This page intentionally left blank

Preface

> We also now appreciate that molecular biology is not a trivial aspect of biological systems. It is at the heart of the matter. Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself. All approaches at a higher level are suspect until confirmed at the molecular level.
>
> Francis Crick, *What Mad Pursuit*, 1988

This is a book about DNA, the most central substance in the workings of all life on Earth. It is a book about the way in which DNA works at a molecular level. We have used the title *Understanding DNA*… because our subject has now reached the stage where many aspects of it are well enough understood for us to be able to give a clear and uncluttered presentation of the main ideas. But we shall not disguise the fact that there is still a great deal which is not known or understood.

The book can be read at two different levels. First, it can be taken as an easy-to-read textbook for undergraduate or graduate students of chemistry and biology at university. Second, it may be read by ordinary people who have no prior knowledge of biochemistry, but who want to understand something of the fundamental processes of life. The sort of people we have in mind here are those who have learned something about DNA from popular magazines, newspapers, and TV programs. They know, for example, that DNA contains the 'genes' of classical genetics – those units of inheritance which pass on characteristics such as red hair or a long nose from parent to child, or even crippling diseases such as sickle-cell anemia or thalassemia. They probably also know that DNA is a long molecule, like a computer tape – the tape which tells our bodies how to grow and how to digest food and (perhaps) how to behave. And they may even know, if they are into quiz games and the like, that the initials 'DNA' stand for 'Deoxyribo-Nucleic Acid,' a certain kind of acid found in the cell nucleus, which was first identified over 100 years ago. People like this, who are curious to know more, will be able to learn a lot from this book about how DNA performs its tasks in our bodies at a molecular level.

This third edition of the book comprises 11 chapters. Chapter 1 is a general introduction to molecular biology: it is aimed at the non-specialist reader, and so it may be passed over by a student who already knows some biology. Chapters 2, 3 and 4 give some lessons about various aspects of the molecular structure of DNA, such as why it is helical, and how it can bend around proteins; this is basic material, which is nevertheless not yet available in other textbooks. Chapters 5 and 6 discuss the three-dimensional structure of DNA at a higher level. These chapters include some mathematics and geometry that may be unfamiliar to non-specialists and biology students; but we take care to present the key ideas by means of clear diagrams wherever possible. Chapter 7 gives an overview of the organization of chromosomes, which are large particles that contain both protein and DNA: there the DNA wraps about the protein into several different levels of structure. Chapter 8 discusses the mechanism of 'direct reading' of DNA sequences by proteins: this is an area that has expanded greatly since the first edition appeared in 1992. Chapter 9 explains the various experimental techniques which scientists use to study DNA. Chapter 10 describes the way in which DNA techniques are increasingly being used in medicine; while Chapter 11, which is new to this edition, summarizes the fast-growing area of cytosine methylation and DNA epigenetics. We end with a Postscript on what we have left out, followed by three Appendices on matters too detailed for the main part of the book. At the end of most chapters we give a bibliography of works to which we have referred in the text; and we provide some further reading for the student, and also some pointers to web-based resources. We have substantially updated the reference lists for this third edition. We have also supplied a few exercises at the end of most chapters.

We have made many changes from the two earlier editions apart from those mentioned above. Thus, we have updated the text and figures where necessary, particularly in Chapters 7 to 10 and the Appendices: the two new members of our author team, B.F.L. and A.A.T., have played a major part in these revisions. As in the earlier editions, we have tried to write in plain English, with minimum use of jargon which might confuse the reader.

Understanding DNA: the molecule & how it works should be suitable as a small text to accompany the very large, general textbooks which are now used widely in university biochemistry courses; or else it may be employed as a main text for a course specializing in DNA structure, provided the students have a background in biology and are willing to pursue more detailed readings in the scientific literature, as suggested. Or, of course, it may just be read as a book.

Many friends and colleagues have helped us greatly in various ways in the preparation of this book. We are grateful to Nick Cozzarelli, Mustafa El Hassan, Malcolm Ferguson-Smith, John Finch, Robert Henderson, Ron Hill, Chris Hunter, Maxine McCall, Garth Nicholson, Dinshaw Patel, Tim Richmond, Masashi Suzuki, David Tremethick, Takeshi Urayama and Sue Whytock for providing photographs and diagrams etc. which we have used; to Dick Dickerson for giving us data on X-ray structures; to Aaron Klug, John Melki, Kiyoshi Nagai, Daniela Rhodes, Deidre Scadden, Chris Smith, Jean Thomas and Michael Waring for commenting freely on various drafts of the manuscript of the previous and the present edition; and to the late Julian Wells for encouraging us in the first place to write a book on DNA. We are grateful to Japan Graphers and Kyoritsu Shuppan Co., Ltd (publishers of the Japanese version of our first edition) for the chapter opening icons, and to Elliott Stollar for help with the cover picture. Tessa Picknett has been a constant source of editorial advice and encouragement. The work of Caryn Wilkinson in revising and adding to the manuscript disk, and of Dennis Halls in updating and making more diagrams, has been beyond praise. Lastly we thank our respective wives, Mary, Maxine, Sandra and Carrie for their help of many kinds over the years; and we dedicate this new edition to them with gratitude.

C.R.C.	H.R.D.	B.F.L.	A.A.T.
Cambridge	Sydney	Cambridge	Cambridge

This page intentionally left blank

CHAPTER 1

An Introduction to Molecular Biology for Non-Scientists

One day two of us were having lunch together at a Cambridge College. We got into a conversation with one of our neighbors at the table, who was a senior historian. After a while he asked us what we did, and we explained that we were scientists, working with the very tiny molecules of biology. Then he said, 'I don't see how you do it'.

'Do what?'

'Work all the time with things that you can't see'.

You see, even people of great intelligence and learning, who spend their lives gathering evidence and pondering it deeply, nevertheless think in ways very different from those of modern biologists; it is hard for them to imagine what atoms and molecules look like. How hard will it be, then, for the beginning student to do the same?

For this reason it is necessary for us to start this book by explaining carefully about the *sizes* of things. A single DNA molecule is too small to be seen by eye. But if you have a big clump of lots of DNA molecules together, then the substance becomes visible, and appears as a clean, white, stringy, and viscous mass, somewhat like molasses sugar. Yet, you *can* see single DNA molecules by using special equipment involving X-rays, or an electron microscope, or an atomic force microscope; and we shall show some pictures of individual DNA molecules later in the chapter.

But first, in order to gain an intuitive feeling for the microscopic world, let us compare the size of DNA to the sizes of things in general, and especially to the sizes of other things that are too small to be seen by the naked eye. Figure 1.1 shows a scale of typical sizes. It is a logarithmic scale, and each division represents a factor of 10. The scale (on the left) covers 10 orders of magnitude from 1 m down to 0.000 000 000 1 or 10^{-10} m. Near the top we have the largest objects

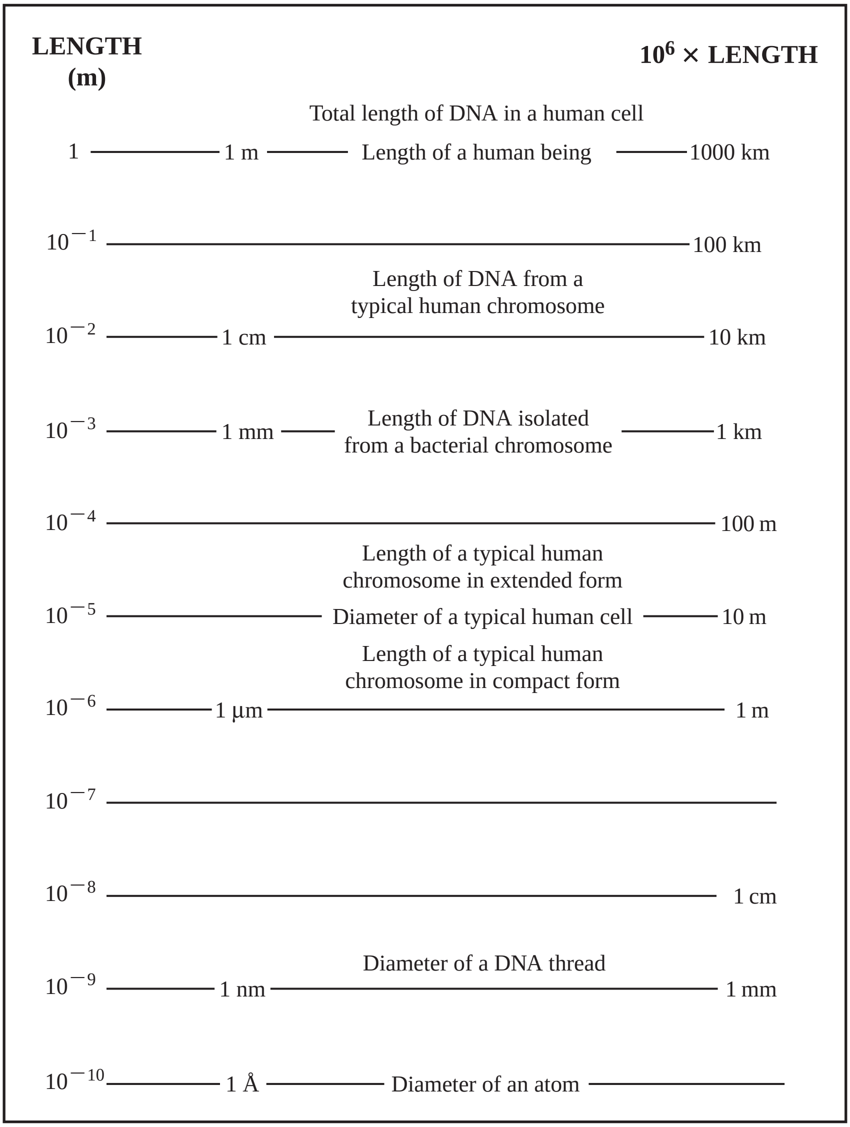

Figure 1.1 The relative lengths of things on a microscopic scale.

that we shall be thinking about: human beings are roughly 1 m long, as an order of magnitude. At the bottom of the scale are the smallest objects that we shall be concerned with: atoms, which are typically of diameter 10^{-10} m (or 1 Ångstrom unit, Å). Exactly halfway between these two extremes, on the present kind of scale, we have the diameter of a typical human cell at about 10^{-5} m or 10 µm.

Some things are larger than a cell on our scale, while others are smaller. It is perhaps surprising that the length of DNA isolated in

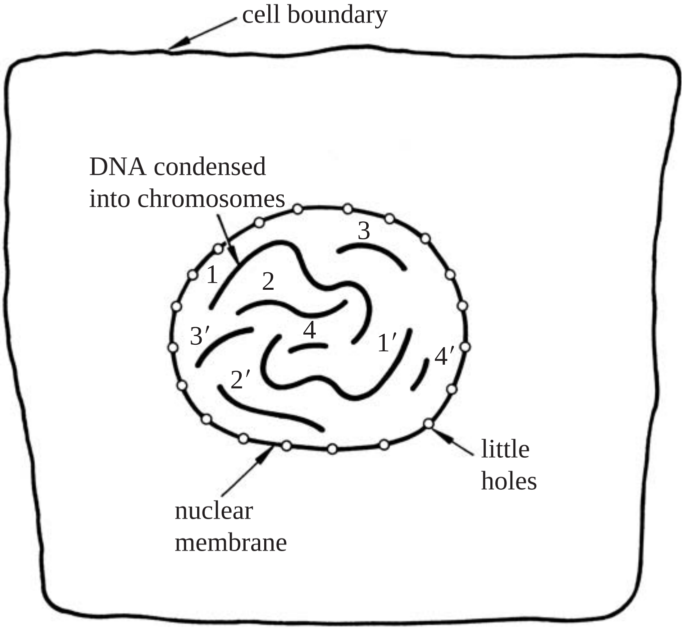

Figure 1.2 Schematic picture of a typical cell from higher organisms. The chromosomes, which contain a mixture of protein and DNA, come in homologous pairs. This picture shows a hypothetical cell with four such pairs that can be distinguished by their sizes. They are labeled 1,1′; 2,2′; etc.

pure, thread-like form from a single human chromosome[1] (3 cm, or 3×10^{-2} m) or from a single bacterium (1 mm, or 10^{-3} m) can be so much longer than the cell from which it came (10^{-5} m). But this illustrates a very important point: the DNA is compacted in length by a factor of as much as 10 000 when it is embedded in a living cell. DNA is a very narrow thread of diameter of just 2×10^{-9} m, and although the DNA from a single human cell has a total length of 2 m, it could conceivably be compacted into a tight ball, like a ball of string, of diameter 2×10^{-6} m. But Nature chooses to pack the DNA into a form somewhat less dense than this, at a length of about 3×10^{-6} m for a single, compact chromosome. After all, if DNA was packed too tightly into a cell, then the information along its length, known as 'genes'[1] (to be discussed below), would probably become inaccessible.

Our bodies are made from billions of individual cells, and DNA is the control center of each and every cell. This DNA is something like a computer tape that stores many programs for a large computer to run. It is present in each cell in the form of a number of chromosomes. Chromosomes are arranged in pairs (see Fig. 1.2), and the

1 See note on p.15

two members of a pair are nearly identical or 'homologous'[1] copies of one another, just like the back-up disks on a computer: one copy saves the functional program if the other copy becomes defective.

The mass of DNA is surrounded in most cells by a strong membrane with tiny, selective holes, which allow some things to go in and out, but keep others either inside or outside. Important chemical molecules go in and out of these holes, like memos from the main office of a factory to its workshops; and indeed the individual cell is in many ways like an entire factory on a very tiny scale. The space in the cell which is not occupied by DNA and the various sorts of machinery is filled with water.

On the right of Fig. 1.1 is a second scale of lengths which would apply if we were to enlarge every linear dimension by a factor of 1 000 000 or 10^6. It can be useful sometimes to make an imaginary enlargement of this sort. When we do so, the relative sizes and proportions of objects remain the same, of course. Note that the length of DNA from a typical chromosome on this expanded scale is about 30 km, while its diameter is just 2 mm. Very few objects in the physical world are so long and so narrow.

Now that we have gained a general idea of the relative sizes of things, and of the importance of DNA in the control of a living cell, let us see what DNA and chromosomes really look like.

Human chromosomes become compact and squat when cells are about to divide, and they can be seen easily by the use of a light microscope. The human chromosomes shown in Fig. 1.3 have been sorted and arranged by size into pairs. Each chromosome, as in Fig. 1.2, has duplicated itself in preparation for cell division, and has then reduced its length 10-fold, so that the duplicate copies can separate from one another without tangling as the cell divides. One half of each X-shaped duplicate chromosome will go to each new cell.

Many more chromosomes are shown here than in the simple, schematic drawing of Fig. 1.2. Human cells contain 46 chromosomes in 22 homologous pairs (numbered 1 to 22) plus the non-homologous X and Y chromosomes that determine sex. All animals and plants have chromosomes that look like these, but in different numbers: for example, a fruit fly has eight chromosomes, in three homologous pairs plus X and Y. Spreads of chromosomes, such as those shown in Fig. 1.3, are very useful for medical purposes; for example, to check the health of an unborn child while still in the womb.

At other stages in the life of a cell, far from cell division, the chromosomes are generally more extended and less condensed, and so they cannot be seen by use of a light microscope. That is why we had to draw the single chromosomes of Fig. 1.2 in schematic form, because no true pictures of such chromosomes exist. Nevertheless,

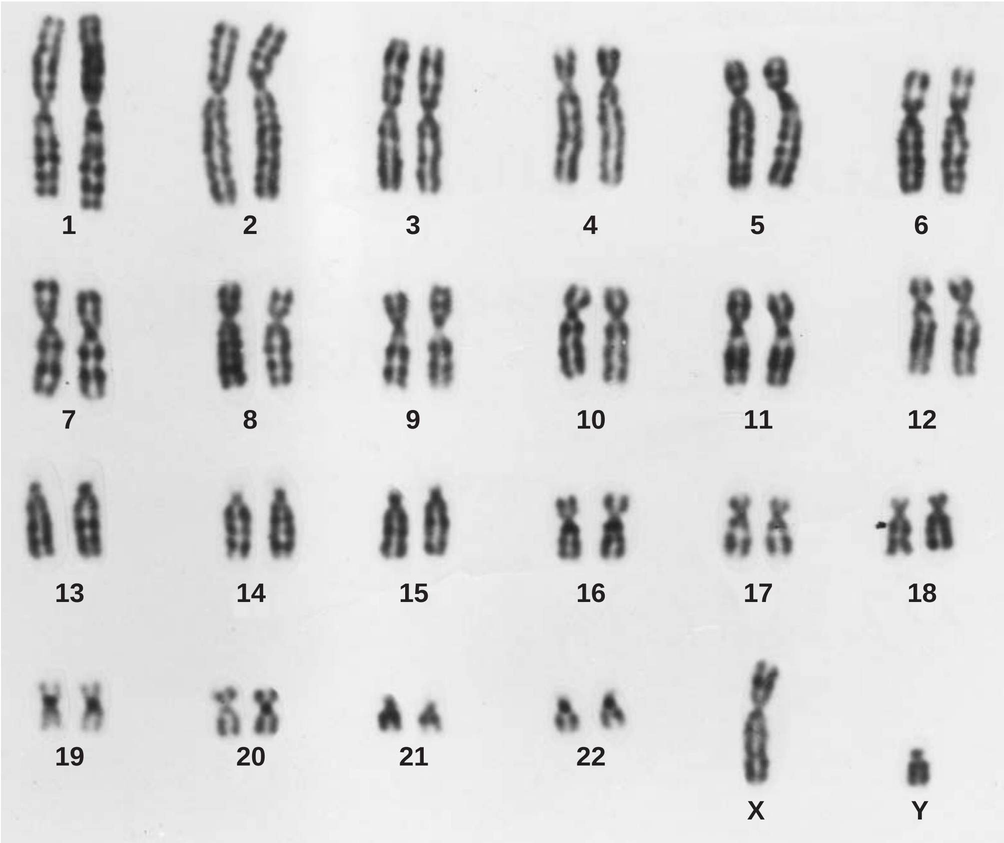

Figure 1.3 Photographs of human chromosomes in duplicate, as isolated just before cell division (at metaphase[1]) and then sorted by length into pairs. Each number identifies the two chromosomes of a homologous pair. X and Y are non-homologous chromosomes that determine a person's sex as female (XX) or male (XY). The two duplicate copies of any individual chromosome form an 'X' shape because they have not yet separated entirely. Scale: chromosomes 3 are approximately 10^{-5} m or 10 μm long. Courtesy of Malcolm Ferguson-Smith.

in a few particular tissues of the fruit fly, these single-copy, extended chromosomes happen to duplicate themselves about 1000 times over, without becoming compact. They eventually contain 1000 identical DNA molecules laid side-by-side, in parallel register. These 'polytene' chromosomes are just like the ones shown in Fig. 1.2, but they are much wider. Due to their greatly increased size, these monstrous, amplified fly chromosomes can be seen clearly by the use of a light microscope, as shown in Fig. 1.4 at different magnifications.

In Fig. 1.4(a) the extended, DNA-containing chromosomes can be seen along with the tight membrane in which they are wrapped; the entire assembly is known as the 'nucleus'.[1] Here the chromosomes