Applied Biophysics
A Molecular Approach for Physical Scientists

Tom A. Waigh
University of Manchester, Manchester, UK

John Wiley & Sons, Ltd

Applied Biophysics

Applied Biophysics
A Molecular Approach for Physical Scientists

Tom A. Waigh
University of Manchester, Manchester, UK

John Wiley & Sons, Ltd

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA
Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA
Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany
John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia
John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809
John Wiley & Sons Ltd, 6045 Freemont Blvd, Mississauga, Ontario L5R 4J3, Canada
Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data
Waigh, Tom A.
Applied biophysics : a molecular approach for physical scientists / Tom A. Waigh.
p. cm.
Includes index.
ISBN 978-0-470-01717-3 (alk. paper)
1. Biophysics. I. Title.
QH505.W35 2007
571.4–dc22 2007011017

British Library Cataloguing in Publication Data

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

ISBN 9780470017173 Cloth, 9780470017180 Paper

Typeset in 10.5/13 Sabon by Thomson Digital, India
Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Contents

Preface

The field of molecular biophysics is introduced in the following pages. The presentation focuses on the simple underlying concepts and demonstrates them using a series of up to date applications. It is hoped that the approach will appeal to physical scientists who are confronted with biological questions for the first time as they become involved in the current biotechnological revolution.

The field of biochemistry is vast and it is not the aim of this textbook to encompass the whole area. The book functions on a reductionist, nuts and bolts approach to the subject matter. It aims to explain the constructions and machinery of biological molecules very much as a civil engineer would examine the construction of a building or a mechanical engineer examine the dynamics of a turbine. Little or no recourse is taken to the chemical side of the subject, instead modern physical ideas are introduced to explain aspects of the phenomena that are confronted. These ideas provide an alternative, complementary set of tools to solve biophysical problems. It is thus hoped that the book will equip the reader with these new tools to approach the subject of biological physics.

A few rudimentary aspects of medical molecular biophysics are considered. In terms of the statistics of the cause of death, heart disease, cancer and Alzheimer's are some of the biggest issues that confront modern society. An introduction is made to the action of striated muscle (heart disease), DNA delivery for gene therapy (cancers and genetic diseases), and self-assembling protein aggregates (amyloid diseases such as Alzheimer's). These diseases are some of the major areas of medical research, and combined with food (agrochemical) and pharmaceutics, provide the major industrial motivation encouraging the development of molecular biophysics.

Please try to read some of the highlighted books, they will prove invaluable to bridge the gap between undergraduate studies and active areas of research science.

TOM WAIGH
Manchester, UK
February 2007

Acknowledgements

I would like to thank my family (Roger, Sally, Cathy, Paul, Bronwyn and Oliver) and friends for their help and support. The majority of the book was written in the physics department of the Universities of Manchester and Leeds. The PhD and undergraduate students (the umpa lumpas etc.) who weathered the initial course and the rough drafts of the lecture notes on which this book was based should be commended. I am indebted to the staff at the University of Edinburgh, the University of Cambridge and the Collège de France who helped educate me concerning the behaviour of soft condensed matter and molecular biophysics.

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

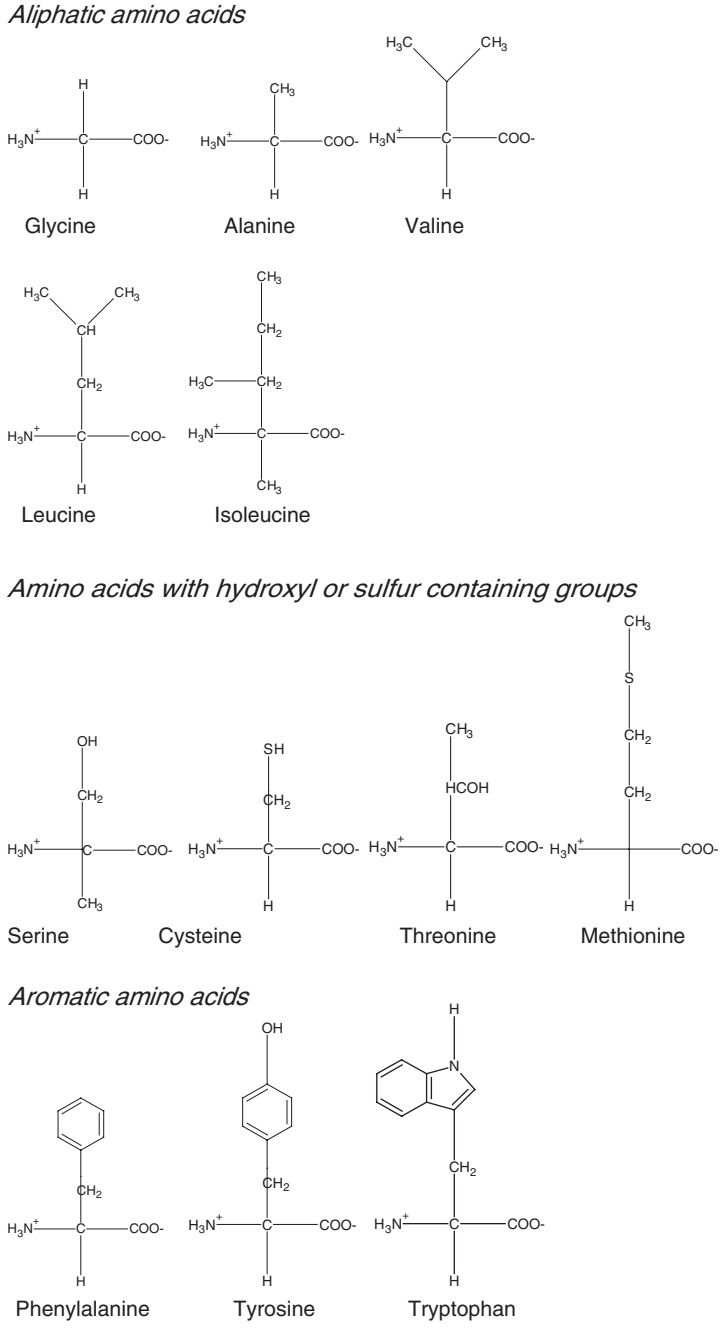
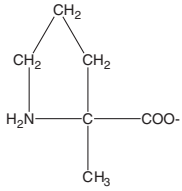


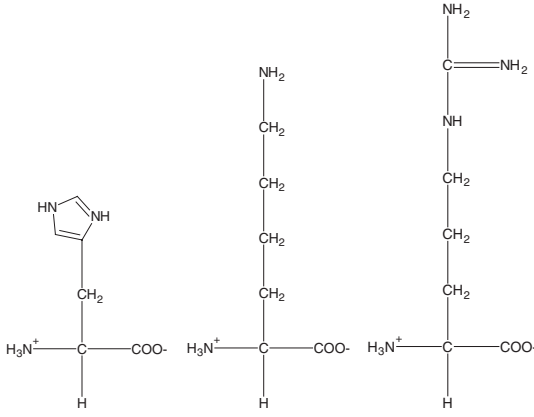
Figure 1.1 The chemical structure of the twenty amino acids found in nature

Cyclic amino acid



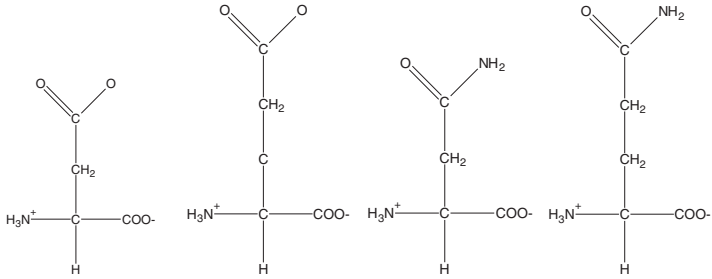
Proline

Basic amino acids



Histidine Lysine Arginine

Acidic amino acids and amides



Aspartic acid Glutamic acid Asparagine Glutamine

Figure 1.1 (*Continued*)

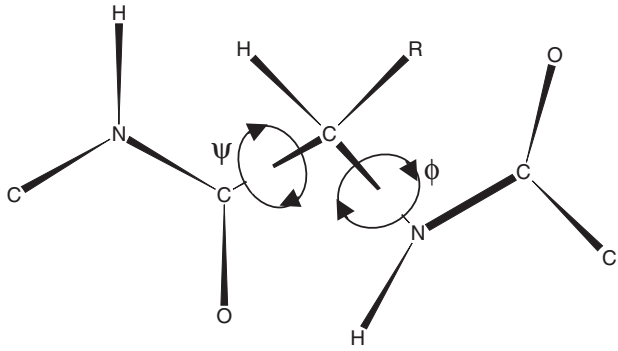


Figure 1.2 All amino acids have the same primitive structure and are connected with the same peptide linkage through C–C–N bonds
(O, N, C, H indicate oxygen, nitrogen, carbon and hydrogen atoms respectively. R is a pendant side-group which provides the amino acid with its identity, i.e. proline, glycine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

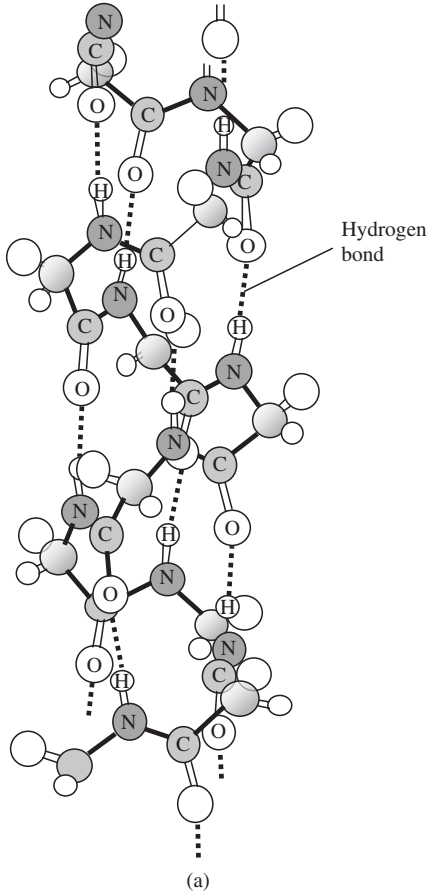
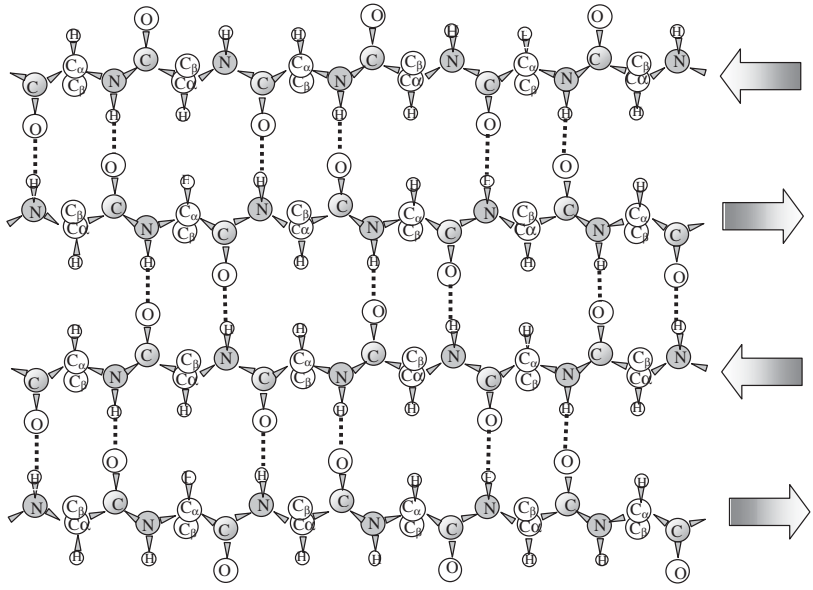


Figure 1.3 Simplified secondary structures of (a) an α-helix and (b) a β-sheet that commonly occur in proteins
(Hydrogen bonds are indicated by dotted lines.)



(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

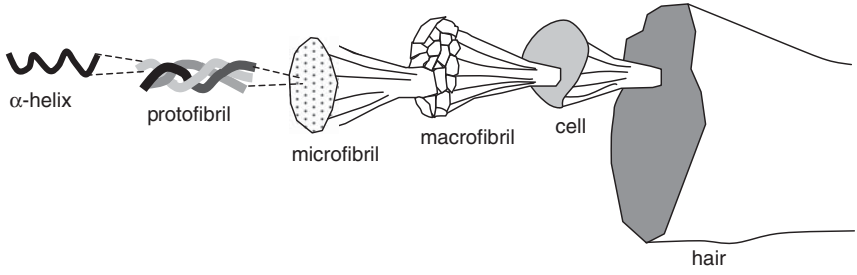


Figure 1.4 The complex hierarchical structures found in the keratins of hair (α-helices are combined in to protofibrils, then into microfibrils, macrofibrils, cells and finally in to a single hair fibre [*Reprinted with permission from J.Vincent, Structural Biomaterial, Copyright (1990) Princeton University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

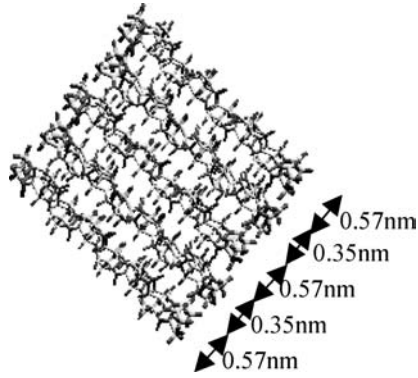


Figure 1.5 The packing of anti-parallel beta sheets found in silk proteins (Distances between the adjacent sheets are shown.)

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

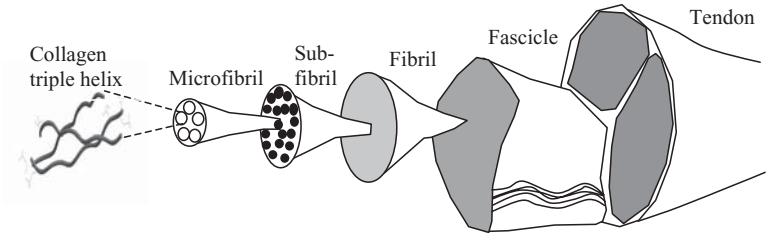


Figure 1.6 Hierarchical structure for the collagen triple helices in tendons (Collagen helices are combined into microfibrils, then into sub-fibrils, fibrils, fascicles and finally into tendons.)

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

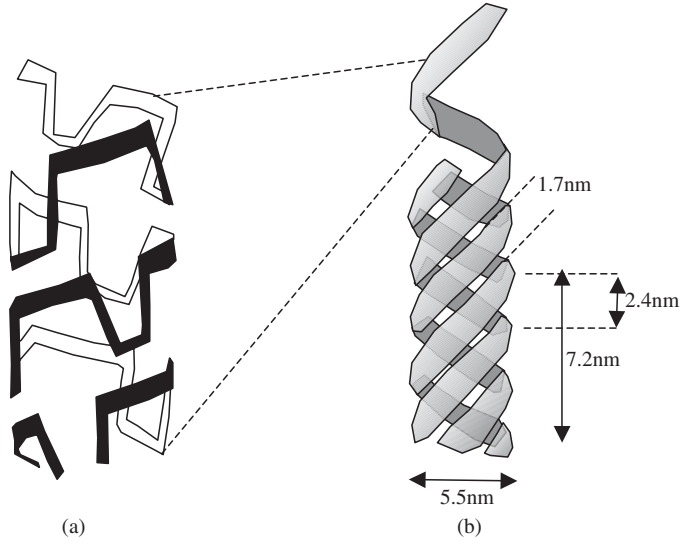


Figure 1.7 The β turns in elastin (a) form a secondary elastic helix which is subsequently assembled into a superhelical fibrous structure (b)

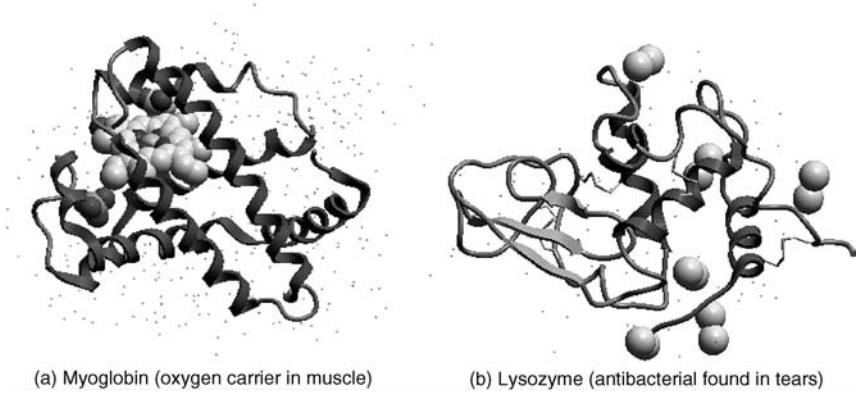


Figure 1.8 Two typical structures of globular proteins calculated using X-ray crystallography data

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

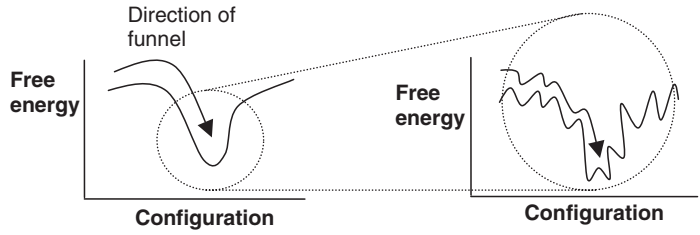


Figure 1.9 Schematic diagram indicating the funnel that guides the process of protein folding through the complex configuration space that contains many local minima. The funnel avoids the frustrated misfolded protein structures described in Levinthal's paradox

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.

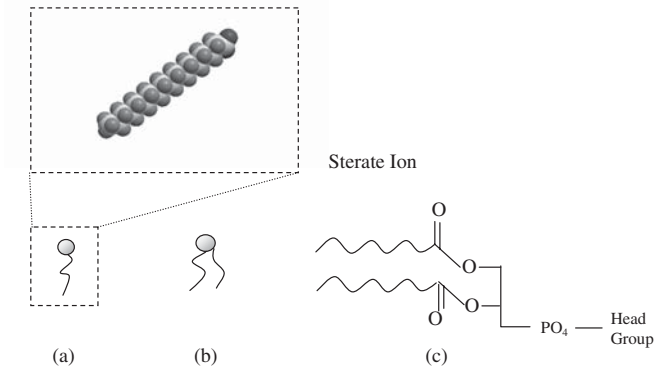


Figure 1.10 Range of lipid molecules typically encountered in biology
(a) fatty acids with one tail; (b) steroids and fatty acids with two tails; (c) phospholipids

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.

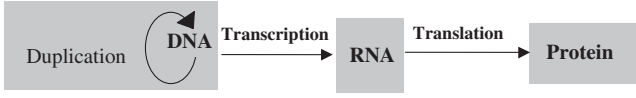


Figure 1.11 The central dogma of molecular biology considers the duplication and translation of DNA. DNA is duplicated from a DNA template. DNA is transcribed to form a RNA chain, and this information is translated into a protein sequence

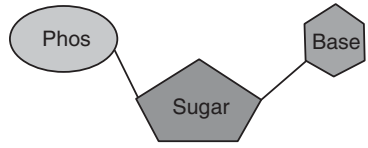


Figure 1.12 The chemical structure of the base of a nucleic acid consists of a phosphate group, a sugar and a base

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

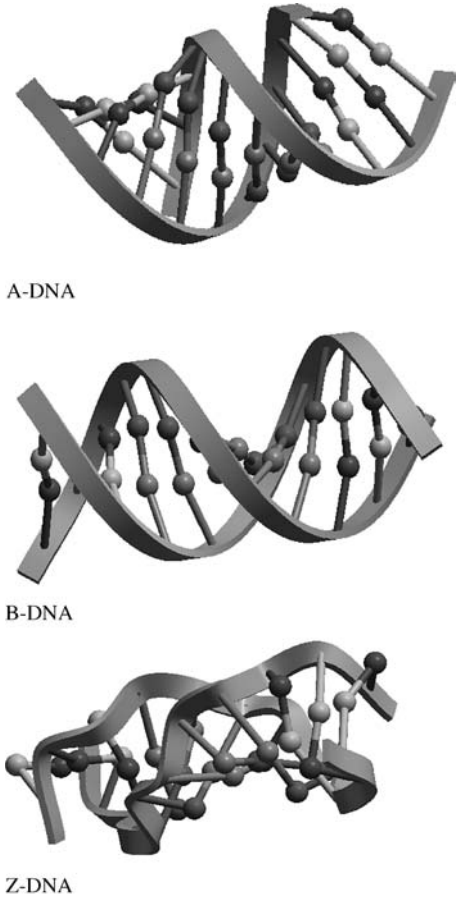


Figure 1.13 Molecular models of A, B and Z type double helical structures of DNA (A and B type helical structures, and their intermediates typically occur in biological systems. Z-DNA helical structures crystallise under extreme non-physiological conditions.)