16

TUTORIAL CHEMISTRY TEXTS

RS•C

Biophysical Chemistry

by ALAN COOPER

TUTORIAL CHEMISTRY TEXTS

16

Biophysical Chemistry

A. T. Austin
Glasgow University

RS•C
ROYAL SOCIETY OF CHEMISTRY

TUTORIAL CHEMISTRY TEXTS

This series of books consists of short, single-topic or modular texts, concentrating on the fundamental areas of chemistry taught in undergraduate science courses. Each book provides a concise account of the basic principles underlying a given subject, embodying an independent-learning philosophy and including worked examples. The one topic, one book approach ensures that the series is adaptable to chemistry courses across a variety of institutions.

TITLES IN THE SERIES

Stereochemistry *D G Morris*
Reactions and Characterization of Solids *S E Dann*
Main Group Chemistry *W Henderson*
d- and f-Block Chemistry *C J Jones*
Structure and Bonding *J Barrett*
Functional Group Chemistry *J R Hanson*
Organotransition Metal Chemistry *A F Hill*
Heterocyclic Chemistry *M Sainsbury*
Atomic Structure and Periodicity *J Barrett*
Thermodynamics and Statistical Mechanics *J M Seddon and J D Gale*
Basic Atomic and Molecular Spectroscopy *J M Hollas*
Organic Synthetic Methods *J R Hanson*
Aromatic Chemistry *J D Hepworth, D R Waring and M J Waring*
Quantum Mechanics for Chemists *D O Hayward*
Peptides and Proteins *S Doonan*
Biophysical Chemistry *A Cooper*
Natural Products: The Secondary Metabolites *J R Hanson*
Maths for Chemists, Volume I, Numbers, Functions and Calculus *M Cockett and G Doggett*
Maths for Chemists, Volume II, Power Series, Complex Numbers and Linear Algebra *M Cockett and G Doggett*
Inorganic Chemistry in Aqueous Solution *J Barrett*

FORTHCOMING TITLES

Mechanisms in Organic Reactions
Molecular Interactions
Biology for Chemists
Nucleic Acids
Organic Spectroscopic Analysis

Further information about this series is available at www.rsc.org/tct

Order and enquiries should be sent to:
Sales and Customer Care, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge CB4 0WF, UK

Tel: +44 1223 432360; Fax: +44 1223 426017; Email: sales@rsc.org

Cover images © Murray Robertson/visual elements 1998–99, taken from the 109 Visual Elements Periodic Table, available at www.chemsoc.org/viselements

ISBN 0-85404-480-9

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge CB4 0WF, UK
Registered Charity No. 207890
For further information see our web site at www.rsc.org

Typeset in Great Britain by Alden Bookset, Northampton
Printed and bound by Italy by Rotolito Lombarda

Preface

Biology is chemistry on an impressive scale. It is a product of evolution, the outcome of countless random experiments, resulting in the exquisite complexity of the biological world of which we are a part. Setting aside any philosophical considerations, living organisms – including ourselves – are simply nothing more than wet, floppy bags of chemistry: complicated mixtures of molecules interacting in a multitude of ways. All this takes place mainly in water, a solvent that most chemists try to avoid because of its complexities. However, we can learn from this. In the course of evolution, biology has had the opportunity to perform vastly more experiments than we can ever contemplate in the laboratory. The resulting chemistry is fascinating in its own right, and we can quite rightly study it for its intellectual satisfaction alone. We can also, if we choose, apply what we learn to other areas of chemistry and to its applications in biomedical and environmental areas.

This book is about the physical chemistry of biological macromolecules and how we can study it. The approach here is unashamedly experimental: this is the way science actually works, and in any case we do not yet have the rigorous theoretical understanding perhaps found in more mature areas of chemistry. This is what makes it a fun topic, and why it poses fascinating challenges for both theoretical and experimental scientists.

The level adopted in this tutorial text should be suitable for early undergraduate years in chemical or physical sciences. However, since this interdisciplinary topic is often postponed to later years, the book will also act as a basis for more advanced study. Students in other areas of biological sciences might also appreciate the less intimidating approach to physical chemistry that I have attempted here.

The term "biophysical chemistry" was brought to prominence by the work of John T. Edsall (1902–2002), who died just prior to his 100th birthday. Together with Jeffries Wyman, he wrote the original classic text: *Biophysical Chemistry, Volume 1* (Academic Press, 1958), but there never was a Volume 2. This book is dedicated to him and to the many other physical scientists who have dared to enter biological territory.

With thanks to my family and other animals who have tolerated me during the writing of this text, and to my students and other colleagues who have checked and corrected some of the material. I did not always follow their suggestions – so just blame me.

Alan Cooper
Glasgow

Contents

1
Biological Molecules

You do not need to know any biology in order to study biological molecules, but it does help to have some background.

Aims

This chapter will briefly review the bare bones of biological (macro)molecules. By the end, and together with previous knowledge and background reading, you should be able to:

- Describe the basic chemical structures of polypeptides, polynucleotides, fats, lipids and carbohydrates
- Explain what is meant by the primary, secondary, tertiary and quaternary structures of proteins
- Describe the behaviour of fats, lipids and detergents in water
- Explain the anomalous properties of liquid water
- Recall the fundamentals of acid–base equilibria

1.1 Introduction

This book is mainly about the experimental methods used to understand the physical properties and function of the molecules that make up living systems.

These molecules – proteins, polynucleotides, polysaccharides, lipids – are not necessarily any different from molecules we study in other branches of chemistry. However, there are some additional factors, arising from their biological origin, that we need to be aware of:

- Biological macromolecules are large molecules formed from many smaller units, and are (usually) polymers of precise length and specific sequence.
- They (usually) fold or associate into specific conformational assemblies stabilized by non-covalent interactions.

- This (usually) happens in water.
- The molecules are the (usually) successful outcomes of biological evolution.

It is this last point that makes things so exciting for the biophysical chemist. The molecules we see today are the results of countless random (more or less) experiments over millions of years, during which living systems have evolved to take advantage of subtle principles of physical chemistry that we barely yet understand. By studying such systems we can learn much about physical chemistry in general, with potential for applications in other areas.

1.2 Proteins and Polypeptides

D-Amino acids are encountered only in special instances such as bacterial cell walls and peptide antibiotics.

Proteins are polymers made up of specific sequences of L-amino acids linked together by covalent peptide (amide) bonds (Figure 1.1). Amino acids are chosen from a basic set of 20 building blocks differing in their side-chain (Figure 1.2), with occasional special-purpose side-chains made to order (*e.g.* hydroxyproline).

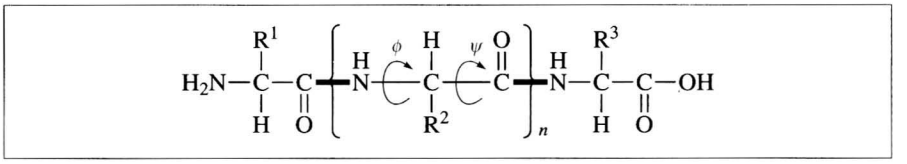

Figure 1.1 Polypeptide structure showing rotatable ϕ/ψ angles. The planar peptide (amide) bonds are shown in bold, and are usually *trans*

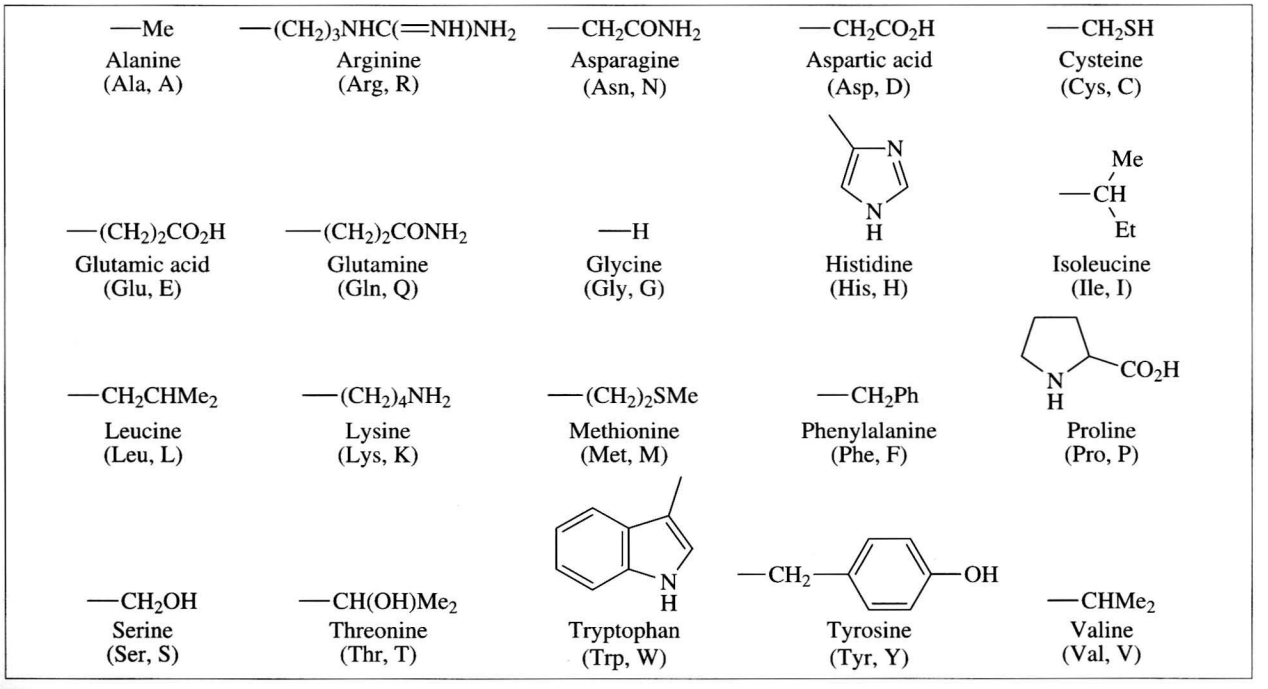

Figure 1.2 The 20 naturally occurring amino acid side chains (residues) with three-letter and single-letter abbreviations

Typical proteins range in polypeptide chain length from around 50 to 5000 amino acids. The average relative molecular mass of an amino acid is around 110, so proteins can have RMMs from 500 to 500,000 (0.5 to 500 kDa) or more, especially in multi-subunit proteins consisting of specific aggregates (see Table 1.1).

The term "molecular weight" is not strictly accurate (why?) but is commonly used, especially in the older (biochemical) literature. The more correct terms are "relative molecular mass (RMM)" (no units) or "molar mass" (kg mol^{-1} or g mol^{-1}). One dalton (1 Da) is equal to 1 amu (atomic mass unit).

Table 1.1 Some common proteins

Name	*No. of amino acids*	*RMM*	*Function*
Insulin	51 (2 chains, 21 + 30)	5784	Hormone controlling blood sugar levels. A-chain and B-chain covalently linked by disulfide bonds. Globular
Lysozyme (hen egg white)	129	14,313	An enzyme that catalyses hydrolysis of bacterial cell wall polysaccharides. Found in egg whites, tears and other biological secretions. Globular
Myoglobin	153	17,053	Oxygen transporter in muscle. Contains haem group. Globular
Haemoglobin	574 (2 × 141 + 2 × 146)	61,986 (2 × 15,126 + 2 × 15,867)	Oxygen transporter in blood stream. Consists of four subunits (two α and two β chains), with haem. Globular
Rhodopsin	348	38,892	Photoreceptor membrane protein in the retina of the eye. Contains 11-*cis* retinal as chromophore
Collagen	3200 (approx. 3 × 1060)	345,000	Connective tissue protein (collagen), skin, bone, tendon. Three-stranded triple helix. The most abundant protein in animals. Fibrillar
RuBISCO (ribulose bisphosphate carboxylase/ oxygenase)	4784 (8 × 475 + 8 × 123)	538,104 (8 × 52,656 + 8 × 14,607)	Carbon fixation enzyme of green plants and algae; 16 subunits (8 large, 8 small). The most abundant protein on Earth

Worked Problem 1.1

Q How many molecules are there in a 1 mg sample of a protein of 25,000 RMM?

A 25,000 RMM $\equiv$ 25,000 g mol^{-1}.

$$1\text{ mg} \equiv 1\times10^{-3}/25{,}000 = 4\times10^{-8}\text{ mol}$$
$$\equiv 4\times10^{-8}\times6\times10^{23} = 2.4\times10^{16}\text{ molecules}$$

Worked Problem 1.2

Q In a 1 mg cm^{-3} solution of proteins with RMM 25,000, roughly how far apart are the molecules, on average?

A Volume per molecule = 1 $(cm^3)/2.4 \times 10^{16} = 4.2 \times 10^{-17}$ cm^3.

So each molecule might occupy a cube of side 3.5×10^{-6} cm (cube root of the volume), or 35 nm.

Worked Problem 1.3

Q How does the answer to Worked Problem 1.2 compare to the size of one 25,000 RMM molecule?

A Mass of 1 molecule = $25000/6 \times 10^{23} = 4.2 \times 10^{-20}$ g which corresponds to a molecular volume of around 4.2×10^{-20} cm^3, assuming a density similar to water. This corresponds to a cube of side 3.5 nm, approximately.

So, in a 1 mg cm^{-3} solution, these molecules are separated, on average, by about 10 molecular diameters.

Proteins function as enzymes (biological catalysts), antibodies, messengers, carriers, receptors, structural units, *etc.* Their chemical structure and molecular conformation are commonly described in terms of:

Primary structure: the sequence of amino acids in the polypeptide chain (see Figure 1.3). This is unique to each protein, and is determined (primarily) by the genetic information encoded in the DNA of the relevant gene.

```
KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR
STDYGIFQIN SRYWCNDGKT PGAVNACHCS ASALLQDNIA DAVACAKRVV
RDPQGIRAWV AWRNRCQNRD VRQYVQGCGV
```

Figure 1.3 Primary structure of a 130-residue protein (human lysozyme) shown using the single-letter amino acid codes

Secondary structure: regular, repeating structures such as α-helix, β-sheets, *etc.* (see Figure 1.4).

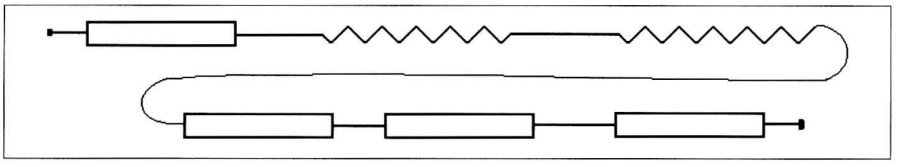

Figure 1.4 Secondary structure

Tertiary structure: the three-dimensional arrangement of secondary structure elements that defines the overall conformation of the (globular) protein (see Figure 1.5).

Quaternary structure: in multi-subunit proteins, the three-dimensional arrangement of the subunits (see Figure 1.6).

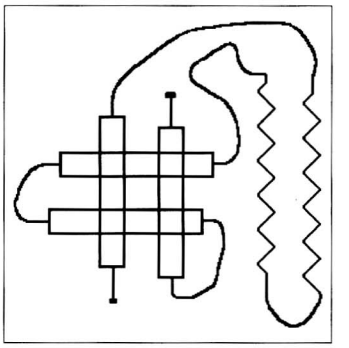

Figure 1.5 Tertiary structure

Haemoglobin, for example, is made up of four globular subunits – two of one kind (α) and two of another (β) – which combine to form a tetramer quaternary structure. Interaction between these subunits is responsible for the delicate control of oxygen uptake and release by the haem groups in this protein.

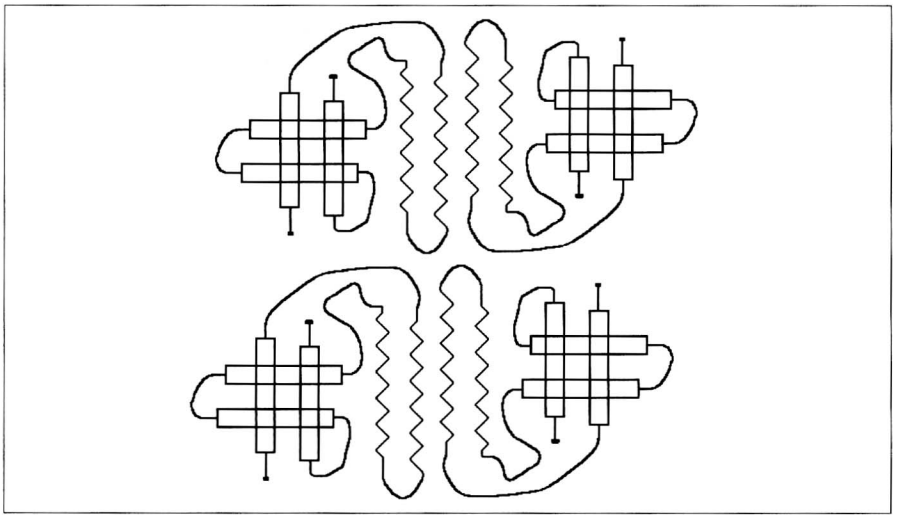

Figure 1.6 Quaternary structure

Because of rotational flexibility in the polypeptide backbone, primarily around the N–C_α (ϕ) and C_α–C (ψ) angles, there is a very large number of possible conformations that any one polypeptide molecule may adopt. Unlike most synthetic polymers, however, proteins have the ability to fold up (under the right conditions) into specific conformations, and it is these conformations (structures) that give rise to their individual properties.

Box 1.1 The "Protein Folding Problem"

Most proteins do not have a problem folding – they just do it. However, *we* have a problem understanding how they do it and predicting what the conformation of a particular amino acid sequence will be.

The complexity of the problem was highlighted some years ago by Levinthal,[1] a computer scientist who was one of the first to tackle the problem.

Each ϕ or ψ angle in a peptide might have roughly three possible values, giving $3 \times 3 = 9$ possible conformers for each peptide (not

You may be surprised to find that your calculator has trouble doing calculations such as 9^{100}. Why? How can you get around it?

counting side-chain conformers). For even a small polypeptide of 100 amino acids, this corresponds to at least $9^{100} \approx 3 \times 10^{95}$ possible different conformations of the polypeptide chain, only one of which (or a relatively small set) will be the "correct" one.

Assuming (optimistically) that peptide conformations can switch on the femtosecond time scale (10^{-15} s), it would take a time of order 3×10^{80} s, or about 10^{73} years, to search through all these possibilities to find the right one. This is a time much longer than the known age of the Universe. Yet proteins actually fold quite rapidly, in microseconds to minutes, depending on the protein and conditions. This is the so-called "Levinthal Paradox".

It is not really a paradox, of course. What it means is that polypeptides do not need to explore all possible conformations before they find the right one. Just as in any other rate process, there are kinetic pathways or reaction mechanisms that direct the system to the required state; and just as the water molecules in a mountain stream do not need to try all possible paths before finding they should flow downhill.

However, what Levinthal was pointing out was that if we do not know these pathways for protein folding, a computational search for the correct fold – no matter how powerful our computers – is doomed to failure.

Repetition of the same ϕ/ψ angles from one amino acid to the next gives rise to a regular secondary structure element, of which α-helix and β-sheet are the most common examples. In these structures the ϕ/ψ angles repeat in such a way that hydrogen bonds may form between different peptide groups to stabilize the structure.

The term "random coil" is sometimes used incorrectly to designate non-regular structural elements within a protein structure. There is, of course, nothing random about this: the ϕ/ψ angles are well defined.

Many structural elements such as loops, turns or other motifs that determine the tertiary structure of the protein do not have a regular repeating ϕ/ψ signature, but are nonetheless unique.

One important feature is that in samples of a particular protein (if pure and properly folded), all the molecules will have the same conformation, give or take a little bit of variation due to thermal fluctuation. This contrasts with the situation normally found in polymer chemistry, where the macromolecules rarely have a well-defined structure, and samples are made up of a heterogeneous mix of conformations, quite often in dynamic interconversion.

A true "random coil" is a hypothetical state in which the conformation (ϕ/ψ angles) of any one peptide group is totally uncorrelated with any other in the chain, especially its neighbours.

Folded proteins are relatively unstable, and can unfold ("denature") easily, especially with a change in temperature, pH or on addition of chemical denaturants such as urea, guanidine hydrochloride or alcohols. Denatured proteins have lost their tertiary and quaternary structure, but

may retain some secondary structure features. They rarely approach the true random coil state.

Unfolded protein is also quite sticky stuff, and has a tendency to aggregate with other denatured proteins or to stick to surfaces.

Traditional animal glues are made from denatured skin and bone. The main connective tissue protein, collagen, takes its name from the Greek word for glue.

This intrinsic stickiness of unfolded polypeptides appears to be one of the causes of prion diseases and other amyloid-related conditions such as mad cow disease, CJD, Alzheimer's, and so forth. In such conditions, unfolded or misfolded proteins aggregate into lumps or "plaques" that interfere with normal cell function.

1.3 Polynucleotides

The genetic information which encodes protein sequences is found in DNA (deoxyribonucleic acid), and the transcription and translation process involves RNA (ribonucleic acid). Both are polynucleotides consisting of long sequences of nucleic acids made up of a phosphoribose backbone, with a choice of four different purine or pyrimidine side-chains or "bases" attached (see Figures 1.7 and 1.8).

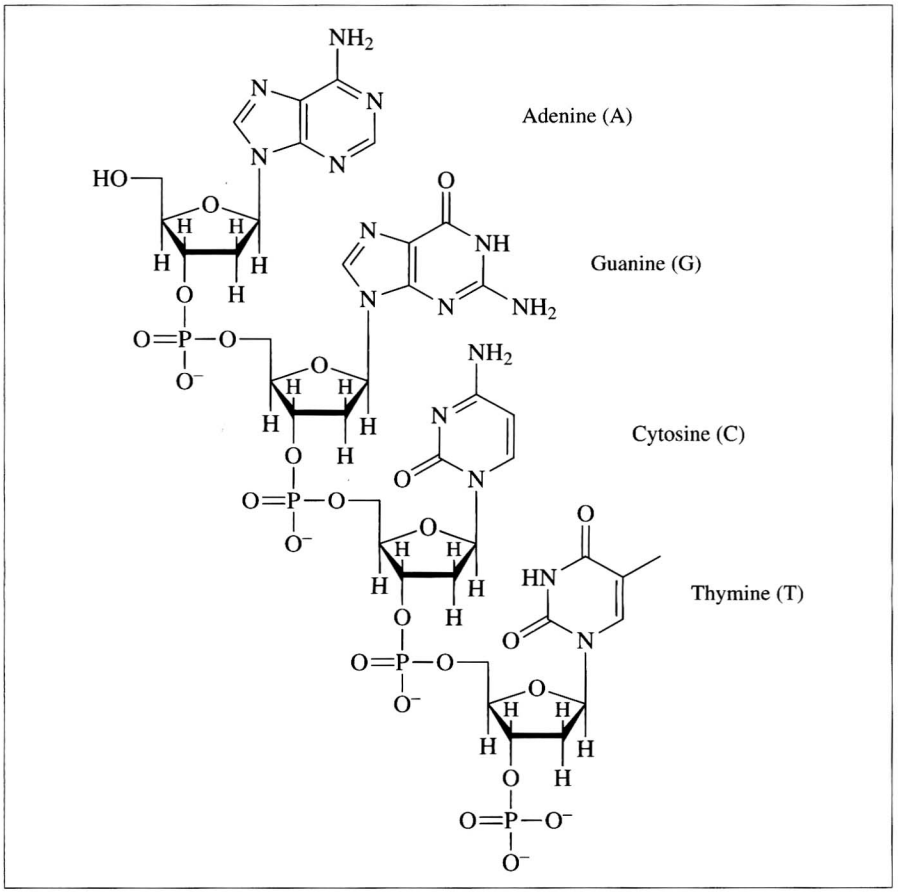

Figure 1.7 DNA structure illustrating the deoxyribose–phosphate backbone, to which may be attached purine (A, G) or pyrimidine (C, T) bases

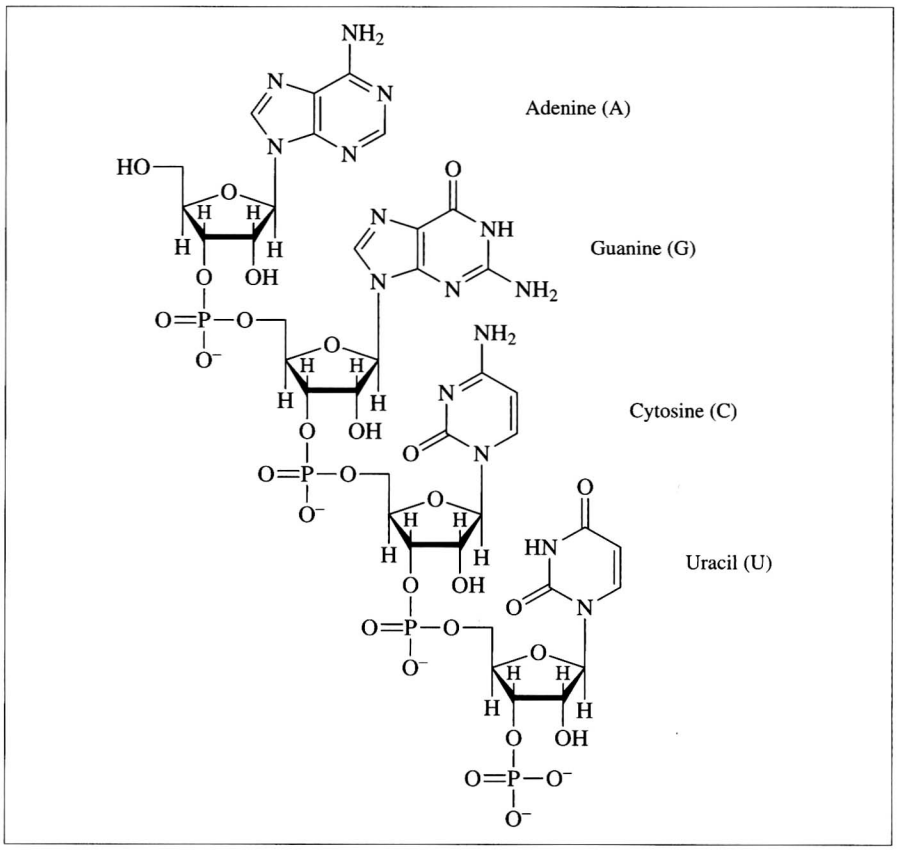

Figure 1.8 RNA structure illustrating the sugar (ribose)–phosphate backbone, to which may be attached purine (A, G) or pyrimidine (C, U) bases

The specific, complementary base pairing in the double-helical structures of DNA and RNA (Figure 1.9) is what gives rise to the ability to translate and proliferate this genetic information.

See *Biology for Chemists*[2] for molecular biology details.

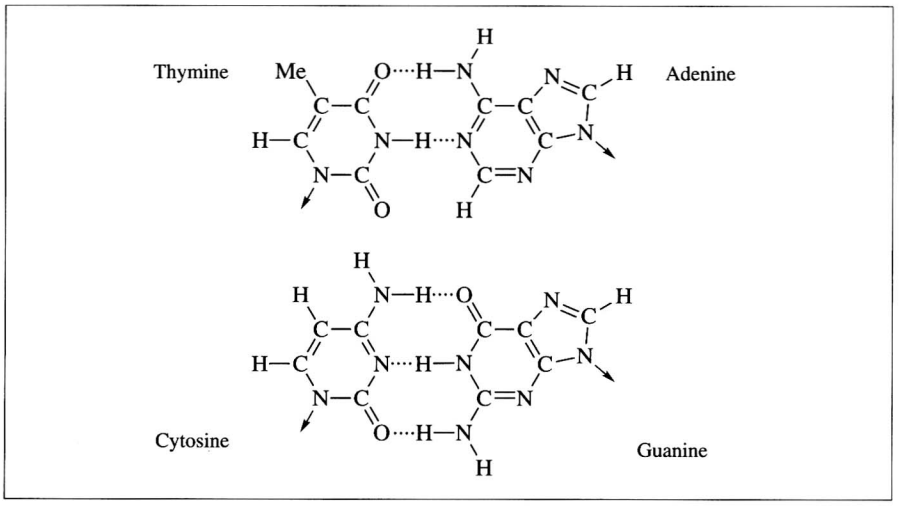

Figure 1.9 Complementary base pairing (Watson–Crick) in DNA (RNA is similar, with uracil replacing thymine)

When complementary strands of DNA and/or RNA come together, they form the characteristic right-handed double-helix structures that lie at the heart of molecular biology. In the most common form ("B-DNA") the base pairs stack in a twisted ladder-like conformation, with the purine/pyrimidine rings lying flat and perpendicular to the helix axis and spaced 0.34 nm apart. The negatively charged sugar–phosphate backbone lies to the outside of this cylindrical structure, which is roughly 2 nm in diameter.

Worked Problem 1.4

Q The DNA in each of your cells (*i.e.* the human genome) contains about 3×10^9 base pairs. How far would this stretch if laid out in a straight line?

A Assume 0.34 nm spacing. Then the distance is:

$$3 \times 10^9 \times 3.4 \times 10^{-10} = 1.02 \text{ m}$$

Many other polynucleotide conformations are possible, including the left-handed helical "Z-DNA" and more complicated structures thought to be involved in chain replication, together with supercoiling and more globular structures in single-stranded transfer RNA.

1.4 Polysaccharides

Complex polysaccharides such as starch, glycogen, cellulose, and so forth, play an important part in biochemistry, both as energy stores and structural components. Many proteins are glycosylated ("glycoproteins"), with oligosaccharide chains (often branched) attached to specific amino acid residues, usually at the protein surface. The carbohydrate portion of glycoproteins is often involved in antigenicity, cell receptor and other molecular recognition processes.

Reminder: **glycosylation** is the covalent attachment of carbohydrate (sugar) groups; an **oligosaccharide** is a short-chain polymer of sugars. Human blood groups are determined by the different oligosaccharides attached to glycoproteins and glycolipids on red blood cells.

Polysaccharides (and the smaller oligosaccharides) are polymers formed by linkage of individual sugar monomers, and may be linear (*e.g.* cellulose) or branched (*e.g.* glycogen).

Although some regular secondary structure is sometimes seen (for example, in cellulose fibres), the complexity of chemical composition and polymer chain branching leads to much more disordered structures (or, at least, structures that are usually too complex to determine). It is for this reason that our understanding of polysaccharide structures and their interactions is still very poor.