73–5. Use of a triple helix with iron (Fe) attached to 'rust' a yeast chromosomal DNA molecule into two pieces.

Watson, J.D. and Crick, F.H.C. (1953) A structure for deoxyribose nucleic acid. *Nature* **171**, 737–8. First proposal of the rules for base pairing of A with T and G with C, and of a two-stranded, double-helical model for DNA.

Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738–40. First companion paper to Watson and Crick (above), giving theory of diffraction of helices.

Franklin, Rosalind, E., and Gosling, R.G. (1953) Molecular configuration of sodium thymonucleate. *Nature* **171**, 740–1. Second companion paper, showing a clear X-ray diffraction picture of 'structure B'.

Exercises

2.1a On planet P all living things are found to contain a DNA-like double-helical molecule just like that found on Earth, except that the molecule on planet P consists only of the two nucleotides A and G. Thus, A–G is the only scheme for base pairing, and the large A–G base pairs impart a separation of 20 Å to the two sugar–phosphate chains. By adapting the calculations shown in Figs 2.4 and 2.6, estimate the angle of helical twist and the number of base-pairs per helical turn for the special DNA on planet P.

b On planet Q the DNA molecule contains four nucleotides, namely A, G, C, and T, just as on Earth; but the sugar–phosphate chain is found to be 7Å long between phosphates (compared with 6 Å on Earth) on account of an extra carbon atom in each nucleotide unit. Given that the bases pair in Watson–Crick style, how many base-pairs do you expect per double-helical turn in the DNA on planet Q?

c On planet R the genetic molecule is exactly like terrestrial DNA, except that the oceans on planet R are slightly acidic (pH 4), such that the base pairings A–T and G–C are mostly in accordance with the Hoogsteen scheme (Fig. 2.12). In consequence, the sugar–phosphate chains spiral about an imaginary cylinder of diameter 16 Å. Estimate the number of base-pairs per helical turn of DNA on planet R.

2.2 Idealized 'B' form DNA has a helical twist of 36° and 'rise' of 3.3 Å, in the axial direction, per base-pair step (see Fig. 2.7). When a molecule of the intercalating drug ethidium bromide inserts itself into a step of DNA in the manner of Fig. 2.9, it increases the length

of the DNA by 3.3 Å, and at the same time reduces the helical twist at the step by 26°, i.e. from 36° to 10°.

Find the overall length of a 100-bp segment of 'B' form DNA, and the total number of helical turns:

a with no ethidium bromide;
b with one ethidium bromide molecule for every 10 base-pairs;
c with one ethidium bromide molecule for every 2 base-pairs.

(Case **c** corresponds to the largest possible uptake of ethidium bromide by DNA.)

2.3 In some circumstances, DNA can make a triple helix by forming planar hydrogen-bonded base triplets in place of the usual base-pairs. Construct such a triplet from one A and two Ts in two different ways, with each thymine T connected by two hydrogen bonds to the adenine A.

a Begin with the A–T pair of Fig. 2.11(a), and add a second T to the 'unoccupied' upper edge of the A by moving the first T base around in the plane of the paper.
b Begin with the usual A–T pair, as before; but now obtain the second T by flipping the first T over onto its other face before moving it around.

Which of the new pairings is equivalent to the Hoogsteen arrangement of Fig. 2.12(a)?

(Hint: Work with a copy of Fig. 2.11(a), and use tracing paper – either way up – for the second T. Note that either oxygen of the thymine can act as a hydrogen-bond acceptor.)

2.4 Investigate possible G–A base-pairings as follows: begin with the G base from Fig. 2.13, along with its hydrogen-bonding scheme; and use A bases with the hydrogen-bonding schemes of Figs 2.11(a) and 2.12(a), but 'flipped over' in each case. Work on tracing paper. Do the two different G–A pairings differ much in their sugar–sugar distances?

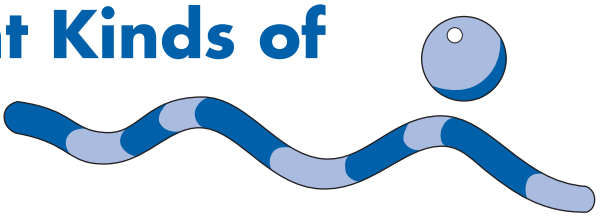
2.5 The three-base code in a DNA or messenger-RNA chain (see Table 1.1) is read with the assistance of various transfer-RNA molecules, as shown schematically in Fig. 2.14. There, one particular transfer-RNA molecule is shown recognizing the messenger-RNA sequence GCU (corresponding to GCT in DNA); but the same transfer-RNA can also recognize the sequence GCC, by using a standard G–C pair rather than an unusual G–U in position 3 of the message.

a Suppose that a particular cell has 32 different kinds of transfer-RNA molecule, each with one of 20 possible amino acids attached at its distal end. On average, how many different sets of base triplets in the messenger-RNA chain would each transfer-RNA have to recognize, in order to make all the proteins necessary for cell growth?

b What is the smallest number of different transfer-RNA molecules that the cell could use and still be viable? How many triplets would any particular transfer-RNA molecule then have to recognize, on average?

CHAPTER 3

Different Kinds of Double Helix



In the previous chapter we learned two things: why DNA forms a double helix, and how the bases interact with one another at the oily, water-insoluble core of the helix to form 'base-pairs'. The driving force for helix formation was shown to be the need for the bases to escape from water by joining with other bases at the core of the helix. Yet they cannot stack directly on top of one another while doing so: rather, they must twist around slightly, because they are attached to sugar–phosphate chains that are twice as long for each base as the thickness of the base itself.

All of these points are what might be called 'first-order' influences on the structure of DNA. Now you might be hoping that you will not need to learn about 'second-order' effects: you can perhaps manage by just knowing the first-order effects. That is a forlorn hope, however, because you will have to learn about the second-order effects before you can understand many of the roles of DNA in biology: for example, how promoters work, how DNA wraps into chromosomes, and even how DNA binds to the 'repressor' proteins, which influence how well promoters work. As you may recall from Chapters 1 and 2, the DNA within any gene makes an RNA copy of itself which then goes on to make protein; a 'promoter' is a short region of DNA near the gene that tells the cell how many RNA copies to make, and hence how much protein.

We shall limit ourselves to three themes in our study of second-order effects on the structure of DNA: first, how the bases undergo 'propeller twist' to make sure that as much as possible of their oily, flat surfaces escape from contact with water; second, how the base-pairs stack on one another in particular ways that depend upon the ordering or sequence of bases; and third, how certain small,

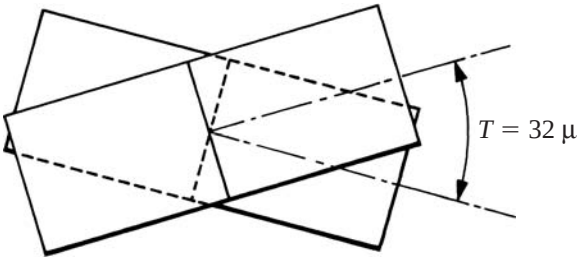


Figure 3.1 Two base-pairs with 32° of right-handed helical twist: the 'minor-groove' edges are drawn with heavy shading, as in Fig. 3.5. As in Fig. 2.11, the base-pairs are shown parallel to the paper.

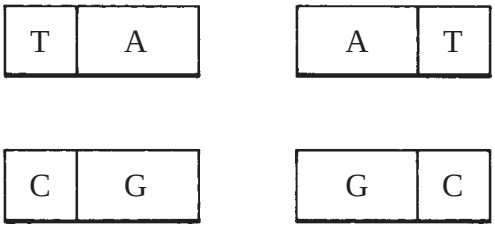


Figure 3.2 Four possible base-pairs of the Watson–Crick type, each of which joins a large purine (A, G) to a small pyrimidine (T, C), and in the same perspective as Fig. 3.1.

subtle motions of the base pairs can accumulate over a series of such base pairs, to make different kinds of double helix.

We said in the last chapter that the bases form ordered pairs at the core of the helix, leaving sugars and phosphates on the outside; and that each base-pair twists with respect to its neighbor by about 32°, as in the treads of a spiral staircase. The sense of this rotation is right handed, or clockwise going forward, as shown in Fig. 3.1: this is the same helical sense as in an ordinary corkscrew.

Almost always, the base-pairs are of a Watson–Crick kind, joining guanine (G) with cytosine (C), and adenine (A) with thymine (T): see Fig. 3.2. Bases A and G are called 'purines', and they are bigger than bases C and T, which are called 'pyrimidines'; yet the overall size of the base-pair is roughly the same in all four possible arrangements. The apparent simplicity of these arrangements once led scientists to conclude that the base sequence of DNA could not influence its three-dimensional structure, because all four kinds of base pair could slot into a perfectly uniform double-helical 'staircase'. This conclusion was not based on any firm evidence, however, and it has proved to be incorrect. For example, typical angles of base-pair twist in real DNA molecules, as determined from the many high-resolution maps of DNA structure collected since 1980, range from

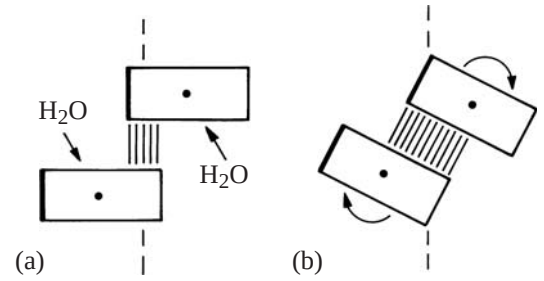


Figure 3.3 Propeller twist, as in (b), allows greater overlap of bases within the same strand and reduces the area of contact between the bases and water.

20° to 50° about a mean of 34°. The mean value of 34° is close to our prediction of 32° from the simple theory presented in Chapter 2; yet our first-order theory did not predict anything about a variation in twist away from 32° to other values. It certainly did not predict a broad range of twist from 20° to 50°, as is typically observed. Evidently, we must develop a theory that includes these second-order effects before we can claim to have any real understanding of the structure of DNA.

Our starting point for the second-order theory is rather subtle: it seems that because of the substantial twist between adjacent base-pairs, less than the entire surface of any base pair can escape contact with water. Thus, only the central overlapping portions of the base surfaces in Fig. 3.1 are protected from water, while the four overhanging triangular portions are not protected. When we view the two right-hand bases in Fig. 3.1 from the perspective of the right-hand margin of the page, edge-on in the plane of the paper, we see the arrangement shown in Fig. 3.3(a). Taken together, Figs 3.1 and 3.3(a) show that the overlap of consecutive pairs is good in the interior of the stack, but is only poor in the outer regions, which are exposed to the water.

What can we do to improve this situation? One solution would be to rotate each of the bases shown in Fig. 3.3(a) in a clockwise sense about its long axis, which points down into the plane of the paper in this view. Such a motion is shown in Fig. 3.3(b), where each base rotates slightly about its end-centerpoint, which is a black dot in the diagram. Stacking is improved, since the water is now excluded from a larger fraction of the surfaces of the two bases. If this motion is not clear to you, stretch out both of your arms in front of you, with your right hand above your left, both hands horizontal, and with the tips of the middle fingers vertically above each other and about 5 cm apart. Then move your right hand by about 5 cm in a left-to-right sense, so that it no longer lies directly over

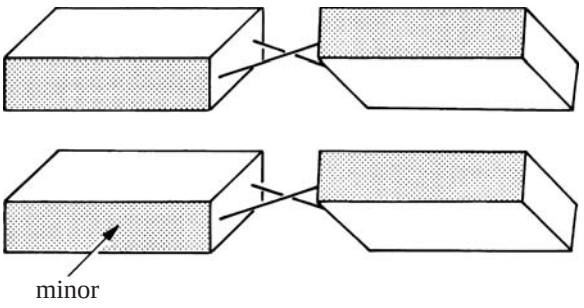


Figure 3.4 Propeller-twisted base-pairs. Note how the hydrogen bonds between bases are distorted by this motion, yet remain intact. The minor-groove edges of the bases are shaded.

your left. Finally, rotate both hands clockwise about the wrists by around 20°; and you will find that your two hands 'cover' each other much better than they did before.

Now the two bases on the left-hand side of Fig. 3.1 can likewise rotate about their long axes to exclude water, but their sense of rotation must be counter-clockwise, when viewed from the perspective of Fig. 3.3, in order to achieve the same result. All four bases from Fig. 3.1 are shown together in Fig. 3.4. In this picture, each base pair looks somewhat like an old-fashioned airplane propeller, since the left-hand and right-hand bases twist in opposite directions. Hence, the overall motion is called 'propeller twist'. Its sense is that of a left-handed screw as one goes forward along the pair from one base to its partner.[1] Propeller twist obviously distorts the hydrogen bonds that hold the two bases together, which are shown schematically by two lines in Fig. 3.4; but these weak bonds can accept some distortion of that kind, provided the degree of propeller twist is not too great.

In general, such propeller twist tends to be higher than average in regions of double helix containing mostly AT base pairs, typically 15° to 25°; but lower in regions of helix containing mostly GC base pairs, typically 5° to 15°.

Also note that we have shaded one edge of each base in Fig. 3.4, and have labeled one of these with the term 'minor'. There is a convention to call one side of a base pair the 'minor-groove side', and the other the 'major-groove side'. Where do these names come from? As shown in Fig. 3.5(a), the two sugars to which a base-pair are attached lie closer to one side of the base-pair than the other. The edge which lies closer to an imaginary line drawn between the two sugars is called the 'minor-groove side', while the other edge is called the 'major-groove side'. By convention, we shall always

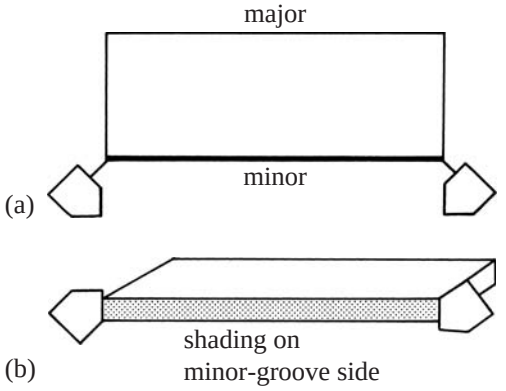


Figure 3.5 Two views of a base-pair, showing directions of the sugar–phosphate chains, just as in Figs 2.3 and 2.4. By our convention, the minor-groove edges are shaded. Here, the base-pair is shown as a single block.

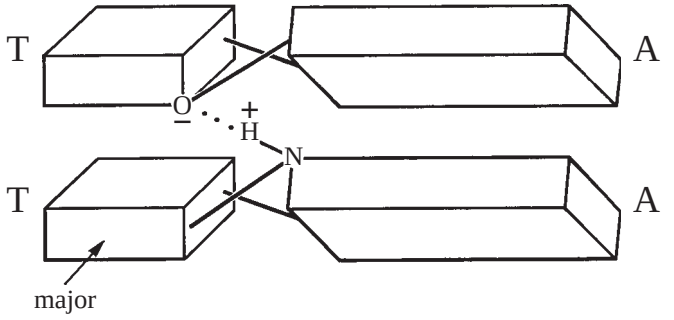


Figure 3.6 Propeller-twisted A–T pairs, showing an additional hydrogen bond between the base-pairs in the major groove, as proposed by Hillary Nelson (cf. Fig. 2.11(a)).

shade the minor-groove side of a base-pair in our drawings, as indicated in Fig. 3.5(b).

Why do we use the term 'groove' in this labeling convention? Early structural models of DNA showed a cylinder with two hollow, spiral grooves lying between the two sugar–phosphate chains. One of the grooves – 'minor' – was smaller than the other. You can see these two grooves in the 'B' model of Fig. 2.7. But for some of the helices which we shall study in this chapter, it turns out that the so-called minor groove is actually as large as or larger than the so-called major groove: for example, see the 'A' model of Fig. 2.7. We can get around this difficulty by talking instead about the minor-groove *side* of the base pairs themselves.

Of course, we can also view the base-pairs from the major-groove side. An example of this is shown in Fig. 3.6, where we see that the near edge, labeled 'major', remains unshaded. Furthermore, in this

drawing we can see that the two bases A and T within either pair are of unequal size. In general, as we have said, purines A and G are larger than pyrimidines T and C. In several previous drawings, for example, Figs 3.1 and 3.4, we have omitted this feature for the sake of clarity; but when we start to consider the interactions of real DNA bases, such as those shown in Fig. 3.6, we have to make our drawings more accurate. Figure 3.6 also shows the detail of a possible hydrogen bond (N–H … O) between adjacent base pairs, from adenine in the lower pair to thymine in the upper. Such a hydrogen bond might be expected to increase the propeller twist, because the distance is right for it to form only when the base pairs are highly twisted along their long axes. Indeed, experiments show that regions of DNA with all adenine bases on one strand and all thymine bases on the other do have an unusually high propeller twist of about 20° to 30° as against 10° to 20° for other sequences.

You can imagine that if we were to study all kinds of two-base-pair arrangements in DNA, or indeed all kinds of three-base-pair arrangements, we should find a lot of unexpected but important contacts between the bases. For that reason it is very hard to understand, on an atomic scale, the behavior of any long DNA molecule such as those found in biological systems. We need some sort of simplified description of base-pair arrangements in DNA, at less than an atomic level of detail, if we wish to understand the many roles of DNA in biology.

The simplification which we make at this point is one which we have already used in some drawings: we construct an imaginary, flat plane that coincides as well as possible with the twisted surface of any base-pair. In other words, we shall pretend that the base-pairs shown in Figs 3.4 and 3.6 have no propeller twist, and treat them as rigid, rectangular blocks, such as those shown in Figs 3.5 and 3.7. Previously, of course, we have said how important it is to

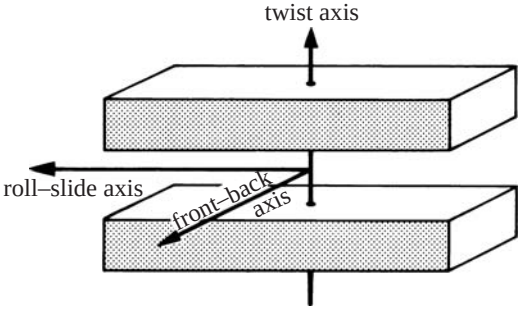


Figure 3.7 Local reference axes for an individual base-pair step, following the mathematics of Leonard Euler.

build propeller twist into a base-pair; but now it seems that we are going to ignore this twist entirely in a simplified model! You must simply remember that the propeller twist is always *there*, but that it is not always *shown* in the diagrams. This is a bit like drawing a pocket watch without showing the gears inside: they are there, but not visible. For some purposes we can understand things more clearly by thinking about a deliberately simple representation of a base-pair.

Our simplified model is based on the work of the famous Swiss mathematician, Leonard Euler (1707–83). He explained that if you have two rigid objects, such as the rectangular blocks shown in Fig. 3.7 – here representing two base-pairs – and you want to describe the position of one block relative to the other, then you will need to use six variables or 'degrees of freedom': three translations and three rotations. A translation[2] is a change of position without any rotation: imagine moving a cardboard box from one place in a room to another, so that every face of the box always moves parallel to itself. The position of the box in the room can be specified completely by the values of three 'coordinates', say x, y, and z, measured from a suitably chosen 'origin'. A rotation, on the other hand, involves a change of angle without a change in position: for example, you could pick up a cube from a table, turn it through 90°, and then put it back on the table in the same position as before.

It is easy to understand why you need six variables to describe completely the position of one solid block relative to another. Suppose that the upper base-pair or block shown in Fig. 3.7 were temporarily replaced by a *point*: then you would only need three translation coordinates, x, y, and z, to say where this point might be located relative to the lower block. Next let us build an upper block around the point, and ask how many kinds of rotation we need to orient the upper block relative to the lower block. The answer is exactly three, as indicated by the three axes of rotation drawn in Fig. 3.7. We call these the 'twist', 'front–back', and 'roll–slide' axes for reasons that will become clear soon.

So it seems that we need only six numbers to describe the local configuration of any two neighboring base-pairs from a mathematical point of view. That is not too bad. But when we look at real DNA structures, the situation becomes even more favorable. To a first approximation, only three of Euler's six possible degrees of freedom are actually mobilized in real DNA double helices. This fortunate simplification comes about because the base-pairs are attached, as we recall from the previous chapters, to sugars and phosphates which limit their range of maneuver in certain directions, notably along the front–back axis of Fig. 3.7. Also, the