CHAPTER 1

BIOPOLYMERS

In this chapter, the basic properties of biopolymers will be briefly discussed. We will group them according to nucleic acids, proteins and polysaccharides and we will summarize their main biological functions. Biopolymers have the unique feature that they exhibit a hierarchy in their molecular structures. Associated with these structures, their biological functions emerge almost naturally. In the latter context, think about the importance of the double-helical structure of DNA for the replication process. It is important to realize that these biological functions are based on the way the building blocks (nucleotides, amino acids, carbohydrates, *etc.*) are assembled. We will subsequently present the primary, secondary and some tertiary structures of nucleic acids, proteins and polysaccharides and show how they are stabilized by interactions. However, a detailed discussion of the chemical composition of the various biopolymers and their biological functions is beyond the scope of this book and for this purpose the reader is referred to the dedicated literature (see, for instance, the textbooks of Mathews, van Holde and Ahern and Bloomfield, Crothers and Tinoco).[1,2]

1.1 Introduction

Biopolymers or biomacromolecules can be roughly classified according to three different categories: nucleic acids, proteins and polysaccharides (carbohydrates). It should be born in mind that this classification is not strict and that there are important exceptions. An example is glycoprotein, which is a combination of protein and carbohydrate and plays a role in, among others, immune cell recognition and tissue adhesion. The biological functions of nucleic acids, proteins and polysaccharides are also different. Nucleic acids are

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

involved with the storage of the genetic code (DNA) and the translation of the genetic information into protein products (RNA). Proteins catalyze biochemical reactions (enzymes), have structural or mechanical functions or are important in cell signalling and immune responses. The structural components of plants are primarily composed of the polysaccharide cellulose. Bacteria excrete polysaccharides for adhesion to surfaces and to avoid dehydration. Examples of these polysaccharides are dextran, xanthan and pullulan, which have found wide-spread applications in pharmacy, biotechnology and the food industry. The classification according to the functioning of the biopolymers is also not unique. An important exception is the ribosome; an organelle on which proteins are assembled. A ribosome contains 65% RNA and 35% protein. It can be considered an enzyme, but its active site is made of RNA. However, the functioning and purpose of biopolymers in the machinery of life is beyond the scope of this book. Here, we intend to explore the extent to which their properties can be understood in terms of concepts from physics and mathematics.

Like every polymer, biopolymers are strings or sequences of monomeric units or monomers for short. In many cases these strings are linear, but sometimes they are closed and circular, branched or even cross-linked. In the latter case, we are dealing with a gel. In this book, we will primarily focus on linear polymers, but we will also discuss star-branched polymers, spherical polymer brushes and closed circular, supercoiled DNA. The structure of any biopolymer is determined by the nature of the building blocks (*i.e.* the monomeric units) in combination with environmental conditions such as the temperature, the solvent (water) and the presence of salts and/or other molecular components. The monomeric units of nucleic acids, proteins and polysaccharides are largely different and will be discussed in the next section. A unique feature of biopolymers is that most of them are essentially heteropolymers, because they may contain a variety in monomeric units. The biological relevance of a biopolymer is ultimately based on the sequence of the monomers, *i.e.* the *primary* structure. In the case of DNA, the primary structure is the sequence of bases attached to the sugar rings, which determines the genetic code. For proteins, it is the amino acid sequence, which eventually determines, together with environmental conditions, their 3–dimensional shapes and biological functions. The properties of polysaccharides are also largely determined by the nature of the monomeric

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

units, more specifically in the way they are connected.

A fundamental characteristic of biopolymers is the formation of hierarchical structures at successive length scales. Starting from the primary structure, the monomeric units are organized in a certain local molecular conformation. This local conformation is commonly referred to as the *secondary* structure. Examples of secondary structures are the famous double-helical arrangement of the two opposing strands in the DNA molecule (the duplex) and α – helixes and β – sheets formed by the polypeptide chains in proteins. At a larger distance scale, a biopolymer can adopt a defined 3–dimensional conformation: the so-called *tertiary* structure. This is particularly relevant for proteins, which largely owe their biological functioning to their 3–dimensional structure, but also nucleic acids and polysaccharides have tertiary structures. An example of a nucleic acid with a tertiary structure is transfer RNA, which has an L–shaped 3–dimensional structure that allows them to fit into the active site of the ribosome (it transfers a specific amino acid residue to a growing polypeptide chain). Eventually, biopolymers can form even larger complexes among themselves and with other macromolecular components in the cell and organism.

Biopolymers have *emergent* properties associated with their hierarchical structures. Here, the meaning of emergence is that the biopolymers have properties that cannot be attributed to the individual building blocks. For instance, the nucleic acid bases are just molecular components made of carbon, nitrogen and oxygen. It is their specific *sequence* in a strand of the DNA or RNA molecule that carries the genetic code. This property cannot be attributed to the individual bases, but it has emerged from the assembling of the bases into the nucleic acid. It is also obvious that the activity of a protein is an emerging property of the hierarchical assembling of the amino acids. Here, it is even possible to replace a selected and limited number of amino acids by other amino acids without losing the biological function of the protein. Emergence is a general phenomenon associated with the assembling of building blocks into larger scale structures, both in civil engineering and in biology.

Throughout this book, we will almost exclusively deal with systems in thermodynamic equilibrium. Although this is of interest in its own right, it is only fair to say that the study of systems in thermodynamic equilibrium has a limited relevance for our understanding of life. It is commonly believed that

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

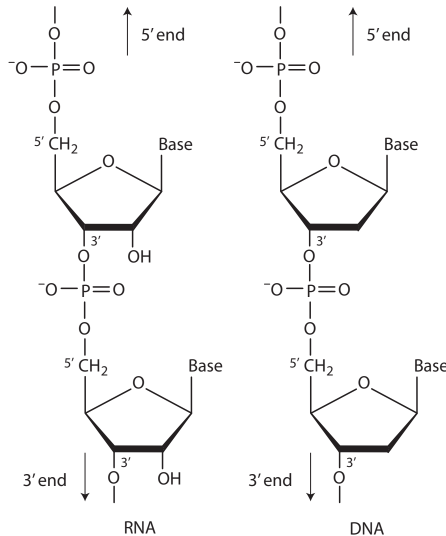

Figure 1.1 Chemical structures of ribonucleic acid (RNA, left) and deoxyribonucleic acid (DNA, right). The phosphate groups and the five carbon sugar rings are shown in detail. The bases are shown schematically, but their chemical structures are depicted in Fig. 1.2.

spontaneous assembling processes, driven by the minimization of the system's free energy (self-assembly), are important in biology. However, one should bear in mind that life exists by the virtue of the dissipation of energy, mainly through the hydrolysis of adenosine triphosphate (ATP). By definition, a living organism is in a non-equilibrium state and it is not always possible to generalize the concepts obtained for equilibrium conditions. Understanding life on the basis of non-equilibrium, dissipative processes is clearly a challenge for the future.

1.2 Primary structures

1.2.1 Nucleic acid primary structure

There are two types of nucleic acids, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). As shown in Fig. 1.1, each molecule is a

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

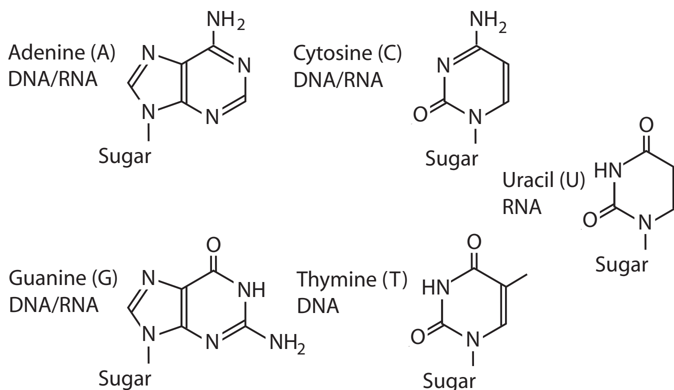

Figure 1.2 Pyrimidine (cytosine, C; thymine, T; uracil, U) and purine (adenine, A; guanine, G) bases in DNA and RNA.

polymeric chain, in which the units are covalently linked by the phosphates. The monomeric units are the nucleotides. Each nucleotide is built around a five-carbon sugar; ribose in RNA and 2'–deoxyribose in DNA. In Fig. 1.1 the five carbon atoms of the sugar are counted from the one to which the base is attached at the right, down through the ring and then up to the fifth carbon at the upper left side. Besides a difference in bases, which will be discussed shortly, the chemical difference between RNA and DNA lies in the replacement of a hydroxyl group by a hydrogen atom at the 2' position in DNA. The nucleotides are linked through the formation of a phosphodiester between the 5' carbon of one nucleotide and the 3' carbon of the next nucleotide. In this way, long nucleic acid chains sometimes contain millions of units which are attached to each other. It is important to realize that the string of nucleotides has a direction from the 3' to the 5' end. The phosphate group is a strong acid with a pK_a of around one. RNA and DNA are thus strong acids and under physiological conditions every phosphate moiety carries a negative charge. DNA and RNA are so-called polyelectrolytes and the presence of charge results in specific properties, such as an electrostatic contribution to the bending rigidity of the molecule. This and other effects of the presence of charge will be detailed in Chapter 3.

The backbone of the nucleic acid molecule is a repetitive structure and by itself it cannot store information. It is clear that the information storage

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

capacity is derived from the sequence of bases, each of which is attached to the 1' carbon of the sugar ring. There are two types of bases: the purines and pyrimidines. In the case of DNA, there are two purines, adenine (A) and guanine (G) and two pyrimidines, cytosine (C) and thymine (T). In the case of RNA, uracil (U) replaces thymine (see Fig. 1.2). DNA and RNA also contain a small fraction of chemically modified bases; some of these can induce alternate secondary structures, as will be discussed in Chapter 5. Note that the bases do not carry charge, but they can form hydrogen bonds.

1.2.2 Protein primary structure

All proteins are polymers and their monomeric units are $\alpha-$amino acids. The amino group is attached to the $\alpha-$carbon, *i.e.* the carbon next to the carboxyl group. Under physiological conditions, the amino acid is in its zwitterionic form; the amino group has picked up a proton and has become positively charged and the carboxyl group has dissociated a proton and is negatively charged. Besides the amino group, a hydrogen atom and a side group are also attached to the $\alpha-$carbon of every amino acid. The amino acids are distinguished by their different side groups. Twenty chemically different amino acids are incorporated in proteins; their structures are shown in Fig. 1.3. In the simplest case, glycine, the side group is just a hydrogen atom. The amino acids can be grouped according to the physical-chemical properties of the side group: aliphatic, hydroxyl or sulphur containing, cyclic (proline), aromatic, basic or acidic. It is clear that the higher order secondary and tertiary structures of proteins are intimately related to these properties, together with environmental factors such as the solvent quality.

With the exception of glycine, there are always four different chemical groups attached to the $\alpha-$carbon of every amino acid. Accordingly, amino acids are chiral and each one can occur in two different stereoisomers: the $D-$ and $L-$forms. The $L-$form of alanine is displayed in Fig. 1.4; it has the amino, hydrogen, carboxyl and methyl groups arranged in a clockwise manner, when the $\alpha-$carbon is viewed from the top with the amino and carboxyl groups pointing downwards and the hydrogen and methyl group pointing upwards. All amino acids incorporated by organisms into proteins are of the $L-$form. The chirality of the amino acids has an important consequence for the

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

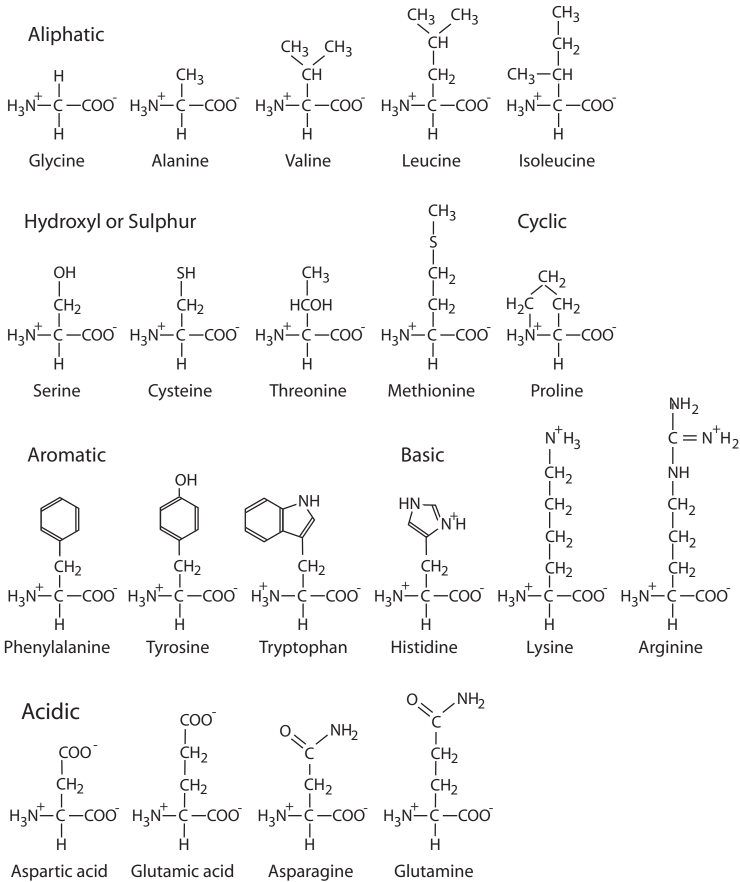

Figure 1.3 The twenty standard $\alpha-$ amino acids found in proteins. Note that they have been arranged according to the properties of the side group. In organisms, more different amino acids are present, but those are not incorporated in proteins.

secondary structure. For instance, owing to the steric interactions among the side groups, only right-handed $\alpha-$ helixes are possible. Left-handed helixes can be obtained by using synthetic amino acids in the D–form.

Amino acids can be covalently linked by the formation of a peptide bond between the $\alpha-$ carboxyl group and the $\alpha-$ amino group. This is illustrated in Fig. 1.4 for the link between alanine and glycine in order to form

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

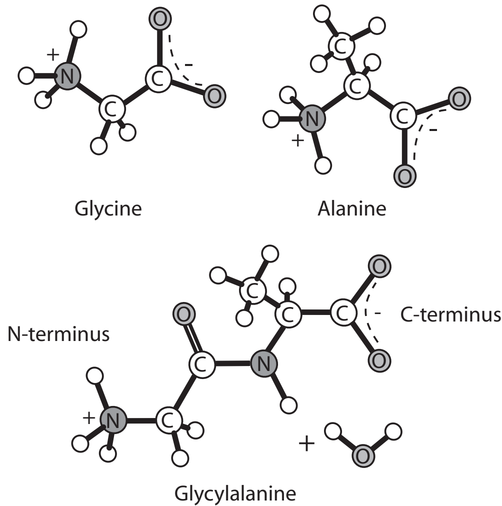

Figure 1.4 Formation of the peptide bond. Here, glycine and alanine are linked to form the dipeptide glycylalanine by the removal of a water molecule. Note that the peptide bond is planar. Redrawn from Ref. [1].

glycylalanine. The carbonyl C=O and the N–H bonds must remain in the same plane with only a little twisting around the C–N bond possible, because of the electron resonance structure of the peptide bond. Furthermore, due to the steric interaction between the side groups, the *trans* form is the favoured configuration. In this way, many amino acids can be linked to form a polypeptide. All proteins are polypeptides of a defined sequence determined by the genetic code. This sequence of amino acids is the primary structure of the protein, upon which all higher levels of organisation are based. As in the case of nucleic acids, the string of amino acids has a direction. At one side there is the amino N–terminus and at the other side the carboxyl C–terminus. Note that the polypeptide backbone is not charged besides the end groups. The charges of a protein are located in the side groups. Since a particular side group can be neutral or charged either positively or negatively, the net charge of a protein depends on the amino acid composition as well as the pH of the supporting medium. Under physiological conditions the protein is usually close to the iso-electric point, so that the positive and negative charges cancel out and the net charge is almost zero.

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

1.2.3 Polysaccharide primary structures

Polysaccharides are polymers of monosaccharides linked with glycosidic bonds. The monomers are cyclic structures, mostly containing 5 (pentoses) or 6 (hexoses) carbon atoms. An example of a pentose is ribose, which is one of the building blocks of nucleic acids. Many polysaccharides are made of hexoses, such as sucrose and galactose. Amylose and cellulose are linear chains of $\alpha - D$–glucose and $\beta - D$–glucose, respectively (see Fig. 1.5). The glycosidic bond is formed between a hydroxyl group on one carbohydrate unit with a hydroxyl group on another unit. In the case of amylose and cellulose the links are formed between the first and fourth carbon atom. It is customary to indicate these bonds by the numbers of the linked atoms and the stereoisomer of the unit, so amylose and cellulose are linked by $\alpha - 1,4$ and $\beta - 1,4$ glycosidic bonds, respectively. The primary structure of a polysaccharide can be more complicated. For instance, pullulan is a linear

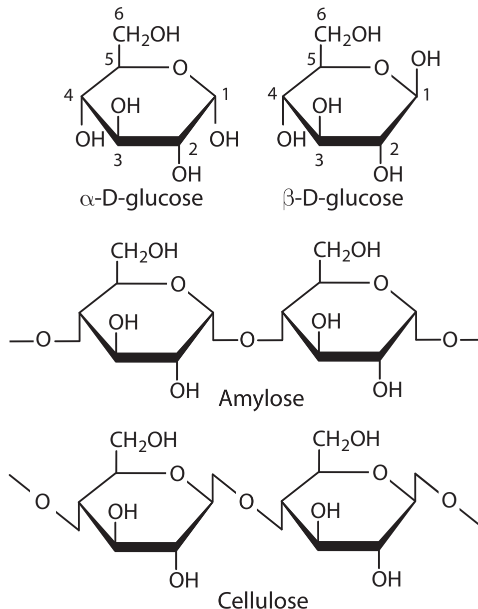

Figure 1.5 Amylose and cellulose are linear polysaccharides made by connecting $\alpha - D$–glucose and $\beta - D$–glucose, through $\alpha - 1,4$ and $\beta - 1,4$ glycosidic bonds, respectively and the removal of a water molecule.

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

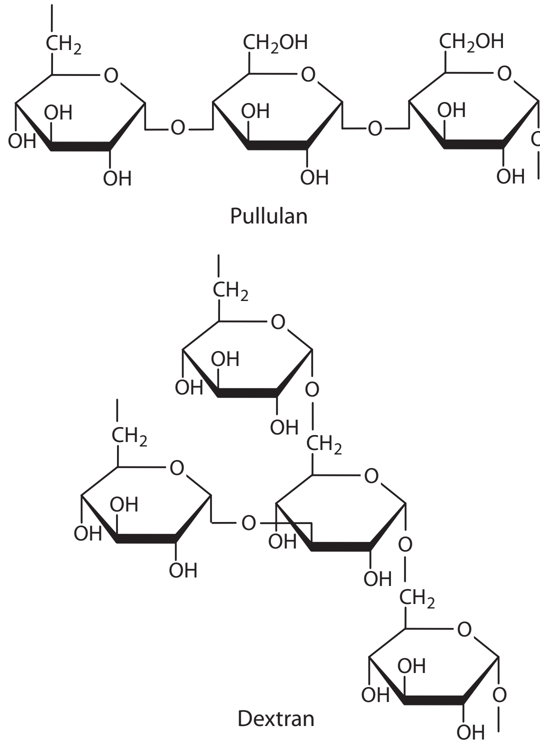

Figure 1.6 Primary structures of pullulan and dextran. Pullulan is a linear polymer with a repeating maltotriose unit of three glucose units. Dextran has a branched structure with on average about 100 monomers between the branch points.

polymer of maltotriose units. Three glucose units in maltotriose are connected by an $\alpha - 1,4$ bond; whereas consecutive maltotriose units are connected by $\alpha - 1,6$ bonds (see Fig. 1.6). Dextran is a branched polysaccharide made of many glucose molecules joined into chains of varying lengths. The linear chain sections are linked by $\alpha - 1,6$ bonds between glucose units, while branches begin from $\alpha - 1,3$ linkages (and in some cases, $\alpha - 1,2$ and $\alpha - 1,4$ linkages as well). Polysaccharides are never as complex as proteins or nucleic acids; they usually contain no more than two kinds of residues. Furthermore, polysaccharide chains have a random degree of polymerisation, in contrast with proteins and nucleic acids which are almost always of a defined length.

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

In their basic form, polysaccharides are uncharged. However, they are often functionalized with carboxyl groups, phosphate groups and/or sulphuric ester groups. The monomeric units contain many hydroxyl groups, which can engage in intra- and inter-molecular formation of hydrogen bonds. This hydrogen bonding keeps the chains together and contributes to the high tensile strength of the polymeric material. In this context, it is interesting to note that other forms of functionalization also occur. For instance, chitin, which is a major component of the exoskeletons of crustaceans and insects, can be described as cellulose with one hydroxyl group on carbon 2 of each glucose unit substituted by an acetylated amino (acetylamine) group. This substitution allows for increased hydrogen bonding, which gives the matrix formed by the polymer increased strength. Some polysaccharides such as cellulose are insoluble in water, whereas for others (*e.g.* dextran or pullulan) water is a moderate to excellent solvent.

1.3 Secondary structures

1.3.1 Secondary structures of nucleic acids

The bases of DNA and RNA can form base-pairs stabilized with hydrogen bonds. As shown in Fig. 1.7, adenine can form two hydrogen bonds with thymine, whereas guanine can form three hydrogen bonds with cytosine. With these pairing arrangements between the purines and pyrimidines, the distances between the 1' carbons of the attached sugars are the same (1.08 nm). In this way, two opposing single strands with a complementary base sequence can form a double helix, which is regular in diameter. This would not be possible if purines pair with purines and/or pyrimidines with pyrimidines. Besides hydrogen bonding, the double helix is stabilized by dispersion forces resulting from correlated electron charge fluctuations in the stack of base-pairs.

RNA is usually single-stranded, but most RNA molecules can form hair-pin structures by base-pairing of self-complementary regions within the *same* molecule. Single-stranded DNA with self-complementary base sequences can also fold back on itself and form single-chain stacked base structures. At elevated temperature and/or in the presence denaturing agents, the single-

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

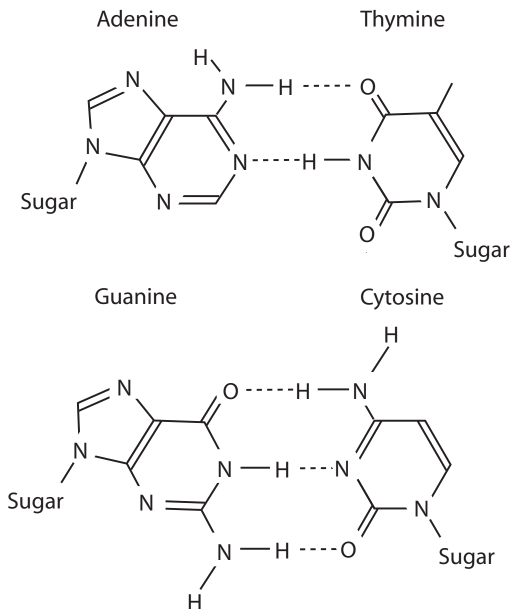

Figure 1.7 Base-pairing of thymine with adenine and cytosine with guanine.

stranded DNA molecule will take a random coil configuration. However, the canonical form of DNA is the double helix made of two complementary strands in an anti-parallel direction (the duplex). In the double helix, each strand can serve as a template for a complementary strand of DNA in the case of replication or for a complementary strand of messenger RNA in the case of the transcription of the genome for the synthesis of protein products.

The bases from the two opposing DNA strands in the duplex are stacked in the interior of the helix, whereas the two anti-parallel sugar-phosphate backbones are extended along the outside. The helix has a major and minor groove. Three secondary structures of the double-stranded DNA molecule have been identified: the *A*–, *B*– and *Z*–forms. The average values of the most important structural parameters are collected in Table I.I and space-filling representations are shown in Fig. 1.8. The main distinguishing features of these different secondary structures of DNA are[2]

- The *A*– and *B*–forms are right-handed and can be found in any sequence. *B* is the dominant form under physiological conditions. The *A*–form is

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

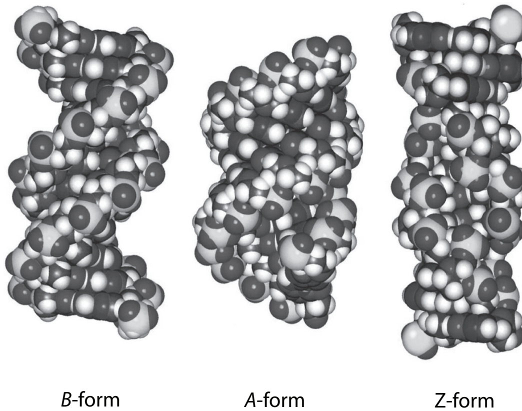

Figure 1.8 Space-filling representations of double-stranded DNA in the *B*–, *A*– and *Z*–form.

found at low hydration levels, such as in spun fibres. The *Z*–form is left-handed and occurs in alternating purine-pyrimidine sequences, particularly guanine-cytosine (GC).

- The double-stranded duplex in the *A*–form is thick and compressed along the helix; in the *Z*–form it is elongated and thin whereas in the *B*–form it is intermediate.
- There are 10, 11 and 12 base-pairs (bp) per turn in *B*–, *A*– and *Z*–DNA, respectively. The corresponding pitches are 3.2, 3.4 and 4.5 nm.

Table I.I Structural properties of DNA in the *A*–, *B*– and *Z*–form.[2]

Geometrical attribute	*A–form*	*B–form*	*Z–form*
Helix sense	right-handed	right-handed	left-handed
Repeating unit (bp)	1	1	2
Rotation/bp	32.7°	36.0°	60°/2
Mean bp/turn	11	10.0	12
Inclination of bp to axis	+12°	2.4°	-6.2°
Rise/bp along axis (nm)	0.29	0.34	0.37
Pitch/turn of helix (nm)	3.2	3.4	4.5
Diameter (nm)	2.6	2.0	1.8
Minor groove depth (nm)	0.28	0.75	0.9
Minor groove width (nm)	1.10	0.57	0.4
Major groove depth (nm)	1.35	0.85	-
Major groove width (nm)	0.27	1.17	-

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

- B–DNA has a wide major groove and a narrow minor groove, both of which are of similar depth. The A–form has a narrow, deep major groove and a wide, shallow minor groove. Z–DNA has a deep and narrow minor groove and no major groove.
- In B–DNA the base-pairs are almost perpendicular to the helix axis; those in A– and Z–DNA are inclined at larger angles.

We will almost exclusively deal with DNA in the regular B–form.

Owing to the presence of the hydroxyl group at the 2' position of the ribose sugar, base-paired RNA adopts the A–form geometry.

1.3.2 Secondary structures of proteins

Proteins show a wide variety of secondary structures. These structures satisfy a number of criteria: the peptide bond is planar because little twisting is possible about the C–N bond; the steric interaction between the side groups of the amino acids is minimal and the structure is stabilized by hydrogen bonding between the oxygen of the carbonyl C=O groups and the hydrogen of the amide C–N groups. Two major secondary structures, which satisfy these criteria, are the $\alpha-$helix and $\beta-$sheet (see Figs. 1.9 and 1.10, respectively). Note that these two structures are by no means the only ones. There exists other well defined, but less abundant secondary structures, such as the 3_{10} helix and some specific sharp turn loop sequences. Furthermore, there are also significant parts of the polypeptide chain which cannot be classified as one of these secondary structures. The latter parts have an irregular structure, but they are not random coils.

The structure of the $\alpha-$helix is shown in Fig. 1.9. There is an almost linear hydrogen bond formed between every carbonyl oxygen with an amide

Table I.II Geometrical attributes of polypeptide secondary structures.[1]

Structure	*Residues/ Turn*	*Rise/Residue* (nm)	*Pitch* (nm)
$\alpha-$ helix	3.6	0.15	0.54
$3_{10}-$ helix	3.0	0.20	0.60
Parallel $\beta-$ sheet	2.0	0.32	0.64
Anti-parallel $\beta-$ sheet	2.0	0.34	0.68

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

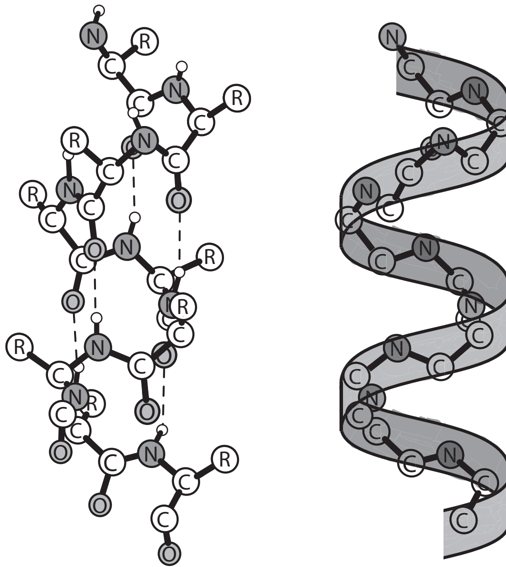

Figure 1.9 Left: Irregular $\alpha-$helical secondary structure of polypeptides. The hydrogen bonds between the carbonyl oxygen and the amide hydrogen are within a single polypeptide chain and almost parallel to the helix axis. The side groups point outwards. Right: Schematic helical ribbon representation showing the atoms of the backbone atoms only.

hydrogen on the *fourth* residue up the chain (separated by two residues). The hydrogen bonds are almost parallel to the helix axis. There is little or no steric interaction among the side groups, because they are pointing outwards away from the central axis of the helix. The $\alpha-$helix has 3.6 residues per turn, which results in a rise of 0.15 nm per residue and a pitch of 0.54 nm per turn. In the less abundant $3_{10}-$helix, there is a hydrogen bond between the carbonyl oxygen and the amide hydrogen of the *third* residue up the chain. Accordingly, the $3_{10}-$helix is less compressed in the longitudinal direction with 3.0 residues per turn and a rise of 0.20 nm per residue.

In the $\beta-$sheet, each residue is flipped by 180 degrees with respect to its preceding one and the polypeptide chain ($\beta-$strand) is folded in a zigzag fashion. As illustrated in Fig. 1.10, the linear hydrogen bonds are now formed

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

between *adjacent* chains almost perpendicular to the strand axis. Due to the consecutive flipping of the residues by 180 degrees, the side groups alternately point upwards and downwards away from the sheet. The $\beta-$sheet can be formed in two ways: parallel and anti-parallel. In the parallel configuration, the $\beta-$strands are all running in the same direction from the N– to the C–terminus. In the anti-parallel configuration, adjacent strands are running in opposite directions (as in Fig. 1.10). In the $\beta-$strand there are just two residues per turn, but the rise per residue differs between the parallel and anti-parallel configuration: 0.32 and 0.34 nm, respectively. The geometrical attributes of a number of secondary protein structures are collected in Table I.II.

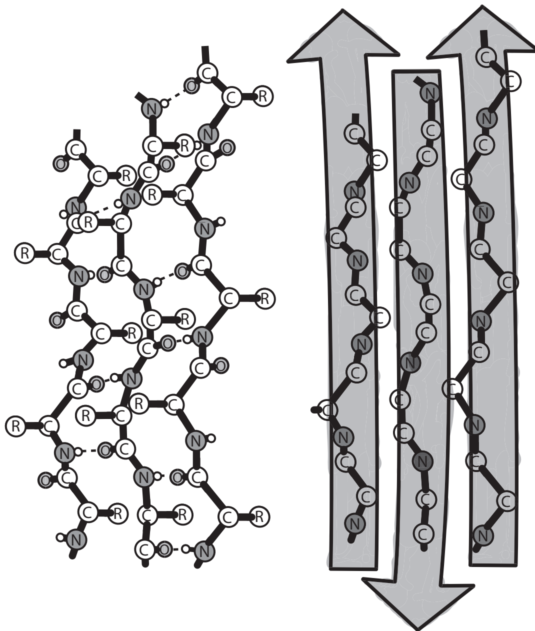

Figure 1.10 Left: Irregular anti-parallel $\beta-$sheet secondary structure of polypeptides. The hydrogen bonds between the carbonyl oxygen and the amide hydrogen are between adjacent chains and almost perpendicular to the chain axis. The side groups alternately point upwards and downwards along the chain. Right: Schematic representation showing the atoms of the backbone atoms only and the coarse-grained arrows which show the directions from the N– to the C–terminus.

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

In proteins, the secondary structures may be deformed by the presence of the side groups. The α – helix and β – strand structures are often depicted by the coarse-grained helical ribbon and deformed arrow shapes as shown in Figs. 1.9 and 1.10, respectively. The arrow heads at the ends of the β – strands point in the direction from the N– to the C–terminus.

1.3.3 Secondary structures of polysaccharides

Polysaccharides with a complex and/or branched primary structure, such as pullulan and dextran respectively, take a random coil conformation when dissolved in a suitable solvent (water). If the primary structure is simple and regular, polysaccharides may exhibit a regular secondary structure. In amylose, the regular orientation of successive glucose residues results in a right-handed helix with six residues per turn. Cellulose can exist as fully extended chains with each residue flipped by 180 degrees with respect to its neighbour in the chain. The cellulose chains form ribbons that are packed side-by-side with hydrogen bonds within and between them; a structure which is reminiscent of the β – sheet. Xanthan is a linear polysaccharide with a repeating unit made of 5 sugar units. To every repeating unit of the main chain a small side-chain is attached consisting of three modified sugar units. Two of these xanthan chains are thought to form a double helix, which gives the molecule a high bending rigidity and accounts for its surprisingly high solution viscosity.

1.4 Tertiary structure and stabilizing interactions

Naked double-stranded DNA, that is DNA not complexed with proteins, behaves as a charged polymer and takes a random coil conformation in water or an aqueous buffer. However, the biological relevance of naked DNA is limited. Inside the capsid of certain bacteriophages, double-stranded DNA is compacted and essentially protein-free, except for the proteins which make up the structure of the capsid itself. In the nucleoid region of bacterial cells, the genome is thought to be compacted by specific interactions with proteins as well as by osmotic, depletion effects exerted by non-binding proteins dispersed in the cytoplasm (the latter effects will be discussed in Chapter 6). In eukaryotic cells, DNA is wrapped around histone proteins and looks like

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

beads on a string when observed with an electron microscope. A section of 146 base-pairs of DNA with a contour length of around 50 nm is wrapped in 1.65 left-handed turns around the histone octamer, which is composed of four identical pairs of histone proteins. This assembly of DNA and protein is called the nucleosome core particle. The nucleosome core particles are connected with sections of 50 base-pairs of 'linker' DNA, together with another histone protein, so that the total repeating unit of the beads on the string is around 200 base-pairs. The core particles are stacked into a higher order structure called chromatin, which is organized in a hierarchical manner up to the level of the chromosome. Besides the structure of the nucleosome core particle, the structure of chromatin is largely unknown.

A special category of double-stranded DNA is plasmid. Plasmids are separated from chromosomal DNA and they usually occur in bacteria. Their size varies from around two to more than 400 kilo base-pairs. Plasmids are widely used as cloning vectors in genetic engineering, because they easily transfer from one bacterial cell to another and it is easy to insert DNA fragments at their restriction sites. Plasmids are often, if not always, circular, but the strands of the duplex are usually twisted a couple of times about their long axes before they are closed in order to form the ring. As a result of this *topological* constraint and the fact that the double-stranded DNA molecule can support twist, the plasmid molecule takes a 3–dimensional, supercoiled configuration. Supercoiling is not exclusive to plasmids; it also occurs in sections of chromosomal DNA as a result of complexation of protein on DNA. We will discuss supercoiling and supercoiling-induced transitions in the secondary structure of double-stranded DNA in Chapter 5.

Unlike DNA, RNA is usually single-stranded and has a much shorter chain of nucleotides. Single RNA strands often have self-complementary bases, which allow them to take a tertiary conformation by intra-molecular base-pairing and the formation of hair-pin structures. The tertiary conformation is stabilized by hydrogen bonding through the hydroxyl group at the 2' position of the ribose ring. The additional hydroxyl group also results in the *A*–form of the RNA double helix with a narrow, deep major groove and a wide, shallow minor groove. RNA molecules can also be packed into larger structures and/or form complexes with proteins. The ribosome is an example of the latter category.

Proteins have very rich tertiary structures, on which their biological

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

functions are based. They can be grouped according to their tertiary structures into two broad categories: the fibrous and globular proteins. The fibrous proteins are elongated and are usually of regular secondary structure. They are often structural elements in the cell and organism. The secondary structures of fibrous proteins can be, among others, $\alpha -$helix ($\alpha -$keratin) and $\beta -$sheet (silk fibroin). An interesting example is elastin, which has elastic properties because it contains cross-linked random coils. The random coils allow for the elastic deformation of the fibre without breaking the polypeptide bonds, like the cross-linked polymers in a natural or synthetic rubber. Globular proteins are compact and more or less spherical. The latter proteins often contain defined domains in which one can recognize structural elements such as bundles of $\alpha -$helixes and assemblies of $\beta -$strands in the form of twisted sheets and barrels.

The folding of the protein from a random coil state with an astronomical number of molecular configurations into its native state with a small number of possible configurations is accompanied by a tremendous loss in configurational entropy. In order to render the native state thermodynamically stable, this loss in entropy should be compensated by stabilizing interactions within the polypeptide sequence and/or an increase in *another* form of entropy. As we will see shortly, both effects are involved in the folding process. In the folded state, the most important stabilizing interactions are:

- *Charge interactions*. Many amino acids contain side groups which are either positively or negatively charged under physiological conditions close to the iso-electric point. The electrostatic attractive forces stabilize the native state. Far from the iso-electric point, the protein acquires a net positive or negative charge (depending on acidic or basic conditions) and the mutual repulsion among these charges will contribute to the instability of the folded structure and might eventually result in denaturation.
- *Hydrogen bonding*. Many side groups contain functional groups which can be involved in the formation of hydrogen bonds with other side groups and if available with the carbonyl oxygen and amide hydrogen on the polypeptide backbone. Although a single hydrogen bond is relatively weak, the sheer number of them can add a significant contribution to the stabilization of the folded state.

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

- *Van der Waals interaction*. The interior of globular proteins is closely packed with many uncharged side groups. The weak attraction resulting from dipole and induced dipole interactions between these side groups adds up and results in a significant stabilizing force.
- *Disulfide bonding*. If the protein is meant to function in an external, oxidizing environment, as opposed to the reducing environment inside most cells, significant stabilization of the folded structure can come from the formation of disulfide bonds between cysteine residues.
- *Hydrophobic interaction*. Despite the fact that the aforementioned interactions stabilize the native state to a significant extent, the main contribution to the stability of the protein comes from the hydrophobic effect. If the hydrophobic side groups are buried in the interior of the globular protein, water molecules that were first restricted in their translational and rotational motions due to the interaction with the protein are released. This release of hydration water molecules results in an increase of the entropy of the whole system including protein and solvent, which partially offsets the tremendous loss in configurational entropy associated with the folding process.

Relatively small proteins fold spontaneously into their 3–dimensional, native tertiary structures. For longer polypeptide sequences, the folding process may be assisted with helper proteins called chaperones, thereby avoiding misfolded states and possibly amorphous aggregation. We will further discuss the scientifically challenging folding problem in Sec. 5.4. Finally, one should bear in mind that many, if not all proteins are multi-unit assemblies and that they form higher order complexes with other biopolymers, such as DNA and RNA in the machinery of life.

1.5 Questions

1. What are the differences between DNA and RNA from a primary structural point of view?

2. Describe the difference in molecular structure of amylose and cellulose.

3. Give a reason why water is a good solvent for dextran and not a good

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html

solvent for cellulose.

4. Why does dextran not have a regular secondary structure as is found for amylose?

5. Why is the right-handed $\alpha-$helix much more abundant than the left-handed $\alpha-$helix in polypeptides of biological origin? Under which condition would the left-handed helix be more abundant?

6. Describe the differences between the parallel and the anti-parallel $\beta-$sheets of polypeptides. Why do they have a slightly different rise per residue?

7. Give a reason why a single-stranded DNA molecule does not take such an intricate tertiary structure as can be found in transfer RNA.

8. What happens to the 3–dimensional tertiary structure of a closed circular and supercoiled DNA molecule when one of the strands of the duplex is cleaved by an enzyme or accidentally cut (nicked)?

9. Why is purine or pyrimidine base-pairing not suitable for the formation of a double helix of two opposing strands of nucleic acid?

10. A protein made of 101 residues in its random coil state can exist in 3 to the power 100 conformations, if each link between residues has three equally probable configurations (see Sec. 5.4).
    a. Estimate the change in configurational entropy if the protein folds into a native structure with only one conformation.
    b. Suppose that the protein folds into a single $\alpha-$helix. Calculate the stabilization energy pertaining to the formation of the intra-molecular hydrogen bonds between carbonyl oxygen and amide hydrogen. Assume that each hydrogen bond contributes 5 kJ/mol to the stabilization energy.
    c. Is the $\alpha-$helix stable at 298 K?

INTRODUCTION TO BIOPOLYMER PHYSICS
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/physics/6644.html