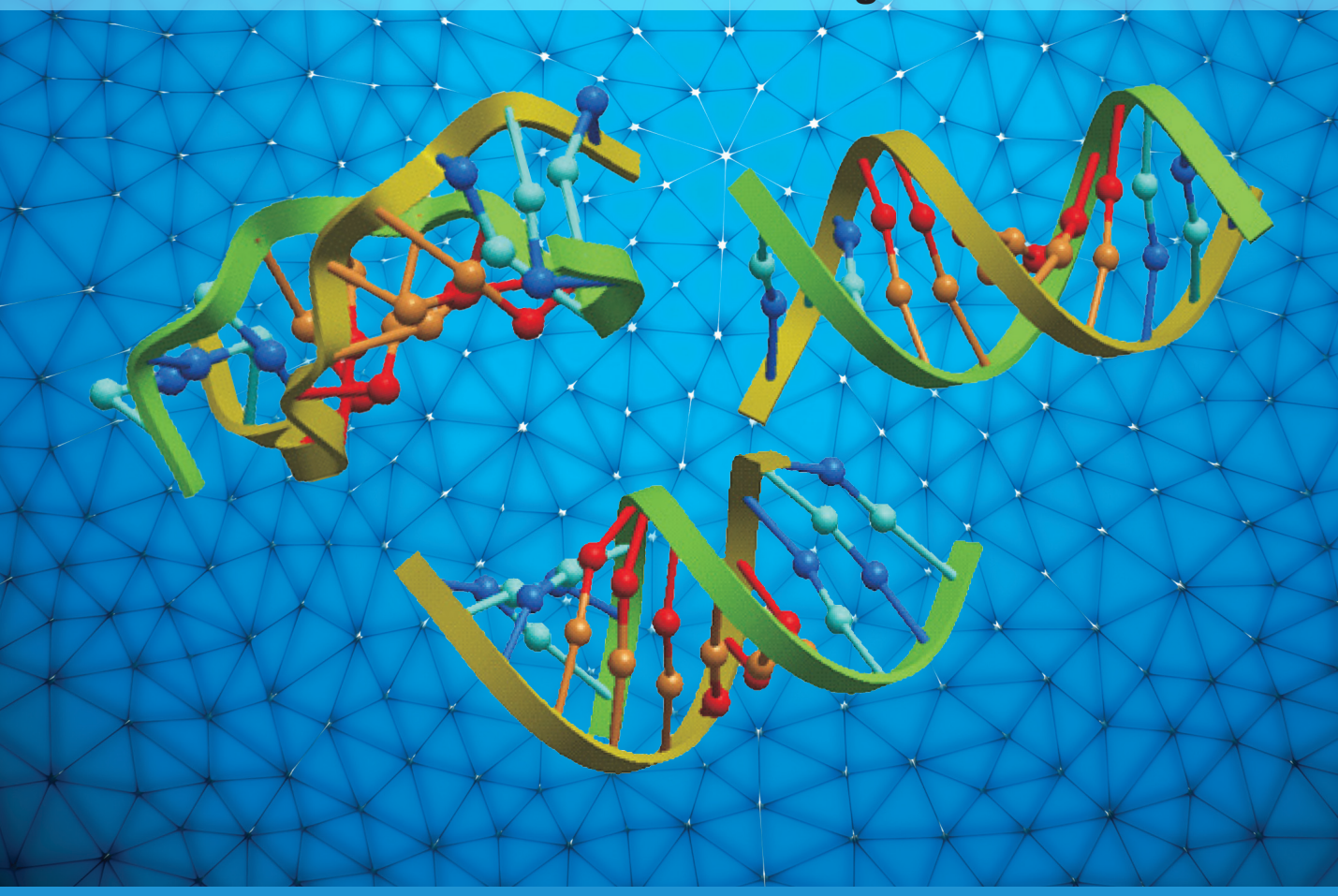
THE PHYSICS OF LIVING PROCESSES

A Mesoscopic Approach

Thomas Andrew Waigh



WILEY

The Physics of Living Processes

The Physics of Living Processes

A Mesoscopic Approach

THOMAS ANDREW WAIGH
School of Physics and Astronomy and Photon Science Institute,
University of Manchester, UK

WILEY

This edition first published 2014

Registered Office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication data applied for.

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

ISBN: 9781118449943

Set in 10/12pt Times by SPi Publisher Services, Pondicherry, India

Printed and bound in Singapore by C.O.S.

1 2014

Contents

Preface

This book is based around a two-semester course on biological physics that has been taught over the past few years at the University of Manchester. Students on the course are predominantly physics undergraduates who have good mathematics and physics foundations, but often lack any advanced biology or chemistry background. Therefore in a bid to make the course self-contained we teach the students the necessary biology, chemistry and physical chemistry as the course goes along. It is therefore hoped that anyone with a reasonably good high/secondary school background in both physics and mathematics can follow the material in this introductory course on the physics of living processes.

The book is divided into five principal sections: *building blocks*, *soft matter*, *experimental techniques*, *systems biology* and *spikes, brains and the senses*. The first section describes the basic building blocks used to construct living organisms. The next four sections introduce a series of useful tools for biological physicists to solve problems in the life sciences. This list of tools is not exhaustive, but it is hoped that they will introduce the reader to some of the possibilities.

The subheading of the book: 'A Mesoscopic Approach' is there to emphasize the range of length scales considered in the solution of biological problems. Here the mesoscale is taken to include the length scales from molecules to the microscale. Thus, for the most part the underlying quantum mechanical details are ignored, i.e. the approach is coarse grained and on the nano/micrometre scale. The quantised details of the molecules in biological problems are neglected in order to make them tractable, i.e. to provide an approximate solution in real time. Again this is a pragmatic approach due to time constraints and the unwieldy nature of current state-of-the-art *ab initio* quantum mechanics simulations. Some fascinating biologically relevant processes certainly do depend sensitively on quantum mechanical details for a satisfactory explanation. Photosynthesis and photodetection are classic examples, but for brevity these areas will be predominantly ignored and readers are directed to additional literature for self-study.

In the first three sections of the book the field of molecular biophysics will be introduced. The presentation will focus on the simple underlying concepts and demonstrate 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 due to the current biotechnological revolution.

The fields of biochemistry and cellular physiology are vast and it is not the aim of the current textbook to encompass the whole area. The first three sections of the book functions on a reductionist, nuts and bolts approach to the subject matter. They aim 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 recourse is taken to the specific chemical details of the subject, since these important areas are better treated in other dedicated biochemistry courses. 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.

The reductionist approach to molecular biology has historically been very successful. Organisms, drugs and foods can all be studied in terms of their constituent molecules. However, knowing all the musicians in an orchestra still leaves us none the wiser with respect to the music they are playing as an ensemble. The concerted motion and interaction of many thousands of different types of molecules gives rise to life. The pulsating

quivering motion of amoeba cells under a microscope, jet lag in bumble bees when they are moved between continents, the kaleidoscope of colours detected in the ape retina when we view a Van Gogh painting, the searing pain when we burn ourselves, and the miraculous healing processes when we have been cut; all these phenomena and many more require explanation. Often such studies have been labelled physiology and in the modern era predominantly have been studied by medics and vets, who often function in an engineering, the-organism-is-broken-how-do-we-fix-it, -type approach. Pharmacologists also study the effect of drugs on the metabolism of creatures, generally concentrating their interest in the development of new medicines, but again practically this takes a pragmatic engineering approach. What are the large-scale effects of the addition of a single chemical on the millions of biochemical processes upon which it could possibly have an impact? People do the experiments and hope they will discover a well-targeted drug (a magic bullet if you will), which will modify only a single faulty mechanism associated with a particular disease with limited side effects. Necessity (time constraints) requires that they ignore most of the holistic effects required for a complete understanding of a disease.

However, increasingly a quantitative understanding of the phenomena involved in living processes is required, returning physiology to the domain of the physical sciences*. This would allow us to make a rigorous connection between the structure and dynamics of biological molecules on the nanoscale and their concerted behaviour in living processes over time at larger lengths scales (at the scale of organs and organisms). At first sight this seems an impossible task due to the complexity of the phenomena involved (a huge intractable many-body problem), but a range of new tools are available in the twenty-first century to explore physiology that give us some hope; high-resolution noninvasive *experimental probes* now exist that will not disrupt or burn the specimens (magnetic resonance imaging, optical coherence tomography, ultrasound, fluorescence microscopy, positron emission tomography, and X-ray tomography, to name just a few); *postgenomic technology* (the human DNA genome has been sequenced, how do we make sense out of all this information?) holds a vast amount of information relevant to living processes that still needs to be properly mined; *network theory* provides mathematical tools to describe the geometry and connectivity of interacting components be they neurons, metabolic processes or individual biochemical products; *soft condensed-matter physics* demonstrates how the tools of conventional physics can be applied to the unusual behaviour of biological soft matter (e.g. statistical mechanics, fluid mechanics, elasticity theory, and novel model biomimetic materials); *systems biology* explores the robustness and diversity of biochemical processes in terms of the circuit diagrams of individual biochemical reactions; and *synthetic biology* allows cells to be completely reprogrammed to test the fundamental requirements for life. All these methods offer new approaches to solve physiological problems.

The discussion in this book is extended on from that previously presented in the textbook 'Applied Biophysics', also written by the current author, and it incorporates some of the same material. Applied Biophysics considered the application of soft-matter physics in molecular biology. The approach is now extended to the study of living processes with additional emphasis on modern deterministic themes concerning the behaviour of cells, action potentials and networks. These ideas and tools are applied to a series of problems in agriculture, medicine and pharmaceutical science. Many of these areas are currently being revolutionised and it is hoped that the text will provide a flavour of the fields that are being developed with relation to their biological physics and provide a bridge towards the relevant research literature.

The connections between the different themes discussed in the book need to be stressed, since they join together many of the different chapters e.g. signalling from **Chapter 23** and motility from **Chapter 7**. This integration of themes lends strength to quantitative descriptions of physiology in the last two sections of the book and highlights many important facets of a physiological question, e.g. how is insulin metabolised;

* The word *physiology* derives from the Greek word *Physis* (nature, origin) and *logos* (knowledge), so we see we have come full circle with the near tautology that is the physics of physiology.

what are the biochemical circuits? How is the motion of a mouse tracked by an owl; what is the activity of its neural networks? And how do the contractions of the heart give rise to the fluid mechanics of a pulse; what is the electrophysiology of the heart?

Much can be learnt from cells for would-be nanobiotechnologists. For example, currently we can make synthetic nanomotors, but switching them on or off when required is an ongoing challenge. Nature achieves this task with great speed and efficiency using virtuoso performances of ion channels and action potentials inside nerve cells. Physiology thus provides a treasure trove of processes that could be borrowed for synthetic biological designs.

Ethical questions are a concern when considering experiments with live organisms. These questions are worthy of careful thought and their solution requires a consensus amongst a broad community. The ethical consideration of such complex subjects as stem-cell research, cloning, genetic engineering and live animal studies should be considered hand in hand with the clear benefits they have for society [1]. Specifically it is a shock to an experimental physicist when samples arrive for experiments that are associated with an actual person: a sputum sample from individual X or a blood sample from individual Y. Thus, for ethical reasons it should be emphasised that, although experimental medical biological research can be fascinating, it needs real support from health-care professionals and should not be attempted independently (not to mention the legal implications)!

A few rudimentary aspects of medical molecular biophysics will be considered in **Parts I** and **II**. In terms of the statistics of the cause of death, heart disease, cancer and Alzheimer's disease are some of the biggest issues that confront modern society. An introduction will thus be made to the action of striated muscle (heart disease), DNA delivery for gene therapy (cancer is a genetic disease) and amyloid diseases (Alzheimer's). These diseases are major areas of medical research, and combined with food (agrochemical) and pharmaceutics provide the major industrial motivation encouraging the development of molecular biophysics. Physiological aspects of some of these industrial biotechnological challenges are also briefly examined towards the end of the text.

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

Further Reading

Reiss, M.J. & Straughan, R. (2001) *Improving Nature: The Science and Ethics of Genetic Engineering*, Cambridge University Press.

Readers can access PowerPoint slides of all figures at http://booksupport.wiley.com

Acknowledgements

I would like to thank the scientists and colleagues who have been an inspiration to me: Dame Athene Donald, Sir Sam Edwards, Eugene Terentjev, Wilson Poon, Peter Pusey, Mike Evans, Jian Lu, Henggui Zhang, David Quere, Claudine Williams, Pierre Giles de Gennes, Ian Ward, Ralph Colby, Rama Bansil, Andy Murray, Mark Dickinson, Tom Mullin, Helen Gleeson, Ingo Dierking, Emanuela di Cola, Ben Cowsil, Paul Coffey, Aris Papagiannopoulos, Stive Pregent, Sanjay Kharche, Mike Gidley, Pei-Hua, Tanniemola Liverpool, Peter Higgs, Andrew Turberfield, Stuart Clarke, Adrian Rennie, Richard Jones, Pantelis Georgiadis, Manlio Tassieri, Alex Malm, James Sanders, Dave Thornton, Viki Allan, Phil Woodman, Tim Hardingham, Alex Horsley, Shaden Jaradat, Matt Harvey, Gleb Yakubov, Paul Pudney, Andrew Harrison, David Kenwright, Peter Winlove, Mark Leake, John Trinick, Amalia Aggeli, and Neville Boden. Prof. Wouter den Otter should be thanked for corrections to 'Applied Biophysics' that were carried through to the current text.

I would also like to thank my family Sally, Roger, Cathy, Paul, Bronwyn, Oliver, Christina and last, but not least, my daughter Emily. The majority of this book was written in the physics department of the University of Manchester. The students and staff members who have weathered the ongoing teaching experiments should be commended. I am indebted to the staff at the University of Edinburgh, the University of Cambridge, the College de France and the University of Leeds for helping to educate me concerning the behaviour of soft condensed-matter and biological physics. The Institute of Physics should also be thanked for their support of the biological physics community in the UK through the creation of a dedicated biological physics group. The final stages of the preparation of the book's manuscript were completed at the Kavli Institute for Theoretical Physics in Santa Barbara, USA. The Institute's organizers should be thanked for making the author's sabbatical visit such a pleasant experience.

Part I
Building Blocks

Every life form discovered so far on the planet Earth follows the same blueprint. The genetic information is carried by nucleic acids that code for a small number of families of other carbon-based molecules. The actual number of possible molecules within these families is huge (in principle infinite due to polymerisation), but many of the general features of the molecules can be deduced by comparison with other similar molecules contained within the same family e.g. the classifications of *proteins*, *lipids*, *nucleic acids* or *carbohydrates*. Furthermore living organisms are constructed from cells that conform to two basic blue prints; *prokaryotes* and *eukaryotes*. The basic nanomachinery is held in common between all members of either of these two cellular forms, although specialisations are possible (human cells come in over 200 varieties). Thus, if you are presented with a new unknown organism on the planet Earth you can rapidly make some educated guesses about how it works, how its cells function and what molecules it is made of. Provided, that is, you have been given a little appropriate training.

A study of the building blocks in cells is also very satisfying with respect to the nanomachinery that life has on offer. Thousands of tiny molecular machines are in action in every cell, elegantly optimised by evolution. There is thus much to be learnt for people building new devices on the nanoscale through the inspection of naturally occurring cellular machinery.

It is hoped that a strong case has been made that time spent learning about cells and biomolecules is extremely well spent. Furthermore, much of the most useful generic information can be acquired relatively quickly, although complete mastery of even a small area of biochemistry can require a lifetime's work. This section of the course is normally greeted with a little scepticism by physics undergraduates, since it is more descriptive in style than they are familiar with, and it is the most biologically and chemically demanding section of the course. However it is surprising the mileage that results from learning a small number of key facts, and it is of course impossible to discuss any biological problems whatsoever without naming the key players; *molecules* and *cells*.

The Physics of Living Processes: A Mesoscopic Approach, First Edition. Thomas Andrew Waigh.

Suggested Reading

Goodsell, D.S. (2010) *The Machinery of Life*, Springer. A simple discursive introduction to biochemistry with some attractive illustrations. The book includes a gallery of pen and ink drawings of different biomolecules. I particularly liked the perspective of the congested environment inside cells.

Goodsell, D.S. (2004) *Bionanotechnology: Lessons from Nature*, Wiley-Blackwell. Another excellent book from Goodsell, similar in style to *The Machinery of Life*.

1

Molecules

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 will therefore be covered. The aim of this chapter is to give the reader 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 in a wide range of cellular processes.

Cells are predominantly composed of water that is structured and organised by inorganic ions and carbon containing (organic) molecules. The extracellular matrix in organisms can, in addition, contain solid crystalline composites such as calcium carbonate and silicates that form bones and exoskeletons. However most of the processes vital for life occur in aqueous solutions, although they are typically highly congested, with a huge variety of competing molecular nanoparticles present. How robust well-regulated living processes occur in such congested environments is still a matter of ongoing research.

There are four main classes of organic macromolecules inside cells: these are the *lipids*, *proteins*, *carbohydrates* and *nucleic acids*. Also, mixtures are possible such as *glycolipids* (carbohydrates fused to lipids) and *glycoproteins* (carbohydrates fused to proteins).

The subject of the molecular structure of materials will be first approached and how this arises from the underlying quantum mechanics. Then, the concept of chirality will be introduced for molecules, cells and organisms. Finally, a rapid tour will be made of the main classes of biological molecules that occur in living cells.

1.1 Chemical Bonds and Molecular Interactions

Ernest Rutherford described the atom as a minute nucleus surrounded by a huge expansive cloud of electrons; 'a mosquito in a cathedral'. Niels Bohr extended this picture, since only certain energy states are permitted by the quantum principle of Max Planck (1900) and de Broglie's wave/particle duality. Electrons are confined to discrete shells as they orbit the nucleus (**Figure 1.1**). Subsequently, Erwin Schrödinger showed how to calculate the energies and the spatial distributions of these electronic orbitals.

The Physics of Living Processes: A Mesoscopic Approach, First Edition. Thomas Andrew Waigh.

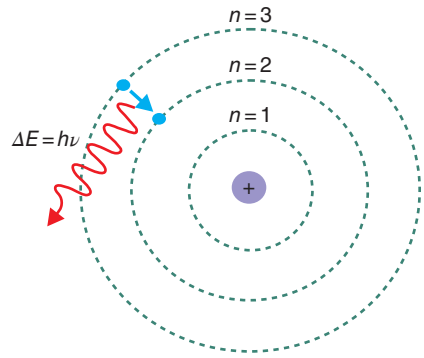


Figure 1.1 *In the Bohr model of an atom electrons move around the heavy nucleus, at a characteristic distance ~0.1 nm. The nucleus acts as a 'mosquito' in the huge empty volume of the electronic cathedral. Transitions between different energy levels (principal quantum number,* n*) can be detected by the emission or absorption of quantised photons (energy* $\Delta E = h\nu$*,* h *is Planck's constant and* ν *is the frequency.)*

Exact calculations of electron wave functions that use the Schrödinger equation will be left to more specialised quantum mechanics courses, as too will be more accurate quantum electrodynamics calculations, which are currently the most accurate theories to simulate the behaviour of atoms and molecules. Both approaches with complex biomolecules tend to be horrendously difficult and require extensive computational power. However, a few more details of quantum mechanics are useful for the development of an intuitive picture of molecular structure and dynamics, and will thus be introduced here.

Max Born postulated a radical reinterpretation of the quantum theory; different quantum states of electron distributions in the vicinity of an atomic nucleus are characterised by different probability distributions. These probability distributions are the fundamental quantities predicted by theory, and intrinsic uncertainties in their values are written in to the laws of nature (*Heisenberg's uncertainty principle*). The Schrödinger equation predicts the probability that an electron is found at a certain position. The full information about a bound electron inside an atom is now thought to be described by four quantum numbers; n (the principle quantum number related to the energy), l (the total angular momentum), m (the z component of the angular momentum) and s (the angular spin momentum). Based on these four quantum numbers, combined with the symmetry of the electrons' wave functions, Pauli deduced his exclusion principle, which is that 'no two electrons can be associated with the same nucleus and have precisely the same values of all four of the quantum numbers'. This principle stops matter collapsing in on itself, since electrons repel each other to form shells such that each electron has a different permutation of the four quantum numbers. This stability is reflected in the repulsive hard-core interactions experienced by all atoms at short distances, i.e. the excluded volume potential.

The quantum theory also explains the formation of molecules. Neighbouring atoms share or transfer electrons to create more energetically stable quantised electronic orbital structures. The geometries of atomic orbitals are classified as s (spherical), p (double lobed), d (double lobe threading a doughnut) and f (double lobe threading two doughnuts), **Figure 1.2**. Molecular orbitals tend to be even more complicated, since they require hybridisation of neighbouring atomic orbitals.

The prediction of chemical bonding patterns tends to be a job for the intuition of a good synthetic chemist with many years of experience, or an extremely hard *ab initio* quantum mechanical calculation for a computer. Chemical bonding can be classified under the broad (and often overlapping) headings of *ionic*, *covalent*, *metallic* and *hydrogen*.

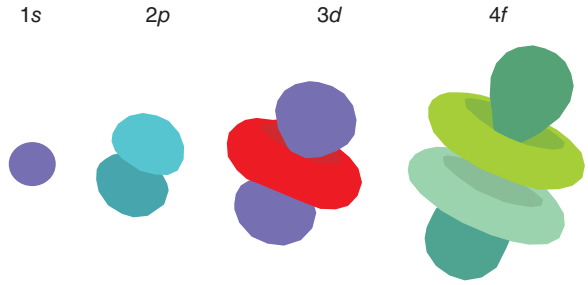


Figure 1.2 *Schematic diagram of some varieties of probability distribution geometries found in the electronic orbitals of atoms. 1*s *spherical, 2*p *double lobed, 3*d *double lobed threading a doughnut, and 4*f *double lobed threading two doughnuts.*

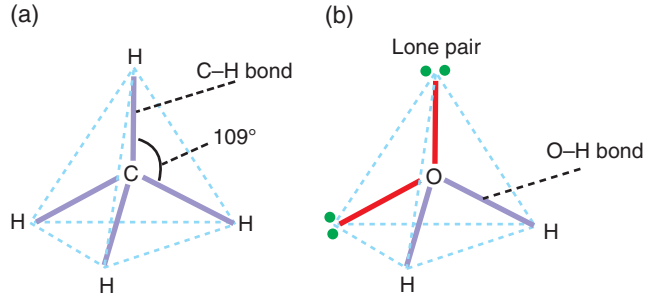


Figure 1.3 *Schematic diagram that shows a comparison of the molecular bonding geometries. (a) Methane (CH_4) is symmetrical with a constant 109° angle between C–H bonds. (b) Water is asymmetrical with a 104.5° between the H–O–H bonds.*

Ionic bonding occurs when atoms ionise to form electrolytes with the classic example being table salt, NaCl. An ionisation energy is associated with the movement of an electron from one atom to another, so their electronic structures are more energetically favourable.

When the difference in electronegativity is small between two atoms, they may form molecules through *covalent bonding*, e.g. H_2, HF, H_2O or *metallic bonding*, e.g. Na, K, Fe. Covalent bonding requires electrons to be shared and they are closely localised between the two bonded atoms. Metallic bonding involves delocalised electronic wave functions that allow rapid mobility of electrons through the crystalline lattices.

The development of a good intuition for the relationship between molecular structures and chemical formulae requires years of experience, but it is instructive to ask a simple question on tetrahedral bonding to start to build some understanding. It is useful to ask why the bonding pattern in methane (CH_4) is symmetrical while water (H_2O) is not (**Figure 1.3**). The answer is found in the quantum mechanics, because carbon has two fewer electrons than oxygen. In water 2*s* and 2*p* orbitals of oxygen hybridise (four orbitals in total); two electrons bond with two Hs, and the other two electrons become lone pairs. The lone pairs are negatively charged and attract the H atoms from neighbouring water molecules (*hydrogen bonding*). Water thus has a distorted tetrahedral structure due to the different interactions between the lone pairs and the hybridised orbitals. In contrast methane's carbon atom forms covalent bonds with four hydrogen atoms in a perfectly symmetrical tetrahedral structure.

Molecules assemble together to form different phases of matter (gas, liquid, solid, etc.), but retain their individual identities at the atomic scale. The exact phase adopted is determined by intermolecular forces between the molecules and they are weaker than the intramolecular forces already considered (ionic bonding, covalent bonding, etc.).

The most common intermolecular force is that of van der Waals. This can be thought of as a default interaction that occurs between all molecules and dominates the interactions if all the other forces are switched off, e.g. with liquid helium at low temperatures. In fact van der Waals forces correspond to a family of three or more types of force. *Debye*, *Keesom*, and *van der Waals* interactions are the principle subclassifications, but higher-order multipolar interactions also occur (**Chapter 4**). In general, van der Waals forces arise from instantaneous stochastic dipole moments associated with the motions of individual electrons. The instantaneous distribution of electrons can influence those of surrounding atoms, and the net outcome can be a weakly positive attraction. In biology, this force can explain the miraculous manner in which geckos can crawl up windows and flies can stick to ceilings. It is, however, not a universal mechanism of adhesion, e.g. tropical frogs use capillary forces to hold themselves onto surfaces, which in turn depends on the humidity.

Unlike *covalent bonding*, the van der Waals force does not have a unique direction, it cannot be saturated (the number of atoms or molecules involved is based on geometrical conditions, not on the electronic structure of the orbitals), it depends on the sample geometry and requires quantum electrodynamics (QED) calculations for a quantitative treatment. Although the derivations are complicated, analytic QED solutions do exist for standard geometries, e.g. see Adrian Parsegian's book, 'van der Waals forces in biology', for more details (also **Chapter 4**).

The importance of *hydrogen bonding* in biology cannot be more emphasised (**Figure 1.4**, **Section 1.7**). All living organisms contain a large amount of water, and water has its unique properties due to hydrogen bonding. For example, water expands when it freezes, which helped life's origins in the sea immeasurably. Water is also important for the self-assembly of many biomolecules. Indeed from a certain perspective, humans are predominantly structured self-assembled water. There is a nice analogy with a wobbly children's party jelly that is 98% water and 2% protein, but appears (at the low frequencies people are familiar with) to be a solid. The water component in human cells is slightly lower at ~70% w/w, but the jelly analogy helps the development

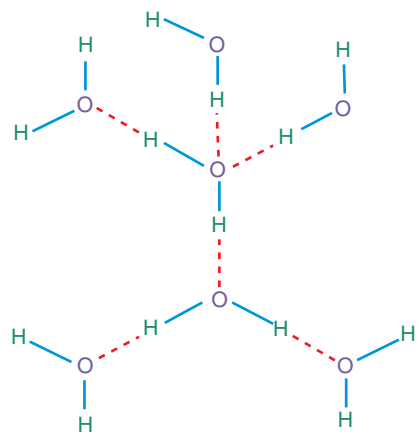


Figure 1.4 *Schematic diagram of the network structure formed by water molecules in liquid water. Dashed red lines indicate hydrogen bonds. Chains of hydrogen-bonded water molecules occur over a wide range of angles for liquid water, but become more restricted in ice crystals.*

of an intuition of the cell's varied dynamic properties at different time scales (its viscoelasticity). At long time scales cells appear to be solid, whereas at short time scales they are liquid-like.

Hydrogen bonding in water has a number of subtle secondary interactions that are important in the determination of its effects. One phenomenon is the hydrophobic effect that is primarily due to entropy. Consider a fat chain surrounded by water molecules. The normal hydrogen-bonded network in liquid water is disrupted and the water molecules become structured due to the loss of mobility as they orient themselves away from the fat chains. There is consequently a penalty in the free-energy term due to the loss of entropy ($-TS$ in $F = U - TS$, the Helmholtz free energy, **Section 3.4**). Entropy will be considered in more detail in **Chapter 3**. In a hand-waving manner, entropy is a measure of the randomness of a system, or equivalently the number of accessible states to a system. The hydrophobic effect is the molecular origin of why oil and water do not mix on the macroscale.

In general, all the forces between different biomolecules, subcellular compartments, cells, organisms and nanoscaled particles are of interest to biological physics. These interactions lead to relatively weak forces compared to chemical bonds, but help explain important physical and biological phenomena such as phase transitions and self-assembly. Such coarse-grained interactions over nanometer length scales are called mesoscopic forces, and will be considered in more detail in **Chapter 4**.

1.2 Chirality

Chirality is a symmetry operation in which molecules do not superpose with their mirror images (**Figure 1.5**), e.g. right-handed B DNA molecules (that exist naturally) do not superpose with left-handed B DNA molecules (an artificial construct). The occurrence of chirality can be determined for many organic molecules directly by inspection of their molecular structure. A carbon atom with four different groups attached to it can immediately be deduced to be chiral (**Figure 1.5c**) and this chirality could in turn affect the macroscopic behaviour of the material. All amino acids have chiral centres, except glycine, and so too do the nucleic acids, DNA and RNA, many carbohydrates and many lipids. This implies a fundamental molecular cause for the macroscopic chirality observed in biological materials, although spontaneous chiral symmetry breaking is also possible (in this case both chiralities occur on average in equal amounts).

Chiral interactions between molecules often dramatically perturb the phase of matter formed, e.g. the crystalline or liquid-crystalline phase adopted will become twisted with characteristic orientational defect structures (**Chapter 6**). Furthermore, chiral phases of matter often have unusual optical properties when they interact with photons (used commercially in liquid-crystalline displays with synthetic molecules). Cilial chirality in human embryos is thought to lead to the partial left/right asymmetry of the body parts in humans, e.g. the heart is on the left in most individuals.

1.3 Proteins

Polymers consist of a large number of identical subunits (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 1.6**) that can function as monomers and all the monomers are connected together with identical peptide linkages (C–N bonds, **Figure 1.7**). Only twenty amino acids occur in nature that are used to create proteins in eukaryotic cells, although three additional protein forming amino acids occur in bacteria and over 140 nonprotein-forming amino acids are known. A particularly persuasive evolutionary argument of why only twenty amino acids exist in proteins in eukaryotic cells is still lacking. The best partial explanation is that twenty is enough, and reflects the common evolutionary origins of all life on the planet Earth. Synthetic chemists have made new amino acids

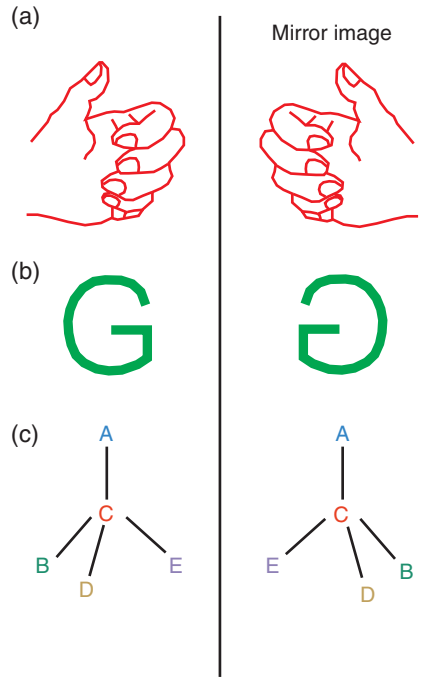


Figure 1.5 Chirality *('handedness') occurs when an object cannot be superposed with its mirror image. (a) A human hand and (c) a carbon molecule with four different substituents are* chiral *objects. (b) A flat letter 'G' is* achiral *in three dimensions, since it can be lifted up and overlayed with its mirror image. (c) Such chiral carbon atoms are found in many organic molecules and give rise to their macroscopic chiral phases, e.g. amino acids in the α helices of proteins can form twisted cholesteric liquid-crystalline phases. The letters A, B, D, and E denote arbitrary distinct chemical substituents, whereas C is a carbon atom.*

and connected them together to form novel synthetic proteins, so there is in principle no clear chemical barrier that restricts the number of possibilities (which is explored in the field of synthetic biology), but natural life only uses twenty (or twenty three if bacterial life is included in the list).

The twenty eukaryotic protein forming 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 making them water loving: serine and threonine. Three amino acids all have very polar positively charged side chains: lysine, arginine and histidine. Two amino acids form a family with acidic negatively charged side groups and they are joined by two corresponding neutral counterparts that have a similar chemistry: aspartate, glutamate, asparagine, and glutamine. More generally, the protein forming amino acids can be separated into three principle families; *hydrophobic* (they hate water), *polar* (they like water) and *charged* (they like water and are charged when they are incorporated into proteins).

The linkages between amino acids all have the same chemistry and basic geometry (**Figure 1.7**), which greatly simplifies their classification on the atomic scale. The peptide linkage that connects all amino acids together consists of a carbon atom attached to a nitrogen atom through a single covalent bond. The condensation of two α-amino acids to form a dipeptide is shown in **Figure 1.8**. There is only one way that

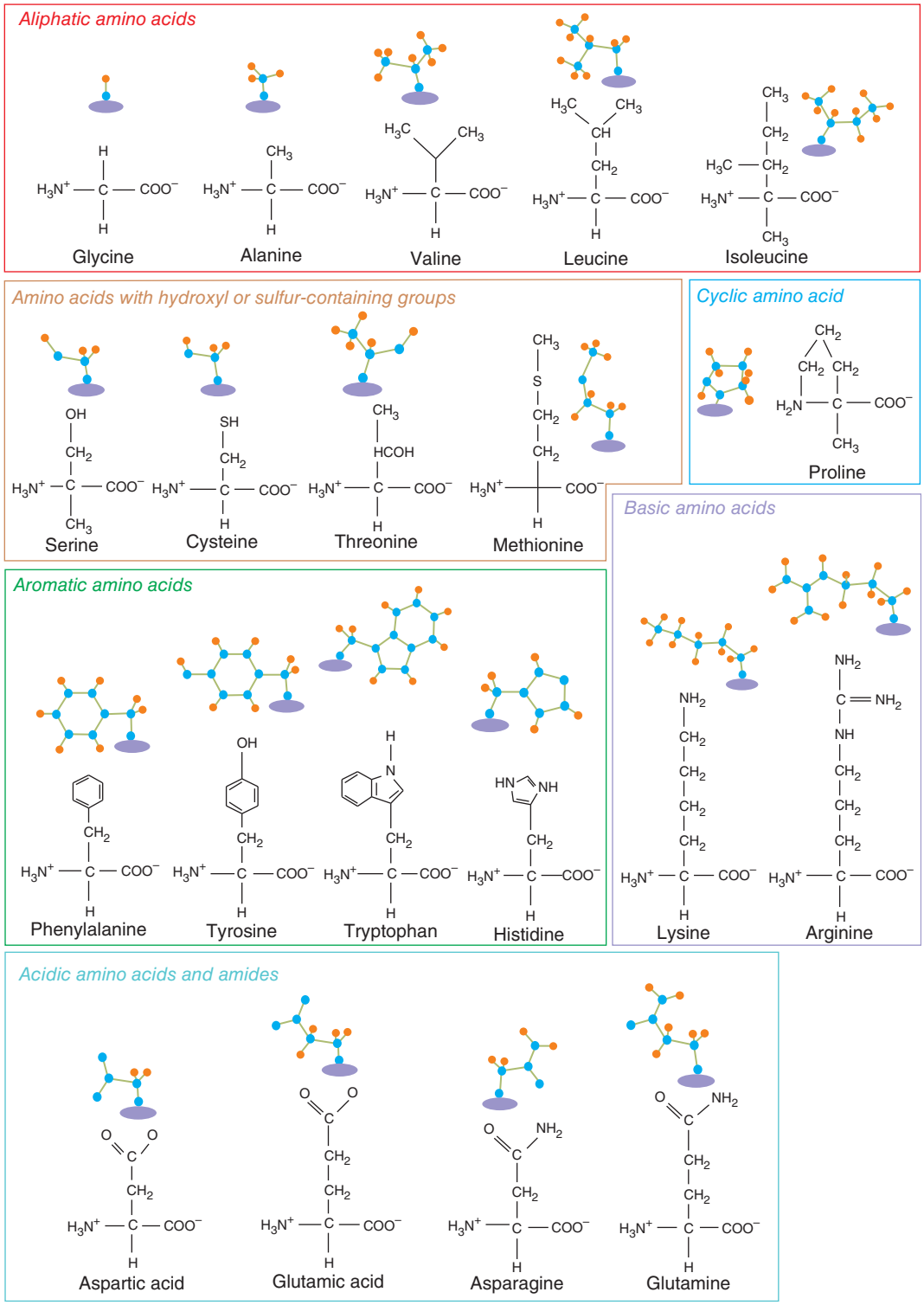


Figure 1.6 *The chemical structure of the twenty amino acids that form proteins in eukaryotic cells.*

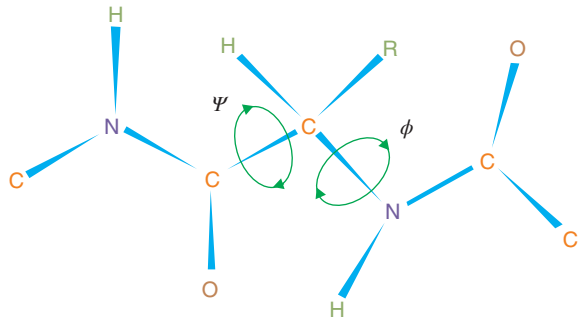


Figure 1.7 *All of the amino acids have the same primitive structure and are connected with the same peptide linkage through C–N bonds (O, N, C, and H indicate oxygen, nitrogen, carbon and hydrogen respectively). R is a pendant side group that provides the amino acid with its identity, i.e. proline, glycine, etc (**Figure 1.6**). Defining the rotational angles (Ψ, ϕ) for each amino acid gives a reasonably compact description of the peptide backbone conformation (it leads to Ramachrandran maps of protein conformation).*

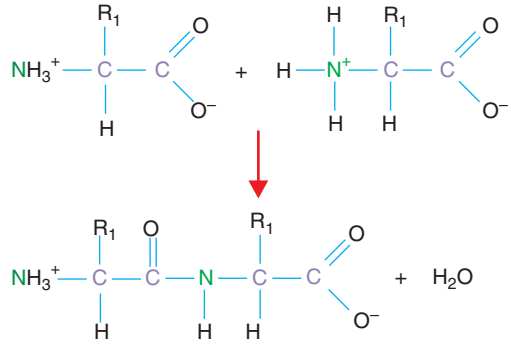


Figure 1.8 *A condensation reaction between two peptides can create a peptide linkage and a water molecule.*

two amino acids can be connected in proteins, the *peptide linkage*, and the chemical formula for four amino acids connected in line is

$$\text{-CR}_1\text{HCONH-CR}_2\text{HCONH-CR}_3\text{HCONH-CR}_4\text{HCONH-}$$

where the Rs are the groups that differentiate the 20 different amino acids and the hyphens are the peptide linkages. The peptide linkage has a directionality and in general -A-B- is usually different from -B-A-. Peptides are thus conventionally written with the N-ending peptide (the unbound amine terminus) on the left and the C-ending peptide (the unbound carboxyl terminus) on the right.

Although the chemistry of peptide linkages is fairly simple, to relate the primary sequence of amino acids (the combination of amino acids along the chain) to the resultant three-dimensional structure of the proteins is very complicated and predominantly remains an unsolved problem. To describe protein structure in more detail it is useful to describe motifs of secondary structure that occur in their morphology. The motifs include *alpha helices*, *beta sheets* and *beta barrels* (**Figure 1.9**). The first beta sheet and alpha helical structures were suggested by William Astbury during the 1930s. He also proposed that adjacent proteins were held together in hair and wool by hydrogen bonds. The first refined models of protein structure required more quantitative