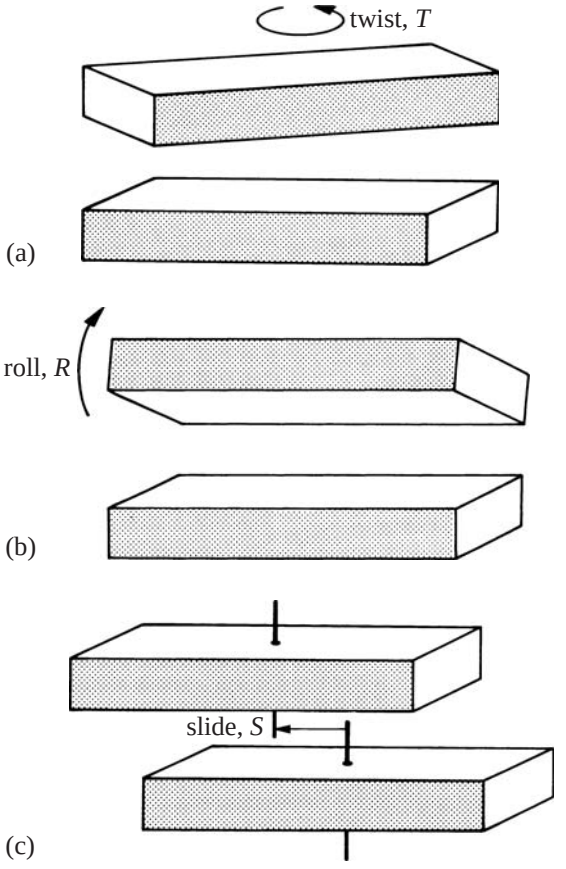

Figure 3.8 Twist, roll and slide motions at a base-pair step. Each drawing defines the positive sense of twist, roll, or slide, as used in this book.

base-pairs cannot be separated along their vertical twist axis without introducing water or a vacuum between them.

Thus, in practice, we need only three variables to describe the motions of any base-pair relative to its neighbor. Two of these are rotations, and one is a translation. All three are shown separately in Fig. 3.8. The first rotation, or 'twist', (Fig. 3.8(a)) is the same twist which we estimated as about 32° in Chapter 2, and which we showed in Fig. 3.1. We shall now give it a rigorous definition: it corresponds to a rotation about the local twist axis that runs vertically through, or near, the centers of any two neighboring base-pairs, as shown in Fig. 3.7. Note that if the DNA happens to be curved, or if the base-pairs stack on one another locally, such that they do not advance directly along the overall helix axis, then the *local* twist axis shown in Fig. 3.7 may not coincide with a *global* twist axis for the whole molecule, averaged over many steps; and the twist

angles measured about these two different axes may also be slightly different. So it is important to define what kind of twist we mean, when we talk about the 'twist' of DNA.

The second kind of rotation, or 'roll' (Fig. 3.8(b)), describes the rolling open of base pairs along their long axes. Angles of roll vary from $+20°$ to $-10°$ in the usual DNA structures. By convention, we say that the roll is positive if base pairs open up towards the minor-groove side, as shown in the diagram. Actually, the surfaces of individual bases do not come apart from one another very much in the roll motion: they only appear to do so because we have drawn the two base-pairs as uniform, rigid blocks. This simple kind of drawing conceals details in much the same way that the cover of a pocket watch conceals the gears.

The last commonly observed kind of motion is a translation or 'slide' (Fig. 3.8(c)). It describes the relative sliding of neighboring base pairs along their long axes. Slide is defined as positive if the upper pair goes further to the left than the lower pair, when we look at the minor-groove edges. Typically, values of slide range from $+3$ Å to -2 Å in real DNA; the sugar–phosphate chains do not easily allow any further motion. Still, these sugar–phosphate chains do allow a great deal more flexibility about the roll–slide axis, as epitomized by 'roll' and 'slide', than they do about the front–back axis shown in Fig. 3.7.

In summary, there are three significant relative motions of the base pairs at any base-pair step. They are called roll, slide, and twist, and they may be abbreviated to *R*, *S*, and *T*, respectively. The positive sense of each is shown in Fig. 3.8(a), (b) and (c).[3] You cannot forget the series *R*, *S*, and *T*, although you have to remember which letter stands for which degree of freedom.

The propeller-twist motion, which we were talking about earlier, is a property of a single base-pair, and not of two base-pairs that lie over each other. Thus, we can imagine two short pieces of DNA that have different values of propeller twist but the same values of roll, slide, and twist. As shown in Fig. 3.9, if we stack two base-pairs directly on top of one another, at zero slide and zero twist for the sake of simplicity, then the roll remains zero no matter how much propeller twist we add equally to both pairs. At the level of the 'rigid block' drawings shown in Figs 3.7 and 3.8, both parts (a) and (b) of Fig. 3.9 would look identical, despite the change in propeller twist on going from (a) to (b). Bases on the right-hand strand tilt upward along their minor-groove edges by 10° and those on the left-hand strand tilt downward by 10°; nevertheless, the two effects cancel when we calculate a mean plane for the entire base-pair.

Drawing base-pairs as rigid blocks, and then identifying roll, slide, and twist motions at the steps between base-pairs may seem

to be a strange way of thinking about DNA; but it is the most successful way so far devised. Therefore let us continue, and see how the three parameters roll, slide, and twist might be useful for understanding different arrangements by which base-pairs stack onto each other in different types of two-base-pair steps. Even though we have pushed the idea of propeller twist out of the limelight for the moment, we must not forget about it, because propeller twist is partly responsible for certain relations between *R*, *S*, and *T* which are found in real DNA.

Previously, in Fig. 3.6, we noted one important feature of DNA structure, which was a hydrogen bond between neighboring base-pairs in sequences of the kind AA or TT. The high propeller twist of

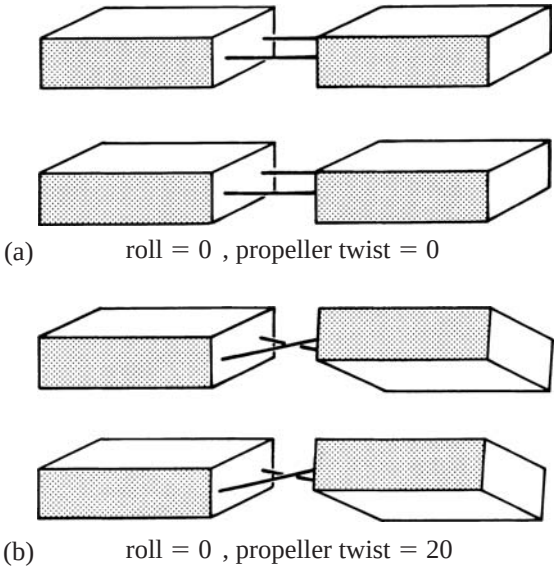

Figure 3.9 Propeller twist need not alter roll. The roll angle remains zero in part (b), because the mean planes of the base-pairs remain parallel.

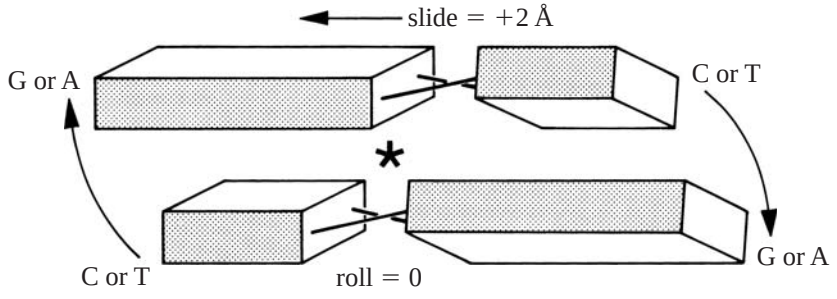

Figure 3.10 A pyrimidine–purine step with zero roll: positive slide is needed to avoid a steric clash at * if the base-pairs have propeller twist.

the two base-pairs allows a cross-chain link between an oxygen atom of T and a nitrogen N–H of A on the major-groove side. It also causes a close contact between an oxygen atom of T and a carbon atom of A on the minor-groove side. Those two cross-chain contacts help to hold the AA/TT step in a nearly fixed conformation, with $R = 0°$, $S = 0$ Å, and $T = 36°$.

Another important feature of DNA structure can be found at steps which we shall describe as 'pyrimidine–purine'. Figure 3.10 shows an example of such a step. You should have learned from the many pictures shown so far that the two sugar–phosphate chains of DNA run in opposite directions. In Fig. 3.10 these sugar–phosphate chains are not shown, but their directions are marked by arrows. Suppose in a particular step that the upper base-pair is C–G and the lower one is A–T in going from right to left across the page, then the entire step can be described as 'CA/TG': coming down along the right-hand chain, we have the sequence CA, and going back up along the left-hand chain we have TG. The 'slash' between CA and TG indicates the 'jump' from one chain to the other. If you were to turn the book through 180° in its own plane, and use the same procedure for naming this particular step, you would get 'TG/CA' instead of 'CA/TG'. It does not matter which term you use: both describe the same collection of atoms in space.

The drawing in Fig. 3.10 is intended to represent all possible pyrimidine–purine steps, which we may now list as TG/CA (=CA/TG), CG/CG, and TA/TA. These steps are seen commonly with two different kinds of base-to-base overlap, as shown schematically in Figs 3.10 and 3.11.

In Fig. 3.10, the large purine bases G or A slide away from each other by +2 Å, in order to avoid too-close contact at the point shown by the 'star' in the picture. Some positive slide is needed here, because both of the bases concerned are the large purines,

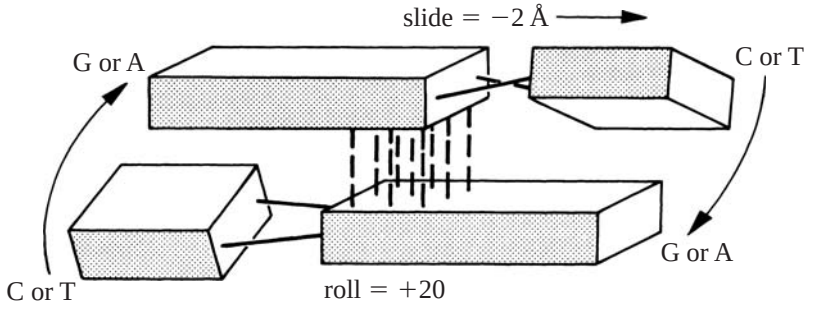

Figure 3.11 A pyrimidine–purine step in an alternative configuration having negative slide and positive roll, due to the cross-chain stacking of purines.

G or A. If the base-pairs were not propeller twisted, then contact of purines at the star would not be a serious problem, and the slide could be zero. Note that the roll angle remains near zero, since all bases remain parallel to their neighbors on the same strand.

In Fig. 3.11, the large purine bases slide on top of one another by −2 Å. In this conformation, the roll angle becomes large and positive, +20°, since the small pyrimidine bases must be inclined with respect to the large purine bases in either strand by +20°, in order to maintain the 20° of propeller twist. Apparently, the close stacking of pyrimidine on purine within either strand of real DNA is less significant than is the maintenance of high propeller twist.

There are two important points about Figs 3.10 and 3.11. First, values of slide lying between +2 Å and −2 Å might be expected to be less stable than the ones shown, because at intermediate values (say 0 to −1 Å) the two large purine bases will neither avoid one another fully, nor stack firmly on top of one another. One might expect, therefore, that pyrimidine–purine steps in DNA would be weakly 'bistable': that is, capable of adopting two extreme conformations but not always a continuous range; and this is indeed what is found. Second, by comparing the two structures shown in Figs 3.10 and 3.11, we can see that roll changes as we change the slide, on account of propeller twist. Thus, slide = +2 Å gives roll = 0°, while slide = −2 Å gives roll = +20°. This is very much like the motion of a bolt-action rifle: the bolt 'rolls' as it 'slides' forward to place a bullet in the chamber.

We have now seen two different ways by which the preferred close contacts of base pairs can influence the roll, slide, and twist values that are adopted at any step. First, as shown in Fig. 3.6, an AA/TT step can be held in a single, preferred position by a possible hydrogen bond on the major-groove side. Second, as shown in Figs 3.10 and 3.11, pyrimidine–purine steps can adopt at least two different types of stacking that cover a wide range of slide *S* and roll *R*. Both of these stacking effects are important in real DNA; but there is also a third stacking effect which is just as important as the other two, and which will be explained fully in Appendix 2 but will be mentioned briefly here. It has to do with how adjacent bases stack on one another, according to partial electric charges within the bases themselves.

The stacking of bases onto one another has been discussed so far in terms of only the van der Waals or hydrophobic effect, which provides for good overlap of any two bases in proportion to the area of contact of their flat, water-insoluble surfaces. The hydrophobic effect applies with equal strength to all bases or base pairs. It thereby contributes a tendency to propeller twist as mentioned above, and it can

be discussed adequately by means of our current pictures in which the base-pairs are drawn as featureless 'blocks'.

However, the usual A–T and G–C base pairs also contain within themselves many precise distributions of partial electric charge, which are spread over their flat surfaces. These are analogous to the partial electric charges used to explain Watson–Crick base pairs in Fig. 2.11. Yet in addition to the partial charges shown in Fig. 2.11, which lie on nitrogen N or N–H or oxygen O atoms that are extended outward from the main parts of each ring, other distributions of partial electric charge are located near nitrogen N or oxygen O atoms within the rings themselves. Such additional electric charges were not shown in Fig. 2.11 for the sake of simplicity; yet they are just as real, and can have almost as much effect on the preferred stacking of bases vertically, as they do when they form the Watson–Crick hydrogen bonds which join one base to its partner.

To a first approximation, the G–C base pair contains a large plus-or-minus electrical 'dipole' along its long axis, specifically as plus on C but minus on G; whereas the A–T base pair contains only small patches of isolated plus-or-minus electrical charge along its long axis, which are relatively dispersed over the entire pair, and hence do not amount to a substantial dipole.

Let us now imagine how two base-pairs will move over one another at different values of roll R and slide S. What effect, if any, will the partial electrical charges have on preferred values of roll and slide for different sequences? In general, we know that 'unlike' electrical charges attract, whereas 'like' electrical charges repel; and so we expect the largest effects of an electrical kind to be observed for steps where two G–C base pairs stack onto one another, such as GG/CC, CG/CG or GC/GC. Much smaller effects due to electrical charge are expected for steps containing both a G–C and an A–T base-pair, such as CA/TG; and hardly any electrical effect is expected for steps containing two A–T base pairs, such as TA/TA.

Thus, for those steps containing two successive G–C base pairs, where each G–C base pair shows a partial plus charge on its C-ring, and a partial minus charge on its G-ring, one expects that the two base-pairs will repel each other near slide $S = 0$ Å, in an arrangement where they are stacked directly on top of one another, due to 'like-to-like' repulsion. Hence, those two successive G–C pairs might prefer to lie slightly offset from one another vertically, at either positive or negative slide, in order to reduce the expected like-to-like charge repulsion. The effect of charge–charge repulsion for any step with two G–C pairs should, therefore, be similar to that shown in Figs 3.10 and 3.11 for pyrimidine–purine steps in general; yet it has a different chemical origin.

Our discussion of base-stacking arrangements in DNA seems to be getting rather complicated. We have three different situations to think about: (a) AA/TT steps, (b) pyrimidine–purine steps, and (c) steps with two G–C pairs. Yet the overall picture may be made clear by examining plots of roll R versus slide S for real DNA as shown in Fig. 3.12. All of the data shown in these plots come from structures of DNA oligomers, determined by X-ray crystallography, and are very precise.

First, Fig. 3.12(a) shows roll R and slide S values as observed for the AA/TT step, and collected from many different X-ray structures in a crystal. We can see there that the AA/TT step allows little variation of either roll or slide away from a mean of $R = 0°$ and $S = 0$ Å, as drawn previously in Fig. 3.6. Next, Fig. 3.12(b) shows the same plot for a CA/TG step, which is an example of the general pyrimidine–purine type. There we can see a wide range of roll and slide values, and also a strong connection between them as drawn schematically in Figs 3.10 and 3.11. Finally, Fig. 3.12(c) shows roll and slide data for the step GG/CC, which is broadly representative of the two G–C pair type. There we can see a clear 'forbidden zone' of slide S near 0 Å, where similar charges in adjacent G and G, or C and C bases, repel one another vertically. The experimental data, therefore, support all that we have said previously, and give us confidence that the structure of DNA can be understood in terms of ordinary chemistry, without any mysterious features!

Finally, to close our discussion of base stacking, could roll R and slide S be coupled with twist T in some general way? From first principles, we might expect to see a relationship between slide and twist if the sugar–phosphate chains were semi-rigid and of constant length: the relevant geometry is shown in Fig. 3.13. In this example, the twist angle decreases from 36° to 28° as the base pairs slide from 0 to −2 Å. There is a direct mechanical coupling involved here, because the sugar–phosphate chains are assumed to be rigid links of constant length. Actually, plots analogous to those shown in Fig. 3.12, but with twist plotted along the vertical axis instead of roll, do show a broad but definite tendency of the sort expected: low slide goes with low twist, while high slide goes with high twist. But the correlation of slide with twist is not so strong as the correlation of slide with roll, perhaps because sugar–phosphate chains are not actually as rigid as the picture of Fig. 3.13 suggests (see Appendix 2). In any case, it seems that roll, slide and twist in DNA are all related to one another.

Thus, having started with Euler's six degrees of freedom, and having eliminated three of these by introducing constraints on the base stacking due to various factors, we find finally that the three

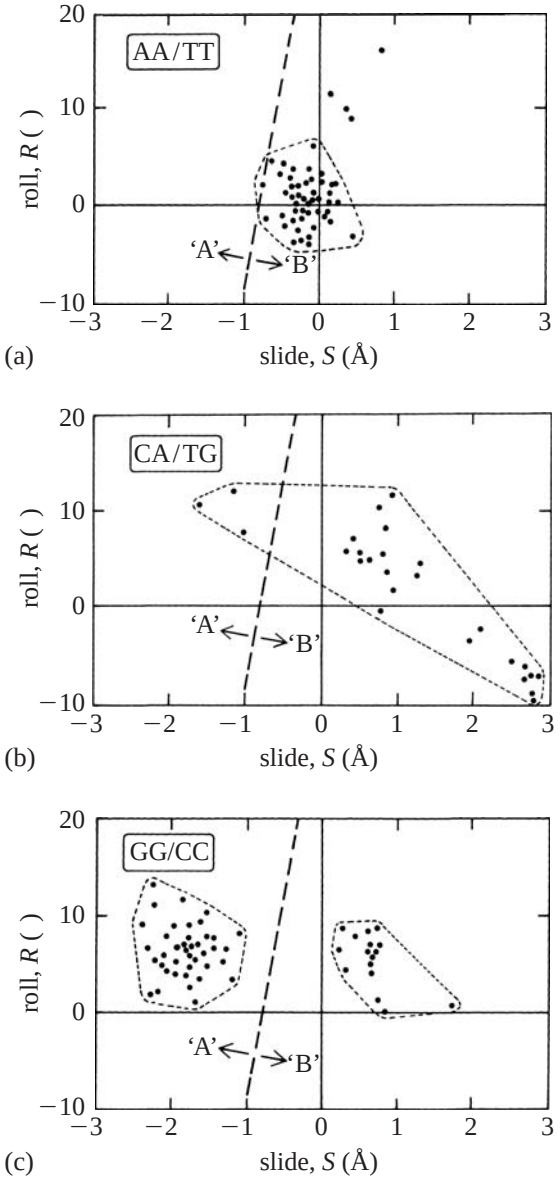

Figure 3.12 Plots of roll versus slide for many base-pair steps of oligomeric DNA as studied by X-ray diffraction. Separate plots are given for three of the ten distinct steps by sequence: AA/TT is a 'rigid' step, CA/TG is a 'flexible' step and GG/CC is a 'bistable' step. From El Hassan and Calladine (1997). *Philosophical Transactions of the Royal Society, A* **355**, 43–100.

remaining parameters *R*, *S*, and *T* are broadly related to each other in ways which depend both on the base composition of the step, and also on the general behavior of the sugar–phosphate chains, which connect the two base-pairs.

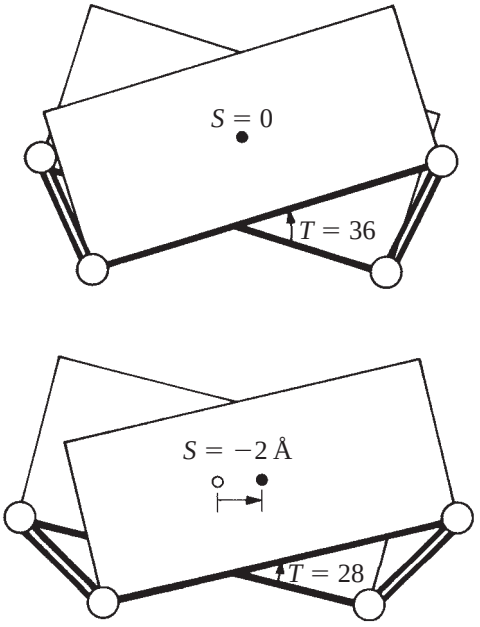

Figure 3.13 Schematic model of a base-pair step (cf. Fig. 2.5(b)) showing a possible mechanical linkage between slide and twist, which is confirmed weakly by experiment.

Now we have explained almost everything that is known today about the internal structure of DNA. In summary, the base-pairs adopt propeller twist to minimize their contact with water; this propeller twist prevents the otherwise flat base-pairs from sliding freely on one another's surfaces, and can sometimes 'lock' certain steps into a nearly rigid configuration. Some base-pairs also contain partial electric charges, which can prevent a step from adopting certain values of slide or can cause it to favor two separate values of slide. There are also some relations between roll, slide, and twist, that come from the connection of base pairs to sugar–phosphate chains of roughly constant length.

Having completed our study of the patterns of base stacking in DNA, our final task is to explain how different values of roll R, slide S, and twist T *generate different kinds of double helix* – as seen, for example, in Fig. 2.7. Thus, suppose that the same values of roll, slide, and twist are repeated over and over again along a significant length of DNA, what kind of double helix will be formed? Once we understand this relationship between the 'internal' variables R, S, and T, which describe the base-stacking relationships, and the 'external' form of the resulting helical structure, we shall be able to understand how different sequences of bases in DNA can generate

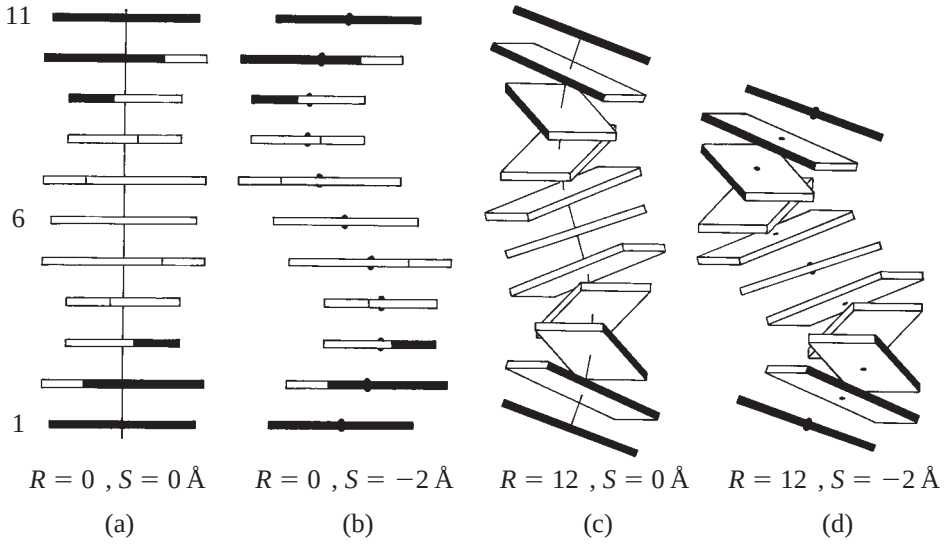

Figure 3.14 One complete helical turn of DNA having $T = 36°$, showing the effects of introducing uniform roll R or slide S at each step. Broadly, (a) corresponds to the 'B' form of DNA, while (d) corresponds to the 'A' form as shown in Fig. 2.7. Parts (b) and (c) correspond to structures intermediate between 'B' and 'A' which have, in fact, been seen recently in DNA crystals by X-ray diffraction.

different double-helical structures by favoring different values of R, S, and T at a local level.

Our analysis will consist solely of three-dimensional geometry. For the present, we shall assume that every step in a given helix has the same values of R, S, and T. An obvious way of proceeding might be to build some physical models of DNA, step by step, by means of a suitable home-made construction kit. That is exactly what we did ourselves in the first instance, when we were struggling to understand the geometry of double helices. From careful study of these models, we were able to derive the relevant equations that describe the geometrical form of the DNA in three dimensions, as functions of R, S, and T. Do not worry if you cannot follow the details of our presentation: many people have difficulty with the simpler two-dimensional geometry that can be drawn on a piece of paper! The crucial point to grasp is that the final results could be established firmly by means of a few hours of practical construction at a woodwork bench.

The various stages of our analysis correspond to the pictures shown in Fig. 3.14(a)–(d). Let us look first at Fig. 3.14(a), which shows the side view of a stack of 11 base-pairs, with $R = 0°$, $S = 0$ Å, and $T = 36°$ at each of the 10 steps. When roll and slide are both zero, then the helical geometry is very simple: neighboring