Judson, H.F. (1979) *The Eighth Day of Creation*. Simon & Schuster, New York. A very thorough history of how molecular biology became a separate branch of science.

McCarty, M. (1994) A retrospective look at the discovery of the genetic role of DNA. *The FASEB Journal* **8**, 889–90. A first-hand account of how it was discovered that DNA is the substance of genes.

Yusopova, G.Zh., Yusupov, M.M., Cate, J.H., and Noller, H.F. (2001) Path of messenger RNA through the ribosome. *Cell* **106**, 233–41. A high-resolution picture by X-ray methods of messenger-RNA bound within a protein-synthesis factory or ribosome.

Zhimulev, I.F. (1996) Morphology and structure of polytene chromosomes. *Advances in Genetics* **34**, 1–490. A comprehensive summary of information about giant 'polytene' chromosomes of flies.

Exercises

1.1 Every human cell, in the non-dividing state, contains a total of about 6×10^9 base-pairs of DNA. The diameter of a typical human cell is $10\,\mu m = 10^5$ Å.

- **a** By treating the DNA as a cylinder of length 3.3 Å per base-pair, calculate the total length of DNA in any cell. Compare this to the diameter of the cell.
- **b** By treating the DNA as a cylinder of radius 10 Å, calculate the total volume of DNA in any cell. Compare this to the total volume of the cell, if it is assumed to be spherical. (Volume of cylinder $= \pi r^2 l$, where r is the radius and l the length; volume of sphere $= \frac{4}{3}\pi R^3$, where R is the radius.)
- **c** Consider a typical chromosome, which contains 1/46 of the total DNA of the cell, on average. Find the total length of the DNA in this chromosome, and then the diameter of a solid sphere into which that volume of DNA could be compacted, in principle. Compare the diameter of this compact sphere with the mean length of an actual, metaphase chromosome, given that the DNA length–compaction ratio for such a chromosome is about 10 000.

1.2a Use Table 1.1 to give the sequence of amino acids in a protein chain which is coded by the following base sequence:

GCCAAGCAACTCATTCAAGGT
1 2 3

Start reading at base 1.

- **b** Now repeat the process by beginning to read at base 2; and then again by beginning to read at base 3. (Observe that the

amino-acid sequence of a protein chain depends critically on which 'reading frame' is used in the DNA.)

1.3a An extra base G is now inserted between bases C and T at positions 10 and 11 of the DNA sequence given in Exercise 1.2. Starting at base 1, read out the new sequence of amino acids.

b The sequence given in Exercise 1.2 is now altered instead by the deletion of C at position 10. Starting at base 1, read out the sequence of amino acids. (These are both known as 'frame-shift' mutations.)

1.4a Translate the following DNA base sequence into a sequence of amino acids for a protein molecule:

ACGCTATGTCACATGGTACCTAACGTAT

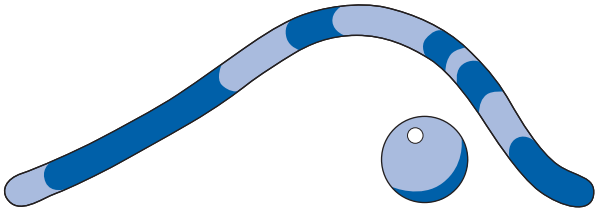
On this occasion, do not begin reading at position 1, but rather in accordance with the true, biologically used start-scheme described in Table 1.1; and do not read beyond a STOP triplet.

b Search the base sequence given below for potential STOP triplets. What amino-acid sequence will be assembled, starting as in a?

GCTCATGGTCATTCGTAACAGTTAGGCCATGACCG

CHAPTER 2

Why a Helix?



It is crucially important, not only in biology but in all fields of science, to understand the inner workings of Nature as well as its external form. For example, consider the funny-looking object shown from two perspectives in Fig. 2.1. It has five different colors, but what *is* it? How would you go about making such an object? A poor scientist would study the object superficially, and give a name to each particular feature, such as the intersection of five edges; he or she would be deeply concerned with measuring the angles between edges on the surface of the object, in order to look for some sort of pattern. But a more perceptive scientist would stop and think: what is the internal structure of the thing?

Now look at the object again, but this time focus on only one color, say yellow: behold, the yellow parts form a cube in space, tilted somewhat on its side. Each of the other four colors also describes a cube. In fact, this toy is called 'The Compound of Five Cubes'.

Spotting the five distinct cubes – it was helpful of the makers to use five different colors – is the first clue to understanding the internal structure of the object. This clue leads us to think about matters of geometrical symmetry; and if we were to follow this line of mathematical thinking, we would eventually understand why there are five cubes rather than, say, six or four. Indeed, we might then be in a position to make other, kindred objects by using the same underlying, structural principles.

This example illustrates an important point in science. By perceiving the internal structure of an object, you learn much about Nature; but by studying only its external form, especially in great detail and with high concern for nomenclature, you learn relatively little that is worthwhile.

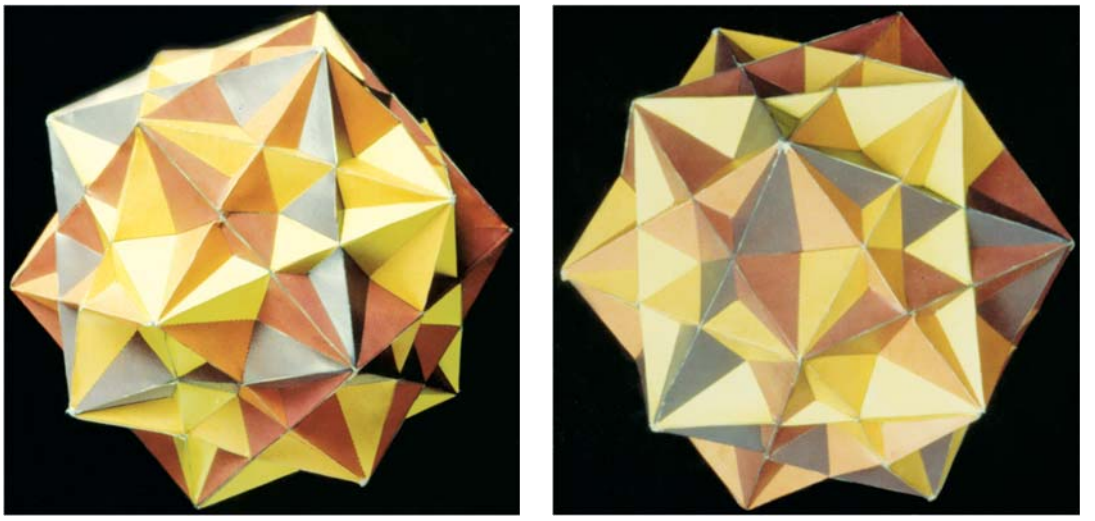


Figure 2.1 A puzzling object in two views. (From a cut-out book by E. Jenkins and M. Bear, (1985). Tarquin Publications, Diss, UK.)

Here we are going to describe the internal structure of DNA. The internal structure is the key to understanding how DNA works. In writing this book, we have tried to make everything as easy as spotting the structure of the five-fold cube. We may sometimes have to do a little mathematics, because mathematics tightens up the various relationships that we shall discover. There are a lot of things to learn about DNA, but we have tried to select the most important and general points. First, we shall learn why DNA makes a *double helix*, as we saw in Chapter 1, instead of some other shape; and then we shall learn how the double helix is held together at its core.

You should recall from Chapter 1 that DNA is made of three things: phosphates, sugars, and bases, and that these components are linked together chemically in a particular way. Now the phosphates are very soluble in water. If farmers use too much phosphate on their land as fertilizer, some of the phosphate runs off into the nearby ditches and rivers, causing algae to grow wildly and so killing the fish. Phosphates are very soluble in water.

Sugars are also soluble in water. There would be no point in putting sugar into coffee or tea if it would not dissolve: if it just sat at the bottom of the cup, it would not flavor the liquid. So sugars, as well as phosphates, dissolve in water.

But what about the bases? People have no intuitive feel for the four DNA (or RNA) bases: guanine, adenine, cytosine, and thymine (or uracil), because they do not recognize them in everyday life. It is easy to find out whether bases dissolve in water; just put some into

a test tube, add water, and watch. You can buy adenine and uracil from any chemical company (they cost little) and use about 50 mg of each. When you do this experiment you find that neither adenine nor uracil dissolves in ordinary water. Yet further simple experiments show that adenine dissolves in weak acid and that uracil dissolves in ammonia, an alkali – but not vice versa.

Now we mentioned in Chapter 1 that the space in our cells which is not occupied by important components such as DNA, RNA, and enzymes, is filled with water. This water is not at all acidic or alkaline. To use a technical term, it is at 'neutral pH'. It follows, therefore, that adenine and uracil (and indeed all of the bases listed above) will be practically insoluble in the aqueous environment of our bodies. Although we are not familiar with bases A, G, C, T, and U in everyday life, we are very familiar with other substances that will not dissolve in water, such as grease and oil. These are all 'water hating' or 'hydrophobic' substances.

The insolubility of bases in water does not really pose a problem for the cell, because these bases do become soluble in water once they are attached to a sugar and a phosphate to form a 'nucleotide', which is the building block of DNA or RNA (see Fig. 1.8). But this insolubility does place strong constraints on the overall conformation of any large DNA or RNA molecule in solution. For such a molecule to be stable in water at neutral pH, the bases will have to tuck themselves into the very center of some folded structure, so as to avoid the water; while the sugars and phosphates, both of which are soluble in water, will have to be on the outside. In fact, this is just what happens. If we take some measurements of the known structure of a DNA sugar–phosphate chain (determined by X-ray analysis, as described in Chapter 9), we see right away how DNA forms a spiral or helix on account of the low solubility in water of the bases. We can even do a first-order calculation to determine what kind of helix it makes.

We know from elementary chemistry that the distance between adjacent sugars or phosphates in the DNA chain is 6 Å (or 0.6 nm) in the usual case (Fig. 2.2). It cannot become much longer than 6.5 Å or shorter than 5.5 Å, or else the strong bonds between the atoms will strain too much. The thickness of the flat part of a DNA base is 3.3 Å, and this distance cannot change much either, because the bases are chemically rigid with strong, inflexible bonds between the atoms. This leaves us with a 'hole' of 2.7 Å between the bases, which some greasy object (and definitely not water) would have to fill, otherwise we will leave a vacuum. In brief, the bases are attached to a sugar–phosphate chain that is twice as long as the thickness of the bases themselves.

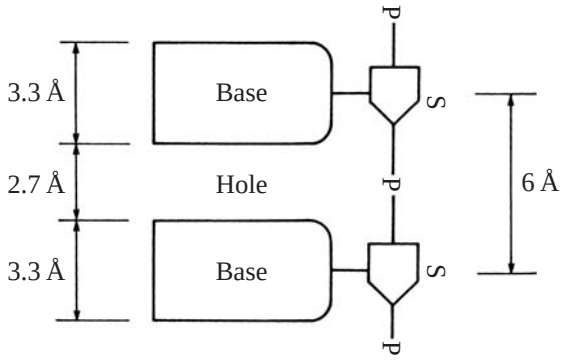


Figure 2.2 Two nucleotides in schematic form, showing key dimensions.

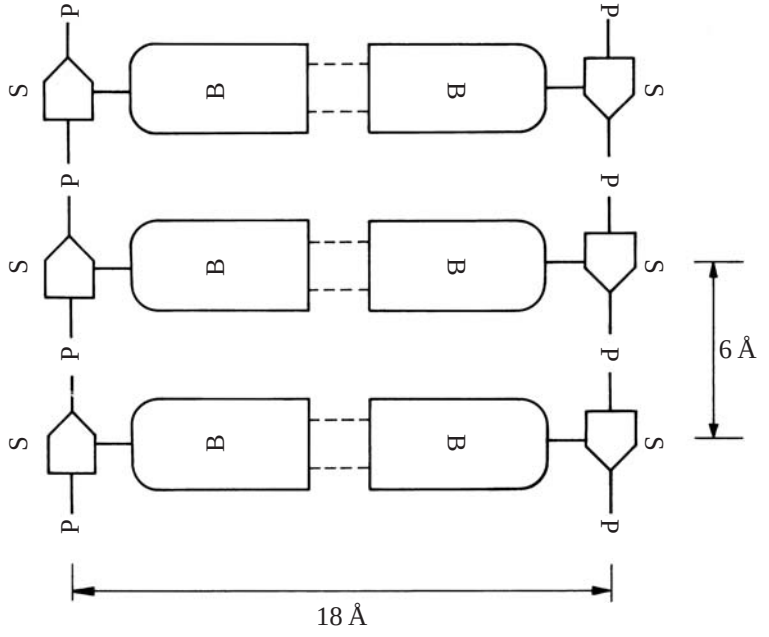


Figure 2.3 Part of a hypothetical DNA ladder, made by the cross-chain pairing of bases.

How can we tuck these insoluble bases into the center of a DNA molecule where they can avoid water, and at the same time be rid of the 'holes'? The most obvious form of DNA, as an assembly of two chains with the bases on the inside, would be a *ladder*. A segment of such a hypothetical ladder is drawn in Fig. 2.3. (A single chain of DNA could also fold back on itself to make a ladder, but this structure would be so similar to the one shown that it need not be considered separately.) In our ladder model, the two bases from opposite strands, joined in ways which we describe below, hold the phosphates 18 Å apart. Also, the two chains run in opposite directions, for reasons that

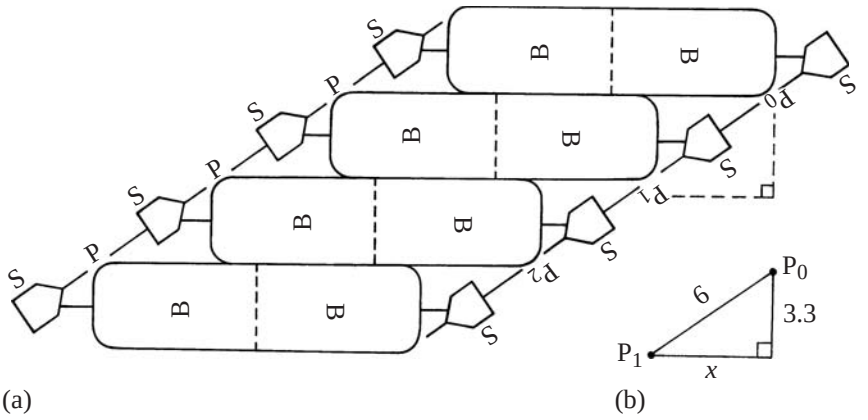


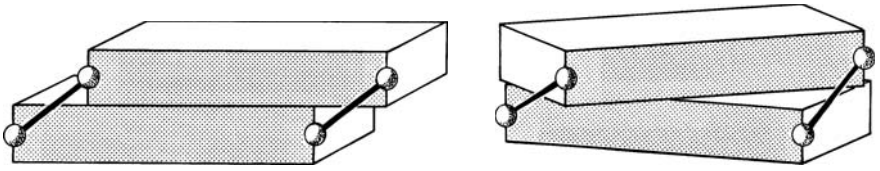
Figure 2.4 A skewed ladder, with no gaps between the paired bases. The plane geometry of this ladder is shown in part (b).

we discuss below. All seems satisfactory in that respect, but we are still left with many 'holes' between the bases within each strand. What can we do to remove them?

As shown in Fig. 2.4(a), one solution might be to skew the ladder strongly to one side. Once the sugar–phosphate chains tilt to an angle of about 30° from the horizontal, then the holes disappear, as shown. The key to the geometry is shown in Fig. 2.4(b), where we see a right-angled triangle that accurately describes the structure. The phosphates are connected along the hypotenuse of this triangle at a distance of 6 Å, while the bases proceed upward by 3.3 Å along the right-hand side. The sideways-shift of the ladder per base-step, left to right across the paper, is therefore given by

$$x = \sqrt{(6^2 - 3.3^2)} = 5.0\,\text{Å}.$$

Our skew-ladder seems perfectly satisfactory as a way of closing up the bases so as to exclude water. But it is not quite the same as the conformation which DNA adopts in Nature. As we have said in Chapter 1, DNA takes the form of a spiral or helix. In fact, the DNA double helix is nothing more than a highly twisted ladder. It provides another, slightly different way of solving the same problem: how to separate the bases by 3.3 Å while leaving the phosphates 6 Å apart. Figure 2.5(a) shows a simplified view of 2 base-pairs from our skew ladder, while Fig. 2.5(b) shows that the bases can stack onto each other just as well, without gaps, if they twist about an imaginary vertical axis into the shape of a helix. The two chains climb from the horizontal at exactly the same angle as before – about 30° – but now they lie on the surface of a cylinder of diameter 18 Å; and the base-pairs within the helix are now arranged as in the treads of a spiral staircase.



(a) (b)

Figure 2.5 (a) Stacking of base-pairs as in the skewed ladder of Fig. 2.4; (b) stacking of base-pairs by means of helical twist.

Why does DNA prefer to form a helix rather than a skew-ladder? Our present model is too crude to answer this question convincingly. When we build a more accurate model which shows individual atoms, we find that the skew-ladder leads to many unacceptably close contacts between neighboring atoms, and so this model has to be abandoned. Nevertheless, the skew-ladder is a useful tool in thinking about the internal structure of DNA, because it has several important geometrical features of the real DNA helix, and yet lies in a plane and is thus easy to visualize.

Now when we go into three dimensions, and consider the shape of a DNA helix, the geometry is almost the same as in our skew-ladder above. We can take a series of the twisted, two-base-pair units shown in Fig. 2.5(b), and stack them on top of each other to get a proper, double-helical model of DNA. Figure 2.6(a) shows such a model schematically. Only the first two base-pairs are shown, but then we show all parts of the sugar–phosphate chains. These chains wrap as spirals around an imaginary cylindrical surface of radius 9 Å, and each sugar ring is represented by a dot. Figure 2.6(b) shows a side view of the cylinder for just one of the two sugar–phosphate chains. Here the phosphates, P_0, P_1, P_2, etc. – counting from the top – are drawn as open circles, and the same lengths of 6.0 Å, 3.3 Å, and 5 Å that were found for our skew-ladder characterize the path of these phosphates through space. Finally, a top view along the vertical axis of the DNA cylinder is shown in Fig. 2.6(c). Again, for the sake of simplicity, only one chain is shown, and the phosphates along it are labeled P_0, P_1, …, P_{10}. Each successive phosphate in this view lies 3.3 Å further away from us than the one before. The chain is shown with a break between P_{10} and P_0, because P_{11} lies directly behind P_0 in this view: it is 11×3.3 Å $= 36$ Å further away from us, when we look down into the plane of the paper.

Simple geometry enables us to calculate the angle θ (*theta*) by which each phosphate turns relative to its neighboring phosphate along the helix. As shown in Fig. 2.6(c), the distance $x = 5.0$ Å is the base of an isosceles triangle, whose vertex lies at the center

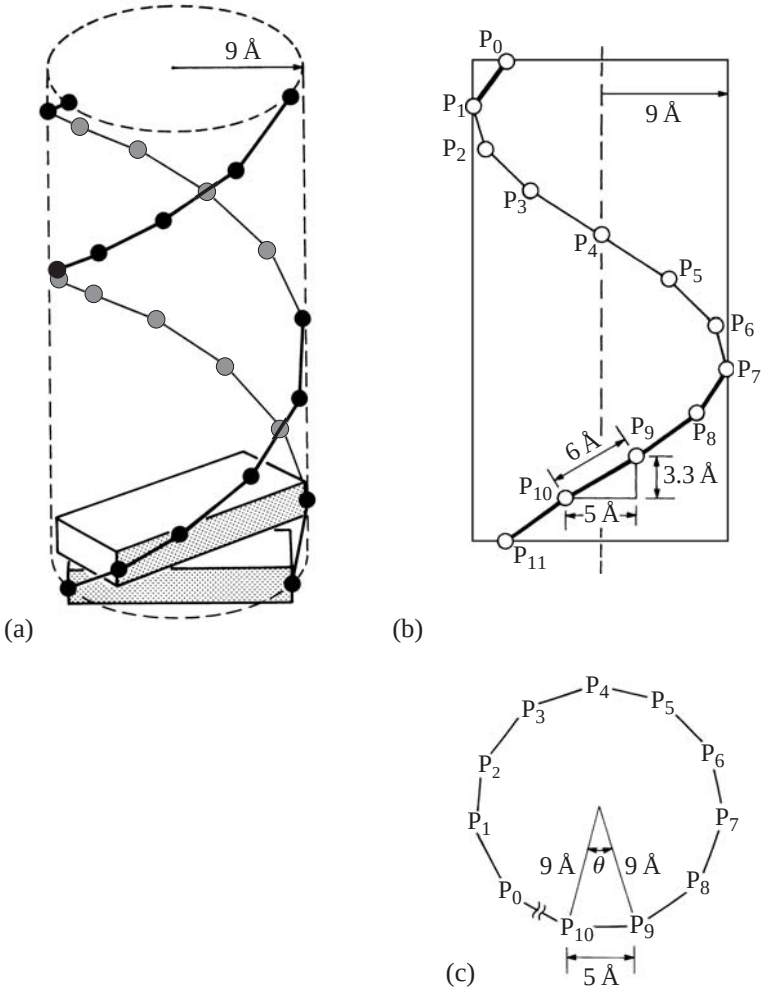


Figure 2.6 Sugar–phosphate chains wrapped helically around a cylinder: three views. In (a), sugar rings are drawn as shaded or filled circles, while phosphates are thin lines. In (b), phosphates are drawn as open circles, while sugars are thin lines. In (c), the view is down the long axis of the cylinder, looking along the dashed line in part (b).

of the cylinder at a distance of 9 Å from any phosphate. The value of θ can, therefore, be found by making a scale drawing, or else calculated as $2 \times \arcsin(2.5/9.0) = 32.3°$. Thus, each phosphate-to-phosphate rotation makes an angle of $32.3°/360° = 1/11$ part of a circle; and that is why we have put 11 phosphates in Fig. 2.6(c), to represent a complete turn of DNA.

This calculation, although relatively simple, tells us something which agrees closely with experiment: almost all DNA double helices have between 10 and 12 phosphates per turn of helix, within each

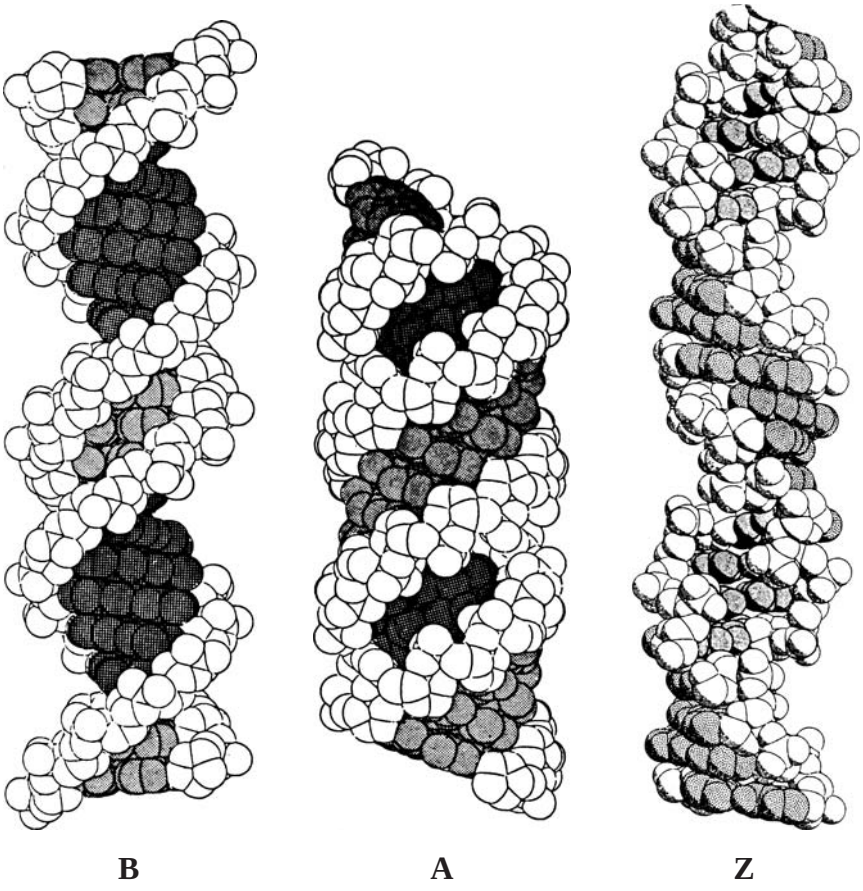


Figure 2.7 Three well-known (but highly idealized) forms of DNA: 'B' and 'A' are right-handed with 10 and 11 phosphates per helical turn, respectively, while 'Z' is left-handed with 12 phosphates per turn. Real right-handed DNA in solution averages about 10.5 phosphates per turn, or halfway between 'B' and 'A'. Pictures of 'A' and 'B' from C.J. Alden and S.-H. Kim (1979) *Journal of Molecular Biology* **132**, 411–34. Picture of 'Z' from H.R. Drew and R.E. Dickerson (1981) *Journal of Molecular Biology* **152**, 723–36 (with atoms shown somewhat smaller).

strand. For example, the well-known 'A' form of DNA (see Fig. 2.7) has 11 phosphates per turn, while the 'B' form has 10, and the 'Z' form has 12. These slight differences are significant in biology, and we shall discuss them later. But the crucial point here is that we have learned something important about the internal structure of the helix. Simply by studying the dimensions of the bases and of the sugar–phosphate chains, and knowing that the bases are insoluble in water (and so must stack directly onto each other), we have been able to determine that DNA will form a helix with about 11 phosphates per turn.

One point has been overlooked so far: how can we decide if our helix should be right-handed or left-handed? That is, as the