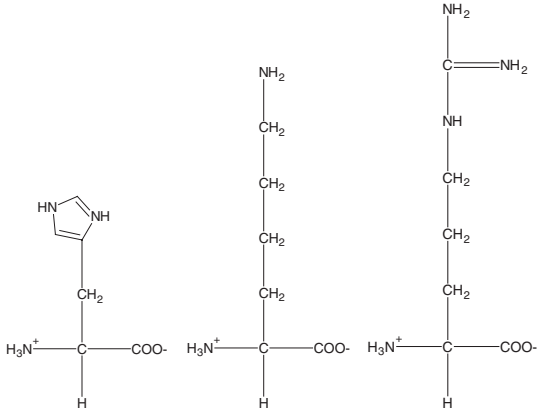
Figure 1.1 The chemical structure of the twenty amino acids found in nature

Cyclic amino acid



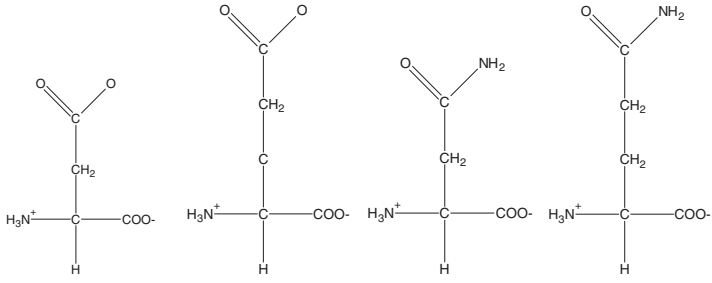
Proline

Basic amino acids



Histidine Lysine Arginine

Acidic amino acids and amides



Aspartic acid Glutamic acid Asparagine Glutamine

Figure 1.1 (*Continued*)

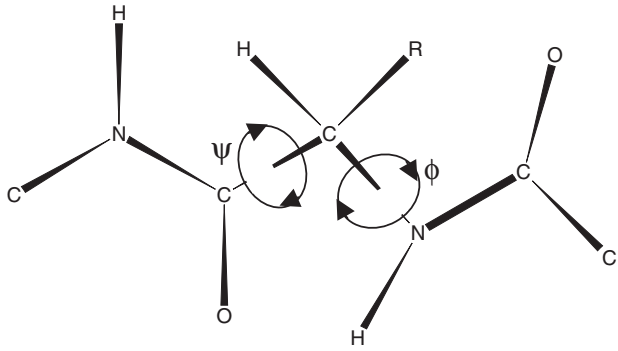


Figure 1.2 All amino acids have the same primitive structure and are connected with the same peptide linkage through C–C–N bonds
(O, N, C, H indicate oxygen, nitrogen, carbon and hydrogen atoms respectively. R is a pendant side-group which provides the amino acid with its identity, i.e. proline, glycine etc.)

side-groups and they are joined by two corresponding neutral counterparts that have a similar chemistry: aspartate, glutamate, asparagine, and glutamine.

The linkages between amino acids all have the same chemistry and basic geometry (Figure 1.2). The *peptide linkage* that connects all amino acids together consists of a carbon atom attached to a nitrogen atom through a single covalent bond. Although the chemistry of peptide linkages is fairly simple, to relate the primary sequence of amino acids to the resultant three dimensional structure in a protein is a daunting task and predominantly remains an unsolved problem. To describe protein structure in more detail it is useful to consider the motifs of secondary structure that occur in their morphology. The motifs include *alpha helices*, *beta sheets* and *beta barrels* (Figure 1.3). The full three dimensional *tertiary structure* of a protein typically takes the form of a compact globular morphology (the globular proteins) or a long extended conformation (fibrous proteins, Figures 1.4 and 1.5). Globular morphologies usually consist of a number of secondary motifs combined with more disordered regions of peptide.

Charge interactions are very important in determining of the conformation of biological polymers. The degree of charge on a polyacid or polybase (e.g. proteins, nucleic acids etc) is determined by the pH of a solution, i.e. the concentration of hydrogen ions. Water has the ability to dissociate into oppositely charged ions; this process depends on temperature

$$H_2O \rightleftarrows H^+ + OH^- \qquad (1.1)$$

The product of the hydrogen and hydroxyl ion concentrations formed from the dissociation of water is a constant at equilibrium and at a fixed temperature (37 °C)

$$c_{H^+}c_{OH^-} = 1 \times 10^{-14}M^2 = K_w \tag{1.2}$$

where c_{H^+} and c_{OH^-} are the concentrations of hydrogen and hydroxyl ions respectively. Addition of acids and bases to a solution perturbs the equilibrium dissociation process of water, and the acid/base equilibrium

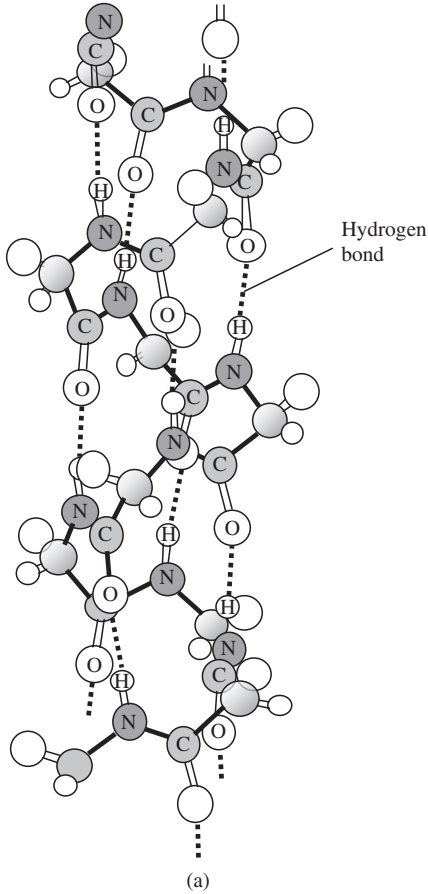
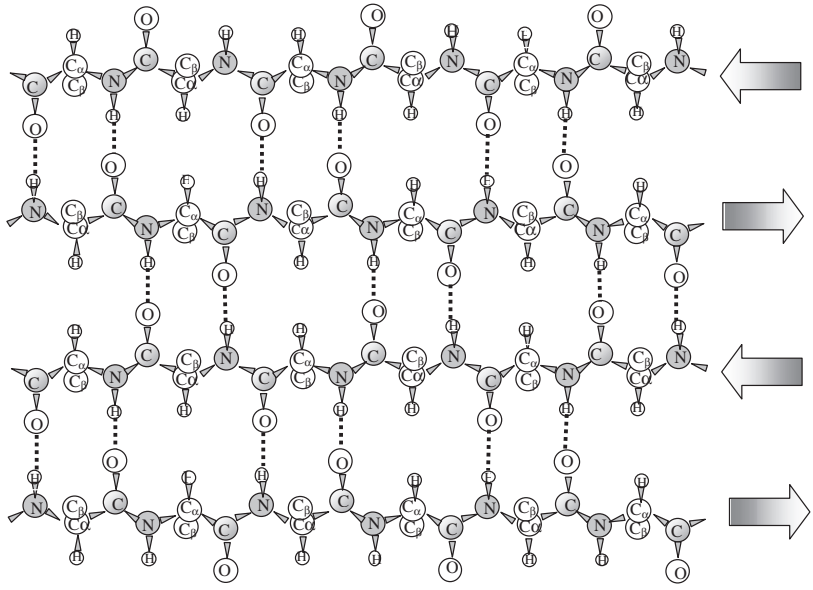


Figure 1.3 Simplified secondary structures of (a) an α-helix and (b) a β-sheet that commonly occur in proteins
(Hydrogen bonds are indicated by dotted lines.)



(b)

Figure 1.3 (*Continued*)

phenomena involved are a corner stone of the physical chemistry of solutions. Due to the vast range of possible hydrogen ion (H^+) concentrations typically encountered in aqueous solutions, it is normal to use a logarithmic scale (pH) to quantify them. The pH is defined as the

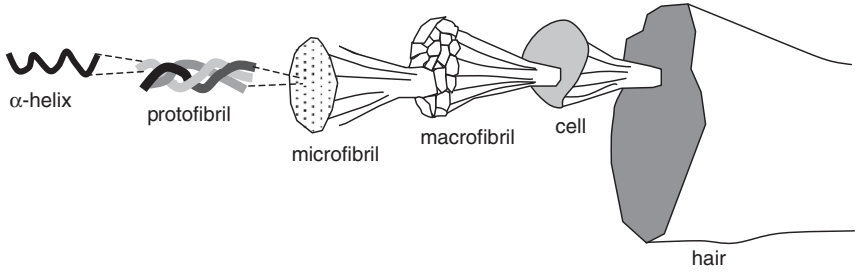


Figure 1.4 The complex hierarchical structures found in the keratins of hair (α-helices are combined in to protofibrils, then into microfibrils, macrofibrils, cells and finally in to a single hair fibre [*Reprinted with permission from J.Vincent, Structural Biomaterial, Copyright (1990) Princeton University Press*])

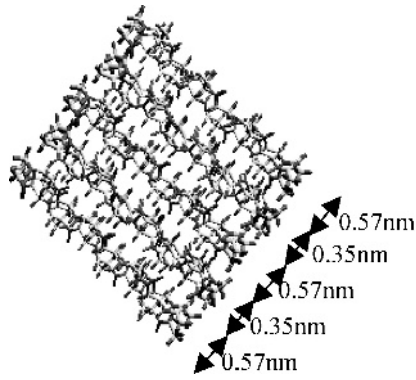


Figure 1.5 The packing of anti-parallel beta sheets found in silk proteins (Distances between the adjacent sheets are shown.)

negative logarithm (base 10!) of the hydrogen ion concentration

$$\mathrm{pH} = -\log c_{\mathrm{H}^+} \tag{1.3}$$

Typical values of pH range from 6.5 to 8 in physiological cellular conditions. Strong acids have a pH in the range 1–2 and strong bases have a pH in the range 12–13.

When an acid (HA) dissociates in solution it is possible to define an equilibrium constant ($\mathrm{K_a}$) for the dissociation of its hydrogen ions ($\mathrm{H^+}$)

$$\mathrm{HA} \rightleftarrows \mathrm{H^+} + \mathrm{A^-} \qquad \mathrm{K}_a = \frac{c_{\mathrm{H}^+} c_{\mathrm{A}^-}}{c_{\mathrm{HA}}} \tag{1.4}$$

where c_{H^+}, c_{A^-} and $\mathrm{c_{HA}}$ are the concentrations of the hydrogen ions, acid ions, and acid molecules respectively. Since the hydrogen ion concentration follows a logarithmic scale, it is natural to also define the dissociation constant on a logarithmic scale (pK_a)

$$\mathrm{pK}_a = -\log \mathrm{K}_a \tag{1.5}$$

The logarithm of both sides of equation (1.4) can be taken to give a relationship between the pH and the $\mathrm{pK_a}$ value:

$$\mathrm{pH} = \mathrm{pK}_a + \log\left\{\frac{c_{\mathrm{conjugate\_base}}}{c_{\mathrm{acid}}}\right\} \tag{1.6}$$

where $c_{conjugate\_base}$ and c_{acid} are the concentrations of the conjugate base (e.g. A^-) and acid (e.g. HA) respectively. This equation enables the degree of dissociation of an acid (or base) to be calculated, and it is named after its inventors *Henderson and Hasselbalch*. Thus a knowledge of the pH of a solution and the pK_a value of an acidic or basic group allows the charge fraction on the molecular group to be calculated to a first approximation. The propensity of the amino acids to dissociate in water is illustrated in Table 1.1. In contradiction to what their name might imply, only amino acids with acidic or basic side groups are charged when incorporated into proteins. These charged amino acids are arginine, aspartic acid, cysteine, glutamic acid, histidine, lysine and tyrosine.

Another important interaction between amino acids, in addition to charge interactions, is their ability to form hydrogen bonds with surrounding water molecules; the degree to which this occurs varies. This amino acid hydrophobicity (the amount they dislike water) is an important driving force for the conformation of proteins. Crucially it leads to the compact conformation of globular proteins (most enzymes) as the hydrophobic groups are buried in the centre of the globules to avoid contact with the surrounding water.

Table 1.1 Fundamental physical properties of amino acids found in protein [*Ref.: Data adapted from C.K. Mathews and K.E. Van Holde, Biochemistry, 137].*

Name	pK_a value of side chain	Mass of residue	Occurrence in natural proteins (%mol)
Alanine	—	71	9.0
Arginine	12.5	156	4.7
Asparagine	—	114	4.4
Apartic acid	3.9	115	5.5
Cysteine	8.3	103	2.8
Glutamine	—	128	3.9
Glutamic acid	4.2	129	6.2
Glycine	—	57	7.5
Histidine	6.0	137	2.1
Isoleucine	—	113	4.6
Leucine	—	113	7.5
Lysine	10.0	128	7.0
Methionine	—	131	1.7
Phenylalanine	—	147	3.5
Proline	—	97	4.6
Serine	—	87	7.1
Threonine	—	101	6.0
Tyrptophan	—	186	1.1
Tyrosine	10.1	163	3.5
Valine	—	99	6.9

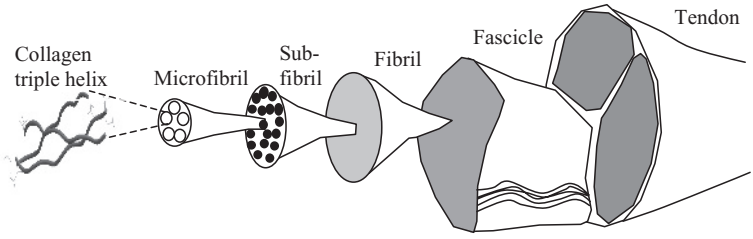


Figure 1.6 Hierarchical structure for the collagen triple helices in tendons (Collagen helices are combined into microfibrils, then into sub-fibrils, fibrils, fascicles and finally into tendons.)

Covalent interactions are possible between adjacent amino acids and can produce solid protein aggregates (Figures 1.4 and 1.6). For example, disulfide linkages are possible in proteins that contain cysteine, and these form the strong inter-protein linkages found in many fibrous proteins e.g. keratins in hair.

The internal secondary structures of protein chains (α helices and β sheets) are stabilised by hydrogen bonds between adjacent atoms in the peptide groups along the main chain. The important structural proteins such as keratins (Figure 1.4), collagens (Figure 1.6), silks (Figure 1.5), anthropod cuticle matrices, elastins (Figure 1.7), resilin

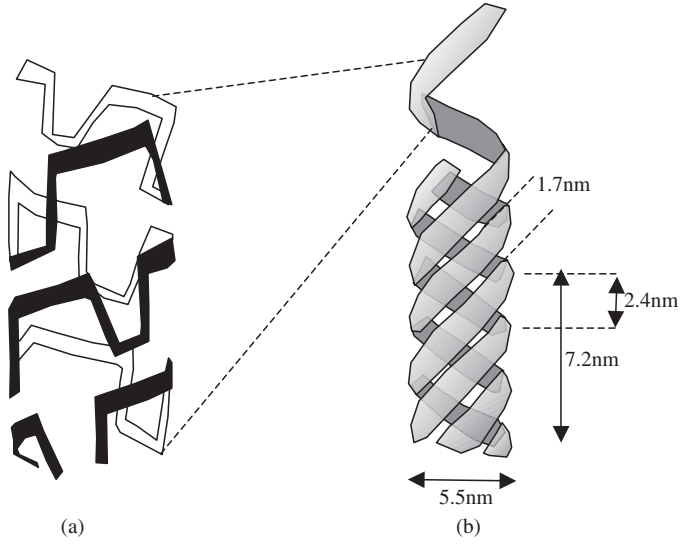


Figure 1.7 The β turns in elastin (a) form a secondary elastic helix which is subsequently assembled into a superhelical fibrous structure (b)

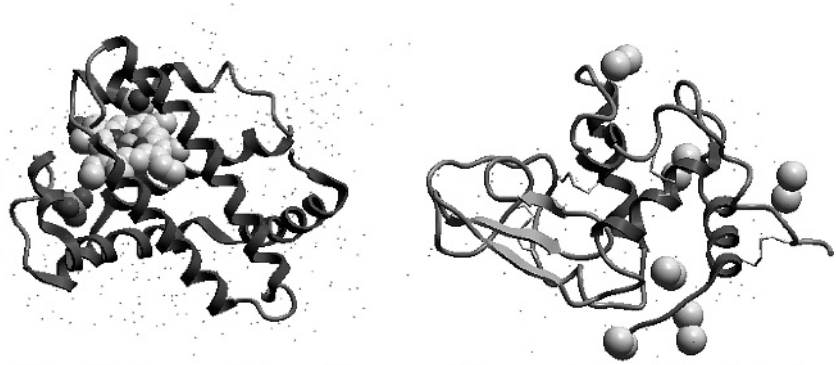


Figure 1.8 Two typical structures of globular proteins calculated using X-ray crystallography data

and abductin are formed from a combination of intermolecular disulfide and hydrogen bonds.

Some examples of the globular structures adopted by proteins are shown in Figure 1.8. Globular proteins can be denatured in a folding/unfolding transition through a number of mechanisms, e.g. an increase in the temperature, a change of pH, and the introduction of hydrogen bond breaking chaotropic solvents. Typically the complete denaturation transition is a first order thermodynamic phase change with an associated latent heat (the thermal energy absorbed during the transition). The unfolding process involves an extremely complex sequence of molecular origami transitions. There are a vast number of possible molecular configurations ($\sim 10^{N}$ for an N residue protein) that occur in the reverse process of protein folding, when the globular protein is constructed from its primary sequence by the cell, and thus frustrated structures could easily be formed during this process. Indeed, at first sight it appears a certainty that protein molecules will become trapped in an intermediate state and never reach their correctly folded form. This is called *Levinthal's paradox*, the process by which natural globular proteins manage to find their native state among the billions of possibilities in a finite time. The current explanation of protein folding that provides a resolution to this paradox, is that there is a funnel of energy states that guide the kinetics of folding across the complex energy landscape to the perfectly folded state (Figure 1.9).

There are two main types of inter-chain interaction between different proteins in solution; those in which the native state remains largely

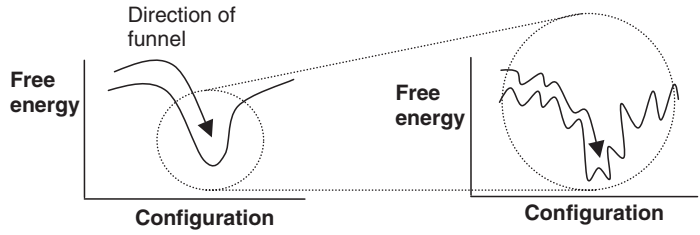


Figure 1.9 Schematic diagram indicating the funnel that guides the process of protein folding through the complex configuration space that contains many local minima. The funnel avoids the frustrated misfolded protein structures described in Levinthal's paradox

unperturbed in processes such as protein crystallisation and the formation of filaments in sheets and tapes, and those interactions that lead to a loss of conformation e.g. heat set gels (e.g. table jelly and boiled eggs) and amyloid fibres (e.g. Alzheimer's disease and Bovine Spongiform Encephalopathy).

1.2 LIPIDS

Cells are divided into a series of subsections or compartments by membranes which are formed predominantly from lipids. The other main role of lipids is as energy storage compounds, although the molecules play a role in countless other physiological processes. Lipids are amphiphilic, the head groups like water (and hate fat) and the tails like fat (and hate water). This amphiphilicity drives the spontaneous self-assembly of the molecules into membranous morphologies.

There are four principle families of lipids: fatty acids with one or two tails (including carboxylic acids of the form RCOOH where R is a long hydrocarbon chain), and steroids and phospholipids where two fatty acids are linked to a glycerol backbone (Figure 1.10). The type of polar head group differentiates the particular species of naturally occurring lipid. Cholesterol is a member of the steroid family and these compounds are often found in membrane structures. Glycolipids also occur in membranes and in these molecules the phosphate group on a phospholipid is replaced by a sugar residue. Glycolipids have important roles in cell signalling and the immune system. For example, these molecules are an important factor in determining the compatibility of blood cells after a blood transfusion, i.e. blood types A, B, O, etc.

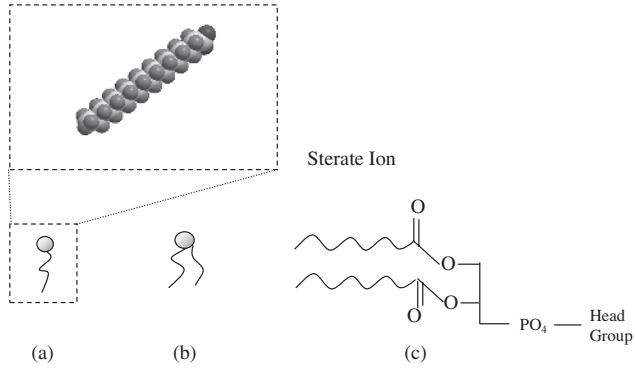


Figure 1.10 Range of lipid molecules typically encountered in biology
(a) fatty acids with one tail; (b) steroids and fatty acids with two tails; (c) phospholipids

1.3 NUCLEIC ACIDS

The '*central dogma of biochemistry*' according to F.C.Crick is illustrated in Figure 1.11. DNA contains the basic blueprint for life that guides the construction of the vast majority of living organisms. To implement this blue print cells need to *transcribe* DNA to RNA, and this structural information is subsequently translated into proteins using specialised protein factories (the ribosomes). The resultant proteins can then be used to catalyse specific chemical reactions or be used as building materials to construct new cells.

This simple biochemical scheme for transferring information has powerful implications. DNA can now be altered systematically using *recombinant DNA technology* and then placed inside a living cell. The foreign DNA hijacks the cell's mechanisms for translation and the proteins that are subsequently formed can be tailor-made by the genetic engineer to fulfil a specific function, e.g. bacteria can be used to form biodegradable plastics from the fibrous proteins that are expressed.

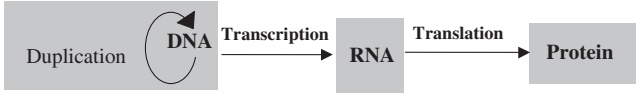


Figure 1.11 The central dogma of molecular biology considers the duplication and translation of DNA. DNA is duplicated from a DNA template. DNA is transcribed to form a RNA chain, and this information is translated into a protein sequence

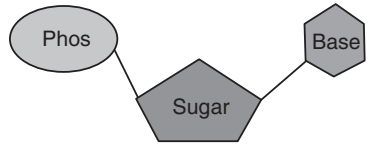


Figure 1.12 The chemical structure of the base of a nucleic acid consists of a phosphate group, a sugar and a base

The monomers of DNA are made of a sugar, an organic base and a phosphate group (Figure 1.12). There are only four organic bases that naturally occur in DNA, and these are thymine, cytosine, adenine and guanine (T,C,A,G). The sequence of bases in each strand along the backbone contains the genetic code. The base pairs in each strand of the double helical DNA are complementary, A has an afinity for T (they form two hydrogen bonds) and G for C (they form three hydrogen bonds). The interaction between the base pairs is driven by the geometry of the hydrogen bonding sites. Thus each strand of the DNA helix contains an identical copy of the genetic information to its complementary strand, and replication can occur by separation of the double helix and resynthesis of two additional chains on each of the two original double helical strands. The formation of helical secondary structures in DNA drastically increases the persistence length of each separate chain and is called a *helix-coil transition*.

There is a major groove and a minor groove on the biologically active A and B forms of the DNA double helix. The individual polynucleotide DNA chains have a sense of direction, in addition to their individuality (a complex nucleotide sequence). DNA replication in vivo is conducted by a combination of the DNA polymerases (I, II and III).

DNA in its double helical form can store torsional energy, since the monomers are not free to rotate (like a telephone cable). The ends of a DNA molecule can be joined together to form a compact supercoiled structure that often occurs in vivo in bacteria; this type of molecule presents a series of fascinating questions with regard to its statistical mechanics and topological analysis.

DNA has a wide variety of structural possibilities (Table 1.2, Figure 1.13). There are *3 standard types* of averaged double helical structure labelled A, B and Z, which occur ex vivo in the solid fibres used for X-ray structural determination. Typically DNA in solution has a structure that is intermediate between A and B, dependent on the chain sequence and the aqueous environment. An increase in the level of hydration tends to increase the number of B type base pairs in a double

Table 1.2 Structural parameters of polynucleotide helices

Property	A form	B form	Z-form
Direction of helix rotation	Right	Right	Left
Number of residues per turn	11	10	12
Rotation per residue	33°	36°	30°
Rise in helix per residue	0.255 nm	0.34 nm	0.37 nm
Pitch of helix	2.8 nm	3.4 nm	4.5 nm

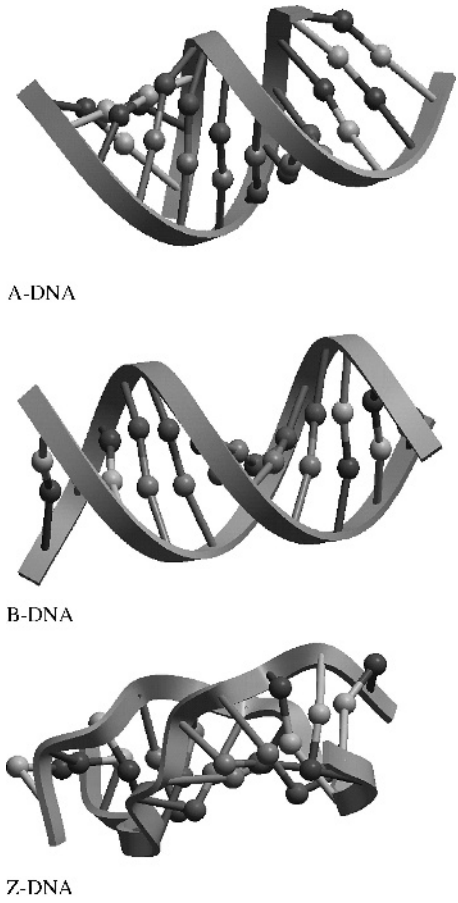


Figure 1.13 Molecular models of A, B and Z type double helical structures of DNA (A and B type helical structures, and their intermediates typically occur in biological systems. Z-DNA helical structures crystallise under extreme non-physiological conditions.)

helix. Z-type DNA is favoured in some extreme non-physiological conditions.

There are a number of local structural modifications to the helical structure that are dependent on the specific chemistry of the individual DNA strands, and are in addition to the globally averaged A, B and Z classifications. The *kink* is a sudden bend in the axis of the double helix which is important for complexation in the nucleosome. The *loop* contains a rupture of hydrogen bonds over several base pairs, and the separation of two nucleotide chains produces loops of various sizes. In the process of DNA transcription RNA polymerase is bound to DNA to form a loop structure. In the process of *breathing* of a double helix, hydrogen bonds are temporarily broken by a rapid partial rotation of one base pair. The hydrogen atoms in the NH groups are therefore accessible and can be exchanged with neighbouring protons in the presence of a catalyst. The *cruciform structure* is formed in the presence of self-complementary palindromic sequences separated by several base pairs. Hydrophobic molecules (e.g. DNA active drugs) can be *intercalated* into the DNA structure, i.e. slipped between two base pairs. Helices that contain three or four nucleic acid strands are also possible with DNA, but do not occur naturally.

DNA has a number of interesting features with respect to its polymer physics. The persistence length (l_p) of DNA is in the order of 50 nm for *E. coli* (which depends on ionic strength), it can have millions of monomers in its sequence and a correspondingly gigantic contour length (L) (for humans L is ~ 1.5 m!). The large size of DNA has a number of important consequences; single fluorescently labelled DNA molecules are visible under an optical microscope, which proves very useful for high resolution experiments, and the cell has to solve a tricky packaging problem in vivo of how to fit the DNA inside the nucleus of a cell which is, at most, a few microns in diameter (it uses chromosomes).

1.4 CARBOHYDRATES

Historically, advances in carbohydrate research have been overshadowed by developments in protein science. This has in part been due to the difficulty of analysing of the structure of carbohydrates, and the extremely large variety of chemical structures that occur naturally. Carbohydrates play a vital role in a vast range of cellular processes that are still only partly understood.

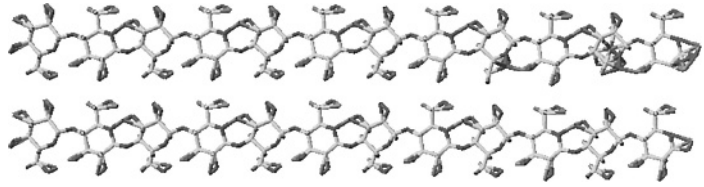


Figure 1.14 Sheet-like structures formed in cellulosic materials
(The $\beta(1 \rightarrow 4)$ linkages between glucose monomers induce extended structures, and the cellulose chains are linked together with hydrogen bonds.)

There are two important glucose polymers which occur in plants that are differentiated by the linkage between the monomers: cellulose and amylopectin. *Cellulose* is a very rigid polymer, and has both nematic and semi-crystalline phases. It is used widely in plants as a structural material. The straight chain formed by the $\beta(1 \rightarrow 4)$ linkage between glucose molecules is optimal for the construction of fibres, since it gives them a high tensile strength in the chain direction (Figures 1.14 and 1.15), and reasonable strength perpendicular to the chain due to the substantial intrachain hydrogen bonding in sheet-like structures. *Amylose* and its branched form, *amylopectin* (starch), are used in plants to store energy, and often amylopectin adopts smectic liquid crystalline phases

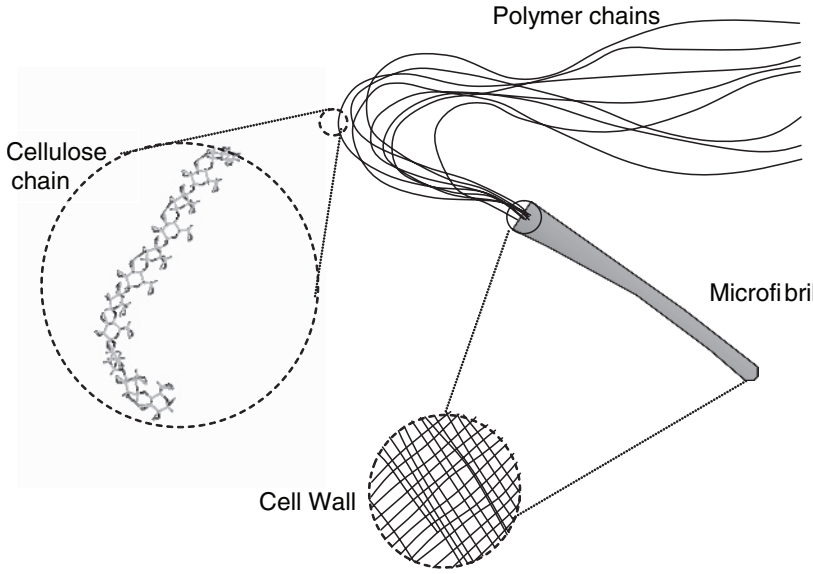


Figure 1.15 The hierarchical structure of cellulose found in plant cell walls
(Cellulose chains are combined into microfibrils that form the walls of plant cells [*Ref.: adapted from C.K. Mathews and K.E. Van Holde, Biochemistry, Benjamin Cummings*])

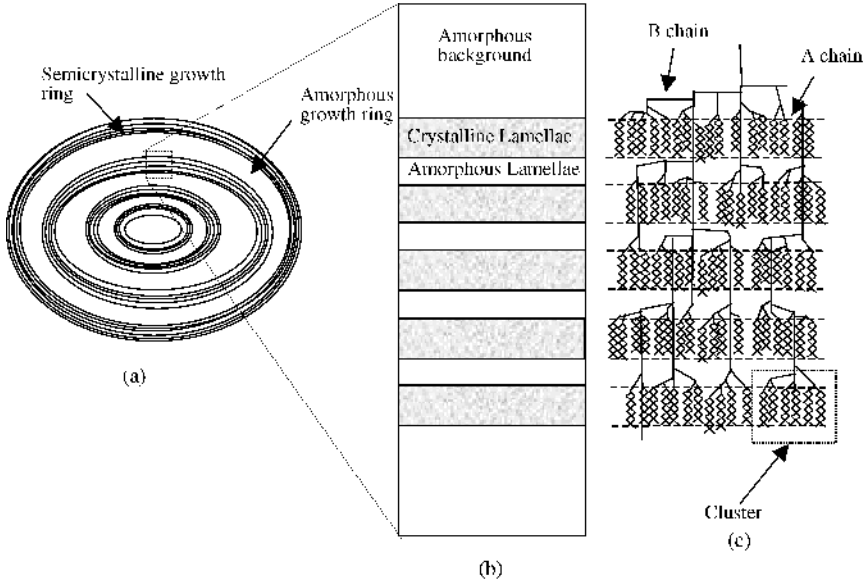


Figure 1.16 Four length scales are important in the hierarchical structure of starch; (a) the whole granule morphology ($\sim$ μm), (b) the growth rings ($\sim$ 100 nms), (c) the crystalline and amorphous lamellae ($\sim$9 nm), and (d) the molecular structure of the amylopectin ($\sim$Å). [*Ref.: T.A.Waigh, PhD thesis, University of Cambridge, 1996*]

(Figure 1.16). Starch, an amylose/amylopectin composite, forms the principle component of mankind's food sources. In amylose the glucose molecules are connected together with an α $(1 \rightarrow 4)$ linkage. α-linkages between the glucose molecules are well suited to the formation of an accessible sugar store, since they are flexible and can be easily degraded by enzymes. Amylopectins are formed from amyloses with additional branched α $(1 \rightarrow 6)$ flexible linkages between glucose molecules (Figure 1.17). Glycogen is an amorphous hyperbranched glucose polymer analogous to amylopectin, and is used inside animal cells as an energy store.

Chitin is another structural polysaccharide; it forms the exoskeleton of crustaceans and insects. It is similar in its functionality to cellulose, it is a very rigid polymer and has a cholesteric liquid crystalline phase.

It must be emphasised that the increased complexity of linkages between sugar molecules, compared with nucleic acids or proteins, provides a high density mechanism for encoding information. A sugar molecule can be polymerised in a large number of ways, e.g. the six corners of a glucose molecule can each be polymerised to provide an additional N^6 arrangements for a carbohydrate compared with a protein

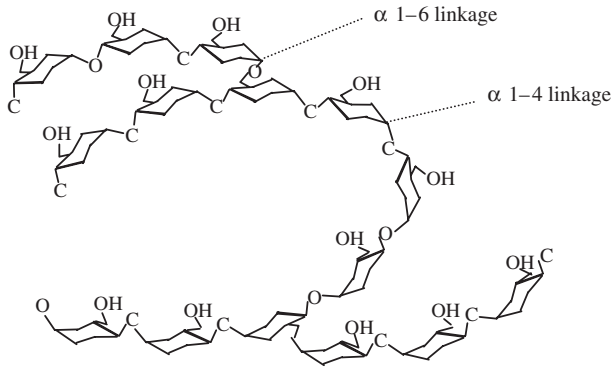


Figure 1.17 The branched primary structure found for amylopectin in starch (Both $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ flexible linkages occur between adjacent glucose monomers.)

of equivalent length (N). In proteins there is only one possible mechanism to connect amino acids, the peptide linkage. These additional possibilities for information storage with carbohydrates are used naturally in a range of immune response mechanisms.

Pectins are extra cellular plant polysaccharides forming gums (used in jams), and similarly *algins* can be extracted from sea weed. Both are widely used in the food industry. *Hyaluronic* acid is a long negatively charged semi-flexible polyelectrolyte and occurs in a number of roles in animals. For example it is found as a component of cartilage (a biological shock absorber) and as a lubricant in synovial joints.

1.5 WATER

Water is a unique polar solvent and its properties have a vast impact on the behaviour of biological molecules (Figure 1.18). Water has a high

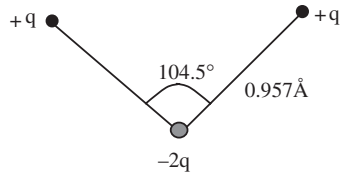


Figure 1.18 The geometry of a single water molecule (The molecule tends to form a tetrahedral structure once hydrogen bonded in ice crystals (Figure 2.2).)

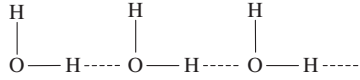


Figure 1.19 Schematic diagram of the network structure formed by water molecules (Dashed lines indicate hydrogen bonds. Such chains of hydrogen bonded water molecules occur over a wide range of angles for liquid water.)

dipole moment (P) of 6.11×10^{-30} Cm, a quadrupole moment of 1.87×10^{-39} Cm^2 and a mean polarisability of 1.44×10^{-30} m^3.

Water exists in a series of crystalline states at sub zero temperature or elevated pressures. The structure of ice formed in ambient conditions has unusual cavities in its structure due to the directional nature of hydrogen bonds, and it is consequently less dense than liquid water at its freezing point. The polarity of the O–H bonds formed in water allows it to associate into dimers, trimers etc (Figure 1.19), and produces a complex many body problem for the statistical description of water in both liquid and solid condensed phases.

Antifreeze proteins have been designed through evolution to impair the ability of the water that surrounds them in solution to crystallise at low temperatures. They have an alpha helical dipole moment that disrupts the hydrogen bonded network structure of water. These anti-freeze molecules have a wide range of applications for organisms that exist in sub zero temperatures e.g. arctic fish and plants.

The imaging of biological processes is possible in vivo using the technique of nuclear magnetic resonance, which depends on the mobility of water to create the image. This powerful non-invasive method allows water to be viewed in a range of biological processes, e.g. cerebral activity.

Even at very low volume fractions water can act as a plasticiser that can switch solid biopolymers between glassy and non glassy states. The ingress of water can act as a switch that will trigger cellular activity in plant seeds, and such dehydrated cellular organisms can remain dormant for many thousands of years before being reactivated by the addition of water.

A wide range of time scales (10^{-18}–10^3 s) of water are important to understand its biological function (Figure 1.20). The range of time scales includes such features as the elastic collisions of water at ultra fast times ($\sim 10^{-15}$ seconds) to the macroscopic hydrodynamic processes observed in blood flow at much slower times ($\sim$seconds).

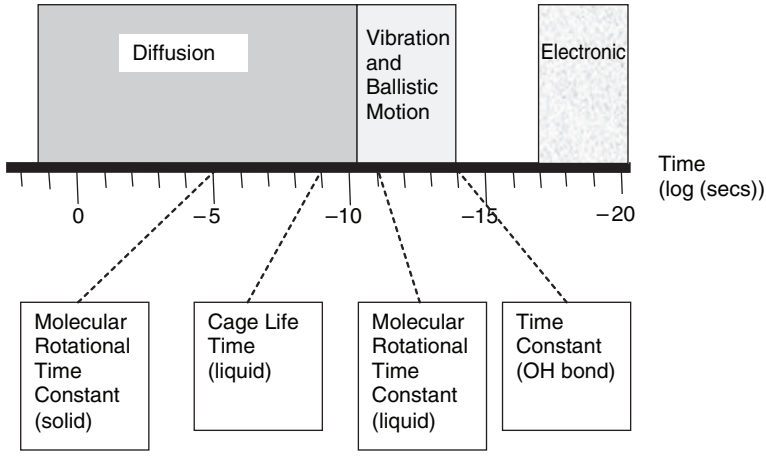


Figure 1.20 The range of time scales that determine the physical properties of water, shown on a logarithmic scale

1.6 PROTEOGLYCANS AND GLYCOPROTEINS

Proteoglycans (long carbohydrate molecules attached to short proteins) and glycoproteins (short carbohydrate molecules attached to relatively long proteins) are constructed from a mixture of protein and carbohydrate molecules (the glycosoaminoglycans). In common with carbohydrates, proteoglycans/glycoproteins exhibit extreme structural and chemical heterogeneity. Furthermore, the challenges presented to crystallography by their non-crystallinity means that a full picture of the biological function of these molecules is still not complete.

Many proteoglycans and glycoproteins used in the extracellular matrix have a bottle brush morphology (Figures 1.21 and 1.22). An

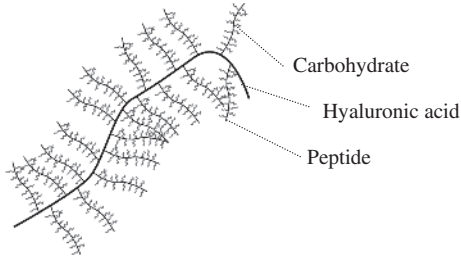


Figure 1.21 A schematic diagram of the aggrecan aggregate
(The aggrecan monomers (side brushes) consist of a core protein with highly charged carbohydrate side-chains. The bottle brushes are physically bound to the linear hyaluronic acid backbone chain to form a super bottle brush structure [*Ref.: A. Papagiannopoulos, T.A.Waigh, T. Hardingham and M. Heinrich, Biomacromolecules, 2006, 7, 2162–2172*])

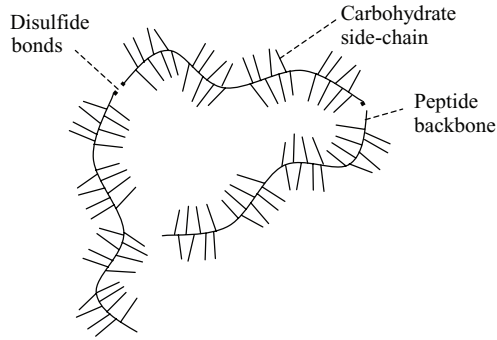


Figure 1.22 Porcine stomach mucin molecules contain a series of carbohydrate brush sections that are connected to a peptide backbone. The ends of the peptide are sticky, and these telechelic bottle brushes form thick viscoelastic gels at low pHs.

example of a sophisticated proteoglycan is aggrecan, a giant polymeric molecule that consists of a bottle-brush of bottle-brushes (Figure 1.21). These materials have a very large viscosity in solution, and are used to dissipate energy in collageneous cartilage composites and to reduce friction in synovial joints as boundary lubricants. An example of an extracellular glycoprotein is the mucins found in the stomach of mammals. These molecules experience telechelic (either end) associations to form thick viscoelastic gels that protect the stomach lining from autodigestion (Figure 1.22).

Other examples of glycoproteins occur in enzymes (Ribonuclease B), storage protein (egg white), blood clots (fibrin) and antibodies (Human IgG).

1.7 CELLS (COMPLEX CONSTRUCTS OF BIOMOLECULES)

Cells act co-operatively in multicellular organisms and are hierarchically arranged into tissues, organs and organ systems. Tissues contain both cells and other materials such as the extracellular matrix.

There are four distinct forms of mammalian *muscle cells*: skeletal and cardiac (which both form striated musclar tissues), smooth muscle (found in blood vessels and intestines) and myoepithlial cells (again present in intestines).

Nerve cells are used to send and receive signals. They are highly branched and this structure allows them to react to up to one

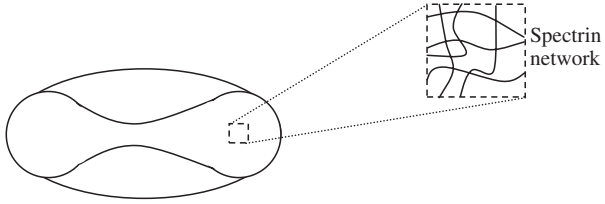


Figure 1.23 The cross-section through a squashed donut shaped blood cell (The spectrin network in the cell wall is a dominant factor for the determination of the morphology of the cell.)

hundred thousand inputs from other cells. The electrochemistry of nerve cells is a fascinating area; the efficiency and time response of these electrical circuits has been carefully optimised by evolution.

Blood cells have a squashed donut shape (Figure 1.23) which is related to the differential geometry of their cytoskeleton. Red blood cells carry oxygen and carbon dioxide, towards and away from the lungs. White blood cells play a role in the fight to remove infections from an organism.

Fibroblast cells are largely responsible for the secretion and regulation of the extracellular matrix, e.g. the production of molecules such as the collagens. *Epithelial cells* control the passage of material across the boundary of organs, e.g. in the interior of the intestinal tract.

1.8 VIRUSES (COMPLEX CONSTRUCTS OF BIOMOLECULES)

Viruses are intra-cellular parasites, biological entities that multiply through the invasion of cellular organisms. In addition to aspects related to their biological role in disease, viruses have attracted a great deal of attention from biophysicists for their physical properties. Viruses self-assemble into well defined monodisperse geometrical shapes (rods and polyhedra) (Figure 1.24) from their constituent components. Such materials have proven ideal model systems for the examination of the phase behaviour of charged colloids and lyotropic liquid crystals (Chapter 4), and allow the processes involved in their self-assembly to be investigated in detail (Chapter 6).

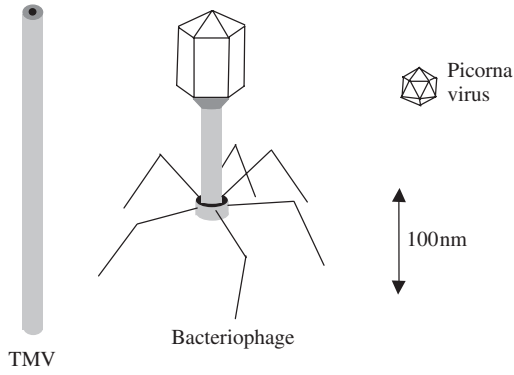


Figure 1.24 Schematic diagram of a range of virus structures (rod-like (TMV), asymmetric (bacteriophage), and icosohedral (picorna))

1.9 BACTERIA (COMPLEX CONSTRUCTS OF BIOMOLECULES)

Bacteria are small structurally simple cellular organisms. Only a minority of bacterial species have developed the ability to cause disease in humans. Bacteria take the form of spheres, rods and spirals. They will be encountered in terms of their mechanisms of molecular motility in Chapter 5 and Chapter 14.

1.10 OTHER MOLECULES

ADP and ATP are the 'currency of energy' in many biochemical processes. Energy is stored by the addition of the extra strongly charged phosphate group in the ATP molecule and can be released when it is metabolised into ADP. There are a vast range of other biomolecules that commonly occur in biology, and the reader should refer to a specialised biochemistry textbook for details.

FURTHER READING

For more exposure to the exquisite detail contained in molecular biophysics the student is directed to:

L. Stryer, *Biochemistry*, Freeman, 1995. Comprehensive coverage of basic biochemical processes.

B. Alberts, A. Johnson, J. Lewis *et al.*, *The Molecular Biology of the Cell*, Garland Science, 2002. A good introductory text to cellular biochemistry that is useful once the contents of Stryer have been fully digested.
D. Goodsell, *The machinery of life*, Springer, 1992. A simple discursive introduction to biochemistry with some attractive illustrations.

TUTORIAL QUESTIONS

1.1) A DNA chain has a molecular weight of 4×10^8 and the average monomer molecular weight of a nucleic acid subunit is 660 Da. For an A type helix there are 11 residues per helical pitch, and the translation per residue is 2.6 Å. For a B type helix there are 10 residues per helical pitch and the translation per residue is 3.4 Å. For a Z-type helix there are 12 residues per helical pitch and the translation per residue is 3.7 Å. What is the length in cm of a duplex DNA chain if it is in the A, B and Z helical forms? What is the average size of the nucleus of a mammalian cell? How does the cell manage to accommodate the DNA in its nucleus?

1.2) Suppose that you isolate a lipid micelle that contains a single protein that normally exists as a transmembrane molecule. How would you expect the lipid and protein to be arranged on the surface of the micelle?

1.3) Calculate the pH of a 0.2 M solution of the amino acid arginine if its pK_a value is 12.5.

1.4) Metals occur in a range of biological processes and form a key component of the structures of a number of biological molecules. Make a list of the biological molecules in which metal atoms occur.