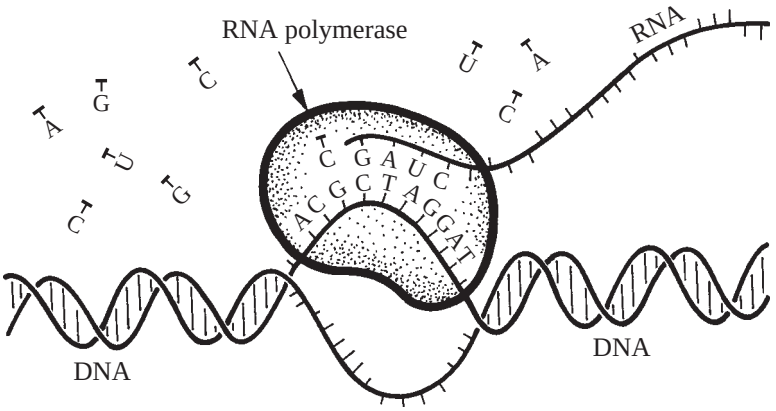
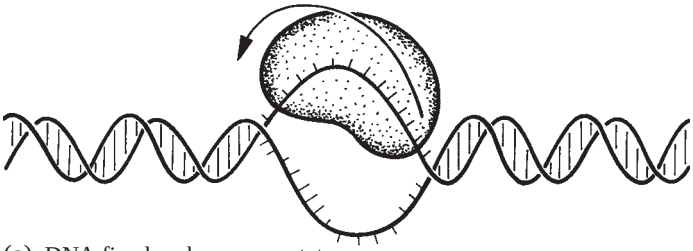


Figure 4.2 An RNA strand in the process of being made from single nucleotides by the enzyme RNA polymerase. One of the two DNA strands serves as a template for accurate synthesis of the new RNA strand, by the rules of Watson–Crick pairing. The actual length of the DNA–RNA hybrid, shown schematically here, is close to 8 base-pairs.

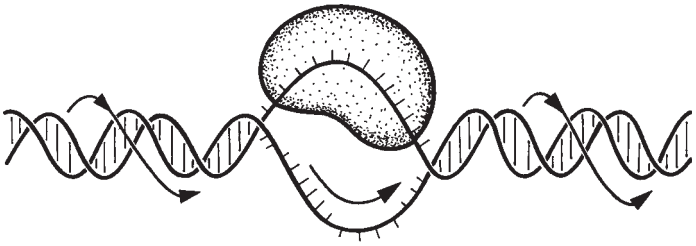
in living cells, where all four kinds of nucleotide are relatively abundant.

Now, you remember that DNA is helical, like a screw: so, as the RNA polymerase and its 'bubble' of unpaired bases move along the DNA, either the polymerase must screw *around* the DNA, or the DNA thread must screw itself *through* a stationary polymerase. Which of these alternatives will be favored in living cells? The two competing models are shown in Fig. 4.3, and it is not obvious on first inspection which might be right. Actually, James Wang and Leroy Liu have shown that the polymerase often remains somewhat stationary, as in Fig. 4.3(b), while the DNA screws through it. There are good reasons why this should be so. First, it will be hard for the big polymerase protein to move rapidly through the sticky, viscous fluid of a cell nucleus, in order to rotate about the DNA by 10 times per second; the thin, wiry DNA can rotate in the fluid much more easily. Second, the polymerase drags behind it a long RNA 'tail' as it goes about its duties, as shown in Fig. 4.2. It would be very hard for this long tail of perhaps 500 to 1000 nucleotides to follow the polymerase round and round the DNA, at a rate of 10 times per second. Indeed, some experiments suggest that in a test-tube the polymerase finds it easier to rotate about the DNA if the RNA tail is cut off by another enzyme (called RNAase) as it is being made.

If you have ever been out fishing, and the current twists your bait around the end of your line, you will know that there will be an unholy mess. Similarly, it would cause big trouble if Nature did not



(a) DNA fixed, polymerase rotates



(b) Polymerase fixed, DNA rotates

Figure 4.3 Alternative schemes for the copying of DNA into RNA by the enzyme RNA polymerase. In (a) the polymerase screws around a stationary DNA, while in (b) the DNA screws through a stationary polymerase.

find some way to relieve the twisting stress of DNA during its passage through a polymerase. In fact, Nature has invented several different kinds of de-knotting enzyme, called 'topoisomerases',[1] to get rid of the excess DNA twist. These enzymes act in very subtle ways, and no one is certain how they work on a near-atomic scale. They do however cut either one or both of the two sugar–phosphate chains to allow some kind of motion in the DNA to relieve torsional stress; then they re-connect the broken parts and thus leave an intact but relaxed DNA. Repeated application of this process to a tangled piece of DNA will eventually unravel it. We shall discuss the tangling (or supercoiling) of DNA further in Chapter 6.

Many clinically-used anti-cancer drugs, for example 'doxorubicin', are directed against the de-knotting enzymes. The cancer cells, which are growing out of control, can't divide if you poison their de-knotting enzymes. Some of these anti-cancer drugs 'trap' the topoisomerases in the middle of their cutting–relaxing–rejoining cycle, so that they cannot re-connect the strands of the DNA. Yet apparently, those cancer cells which survive the first few rounds of treatment with these drugs somehow find a way to grow with only a very low level of de-knotting enzymes; or else turn on new genes which help to expel the drug from the cell; and so they are eventually able to resist further treatment.

So it certainly doesn't look very promising, if you are a biological scientist today, to look for more anti-cancer drugs that poison topoisomerases. In fact, the most successful new anti-cancer drug to be discovered recently is called 'taxol' or 'docetaxol'; and it works by preventing the long cellular structures known as 'microtubules' from disassembling after they are made during cell division, when they help to carry one copy of each chromosome to each new cell – the process of separation mentioned on p. 4.

Apart from doxorubicin and taxol, the third most promising anti-cancer drug is called 'cisplatin' or 'carboplatin'. It was found in 1969 by Barnett Rosenberg, who was studying how bacteria grow in an electric field. He found that certain electrodes cause bacteria to grow long and thin, like spaghetti. By lots of detective work, he eventually worked out that the platinum in his electrode was combining with ammonia in his buffer, to make a platinum–ammonia compound that prevented cell division. This compound had been sitting on people's shelves for over 100 years, but no one had ever thought it would cure cancer. Anyway, he tested it on people, and it worked in some cases. Apparently, the platinum atom uses two of its ligand-binding sites to cross-link purines in GG or GA steps; which causes the DNA to bend, and so prevents replication.

We mentioned above that DNA unwinds into a bubble of 15 to 20 bases, as RNA polymerase copies one strand of the DNA to RNA. How is this accomplished? It turns out that the hard step is to unwind the double helix to begin with, in just one location. After that has been done, the bubble of DNA can travel with relative ease to some other location along the length of the molecule. It is therefore of crucial importance to understand how the cell tells its DNA to unwind in certain, specific locations. These events are then responsible for the large-scale synthesis of RNA from many particular genes on the DNA; and so they are responsible, ultimately, for the kinds of protein which are found in any cell. A similar kind of unwinding takes place at the start-sites for replication, where both strands of the DNA are copied into DNA; but there the enzymes are different.

What would you do if you were faced with the problem of unwinding the DNA from a cell in many specific locations? Would you make a protein with a little shovel attached at one end, to dig or pry a hole in the DNA? Some DNA-unwinding proteins do have little 'shovels', in the form of flat, oily amino acids such as phenylalanine, tyrosine, and tryptophan, which can make holes and then insert themselves between the base-pairs, and thereby effect unwinding of the helix. They would act just like the ethidium bromide molecule shown in Fig. 2.9, which converts the DNA locally from a helix into a partially untwisted ladder. Other DNA-unwinding

proteins perhaps have clefts on their surfaces into which only one of the two DNA strands can fit, so that they bind to a single strand in preference to a double helix, and so unwind the molecule that way. Finally, yet other proteins might curve the DNA around themselves in the form of a telephone cord or a coiled bed-spring, so that the DNA can unwind as it vibrates like a concertina in solution. There is some evidence for all of these kinds of unwinding in the cell; yet only recently has there been much knowledge at a detailed, atomic level of what happens in any case.

All of these unwinding events involve the binding of some protein to DNA. Much progress has been made recently at visualising the complexes of these proteins with DNA, using X-ray crystallography or other methods; yet still more is known about the role of DNA in such interactions than about the role of the protein. As a general rule, Nature makes the double-helical connections very weak in places where unwinding must begin, so that it won't take much energy to break open the DNA base-pairs, and thereby separate the two strands from one another. You will recall from Chapters 2 and 3 that the main determinants of the stability of a double helix are the number of hydrogen bonds between the bases in a base-pair, and the extent of base-to-base overlap. An adenine–thymine pair has only two hydrogen bonds, while a guanine–cytosine pair has three (see Fig. 2.11); so a series of adenine–thymine pairs will make the double helix less stable, as shown in Fig. 4.4(a). Furthermore, it has been shown by many careful experiments that pyrimidine–purine sequences have

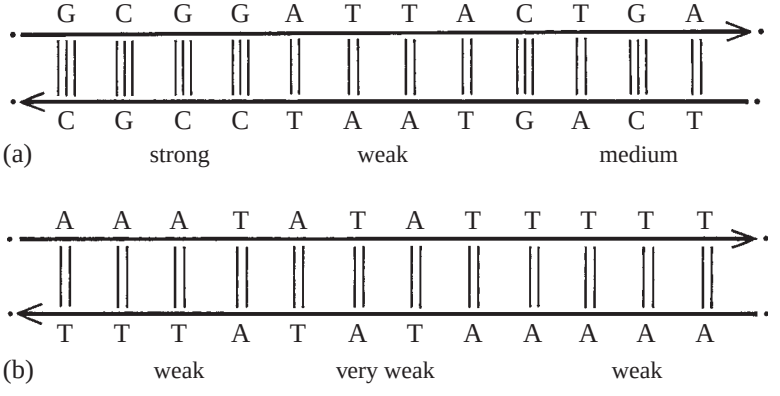


Figure 4.4 A two-part explanation for the ease of DNA unwinding at any TATA sequence. In (a), the double or triple cross-chain lines represent 2 or 3 Watson–Crick hydrogen bonds, respectively; and the double helix is less stable with 2 bonds than with 3. In (b), the TA steps (of kind pyrimidine–purine) cause a further weakening of the cross-chain connections, because they can easily unwind to yield low twist.

the least stability with regard to base-to-base overlap. These are the steps TA, TG, CA and CG. Why should this be so?

In the last chapter, we explained how to describe DNA in terms of the roll R, slide S, and twist T at any base-pair step. Pyrimidine–purine steps are special in this description because they can adopt either of two stackings: a high-slide form where the purines slide apart from one another (Fig. 3.10), or a low-slide form where the purines slide on top of one another (Fig. 3.11). Now the base-pairs are connected to sugar–phosphate chains in such a way that *low slide* leads to *low twist*. Thus, in the example shown in Fig. 3.13, a slide of $S = -2$ Å leads to a low twist of $T = 28°$ or thereabouts, as compared to $T = 34$ to $36°$ for other DNA. So it may be that the amount of energy required to unwind the DNA further, say to $T = 10°$ or $20°$, will be less for a pyrimidine–purine sequence than for other DNA, because the base-pairs there are already unwound to a significant degree; or perhaps they can unwind more easily under stress than at other sequences into a stable, low-twist form.

In any case, there seems to be little energetic barrier to low twist at a pyrimidine–purine step, for reasons that may not yet be fully understood.

Which, then, are the most easily unwound sequences in DNA? Simply those which combine the two characteristics we have been describing: few hydrogen bonds as in A–T pairs, and low twist as in pyrimidine–purine steps. In other words, DNA unwinds most easily at AT-rich regions that have many pyrimidine–purine steps, as shown in Fig. 4.4(b). The prototypic 'weak' sequence is something like 'TATATATA' or 'TAATAATAA', where TA is the pyrimidine–purine step of low twist.

This is not just a theoretical, hand-waving argument: there is actually strong experimental evidence for a low-twist intermediate in the unwinding of TATA-type sequences. It seems that such sequences unwind to 'cruciform' configurations more easily than other DNA. Now a cruciform is just a big 'bubble' of DNA which can be trapped and studied, because each strand of the bubble folds back and pairs with itself to form a double helix, as shown in Fig. 4.5. James McClellan and David Lilley set out to determine the reason for this odd behavior. They found that the unwinding of DNA into a cruciform at TATA-like sequences was being catalysed by partial unwinding of the DNA from 10.5 to about 12 base-pairs per turn, or from $T = 34°$ to about $T = 30°$, before the cruciform appeared. Their observations agree closely with what we expect from theory, as explained above.

Biologists identified the TATA-type sequences almost by accident, by determining the sequences of DNA in places where the double helix must unwind for transcription to begin. These broadly defined

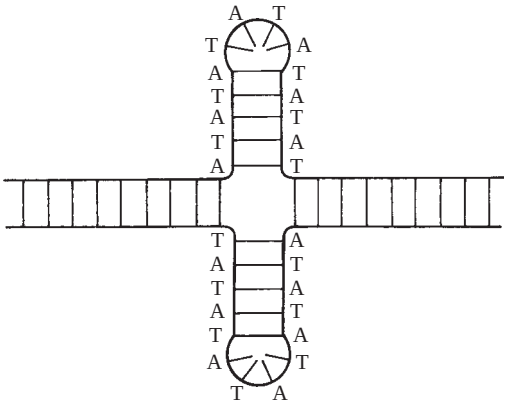


Figure 4.5 A 'cruciform' structure which has been 'extruded' from a suitable sequence of A and T bases. The DNA double helix untwists from 10.5 to 12 base-pairs per helical turn before the cruciform appears, owing to the easy unwinding of TATA sequences. As a consequence of this initial untwisting, the cruciform extrudes there more readily than at other sequences.

regions of DNA are called 'promoters', because they promote RNA synthesis from a nearby gene without being transcribed themselves. The biologists found in many cases a 'TATA' or similar sequence at or close to the required site of unwinding within each promoter. Then they drew a box around this sequence on their computer output, to show how important it might be. Hence, many textbooks call the weak region of DNA in some promoters a 'TATA box'. The majority of promoters contain a TATA or related sequence 'upstream' of the site where unwinding into separate strands begins; and this evidently plays an important part in starting the operation of the polymerase.

Thus, in most *bacteria*, an RNA polymerase protein binds to the TATA region: one part of that polymerase, the 'sigma initiation factor' actually contacts the TATA sequence, 10 base-pairs upstream from the start-point for making RNA. This sequence is where unwinding of the two strands starts.

But in the cells of our body the TATA-like 'box' performs a different function. Here, there is almost always an *additional* protein called 'TBP' (for 'TATA-binding protein') which binds to TATA-like sequences near the start of genes (and is sketched in Fig. 4.14, below). This protein is available in moderate amounts in pure form, and several groups of workers have found that it untwists the TATA sequence, as part of preparation for the polymerase, by about one-third turn. Here, the TATA-sequence preceeds the site of unwinding by about 25 base-pairs. We shall discuss the action of polymerases further in Chapter 6.

The TATA sequence is known to adopt at least two different helical forms: one with low twist and high roll, as seen when bound

to TBP (see below), and some other proteins (also discussed below); and a second form with just 7.5 to 8.0 base-pairs per turn rather than 10 to 12 for normal DNA. That last helical form has been called the 'D' form of DNA in X-ray studies of DNA fibers, at low resolution. Until we establish the molecular structure of such a strange helix, we cannot even guess whether it might be used in biology.

As a last note on the subject of TATA-type double helices, we refer to the three triplets in the Genetic Code (Table 1.1) which code for 'stop', and are usually called 'stop-codons': TAA, TAG and TGA. The process of assembling amino acids into a protein chain comes to a halt at those particular triplets, because there are no transfer-RNA molecules which can recognize them. One of the stop codons is UAA in the messenger-RNA (like TAA in DNA), and if a transfer-RNA molecule for UAA were to exist, it would form a specially weak bond with this triplet (see Fig. 4.6) for precisely the reasons given above. We think that it is no accident that this specially weak-binding triplet does not correspond to a specific amino acid, for the hypothetical transfer-RNA would be unreliable at recognising it. The same considerations probably apply also to the two other stop-codons UAG and UGA, which are moderately weak in their capacity for base-pairing.

That gives an overview of twisting in DNA. Now let us look at the *curving* of DNA. In principle, DNA could curve either in a plane, like a banana, or else in three dimensions, like a coiled bed-spring. It is hard to learn about three-dimensional curvature at the first attempt, however; so we restrict ourselves in this chapter to the

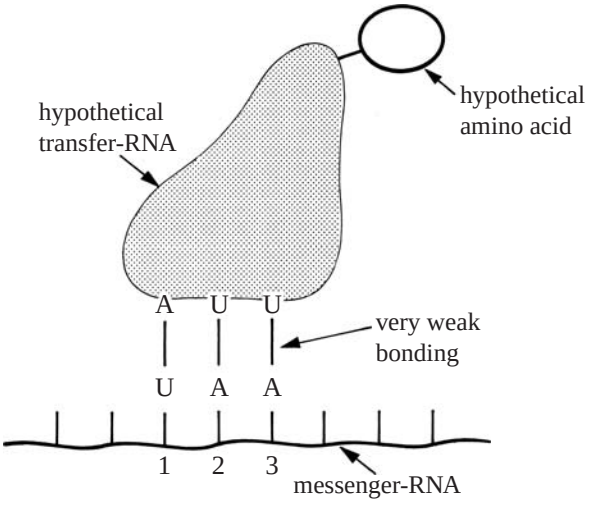


Figure 4.6 All of the 'STOP' signals in the Genetic Code (UAA, UAG, UGA, see Table 1.1) would correspond to very weak or moderately weak pairings between a hypothetical transfer-RNA and the messenger-RNA.

curving of DNA in a plane. The next chapter will deal with curving in three dimensions.

First let us forget about DNA for a moment, and think instead about our gardens. Many people have small, flowering trees in their suburban gardens, and they like to use slabs to build neat paths around them, as shown in Fig. 4.7. There are two ways to do this. One way is to buy special slabs that are made in a wedge-shape, with one end wider than the other. Such slabs will make a circle of a particular radius, just as stone 'voussoirs' will build a round arch of specific radius over a window or a door. The other way is to make the path out of rectangular slabs, and to fill the narrow triangular spaces between them with pebbles. An advantage of this scheme, of course, is that one can change the curvature of the path at will, making it circular, or straight, or of variable curvature as the plan requires, just as in the picture.

What do we mean by the *curvature* of a garden path? In Fig. 4.7 the path is obviously more curved where the angle between consecutive slabs is larger. Where this angle is zero, the path is straight; that is, it has zero curvature. It would therefore be sensible to define curvature in terms of the angle between successive slabs. Thus the curvature of the semicircular part of the path in Fig. 4.7 may be described as 15° per slab (or per step) because the path turns through 180° in 12 steps.

Now, if we lay a garden path with 15° between every pair of slabs, we shall eventually complete a circle. And the radius of the circle will be smaller if we make the angle of curvature larger, and vice versa. We can easily calculate the radius of the semicircle in Fig. 4.7, as

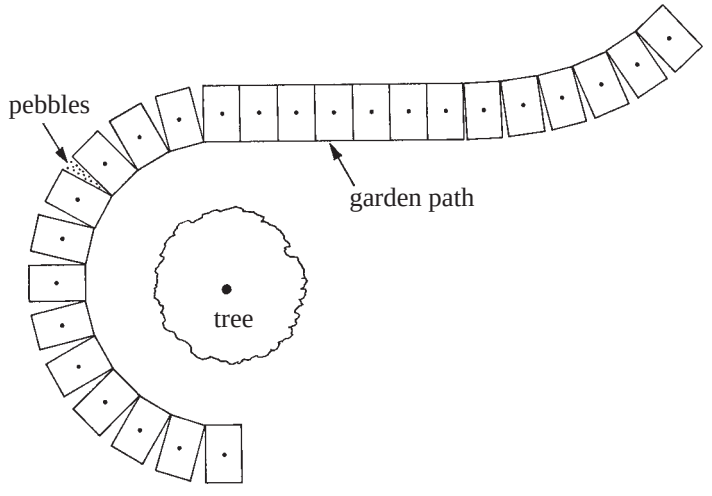


Figure 4.7 Curvature of a garden path in a plane, around a tree.

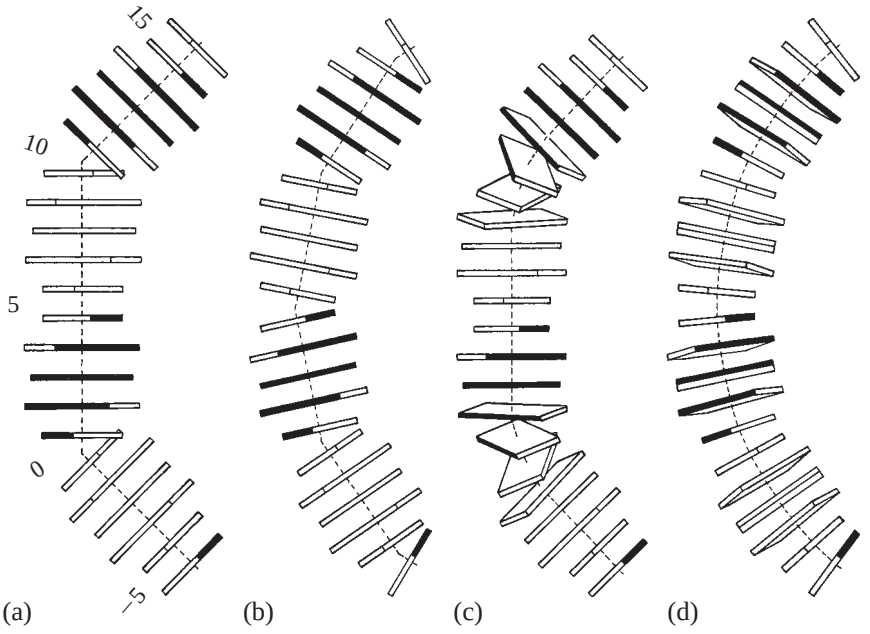


Figure 4.8 Two complete helical turns of DNA, with a curvature of 45° per turn, or 4.5° per step on average. Such tight curvature may be achieved, in principle, by any of the distributions of roll angle shown in parts (a) to (d).

follows. Suppose the width of each slab is 0.5 m. If the angle between slabs is 15°, there will be $360/15 = 24$ slabs in a circle, with a total circumference of $24 \times 0.5 = 12$ m, and hence a radius of $12/2\pi = 1.9$ m. This is the radius of the inner edge of the slabs; similarly we could find the radius of the center-line of the path, by using the center-to-center separation of the slabs, together with the angle between slabs.

The curvature of DNA is more complicated than that of a garden path, because DNA is three-dimensional; yet it follows broadly the same principles, because DNA is also made in discrete steps, being built up from base-pairs as we have seen. Thus, imagine that a piece of DNA of length 80 base-pairs, or 8 double-helical turns, has been bent into a 360° circle. That is just about the degree of curvature by which DNA wraps around proteins in the cell nucleus (Fig. 1.5). If you were to make the same circle from just eight tiny slabs, then the angle between slabs would have to be $360°/8 = 45°$. This angle is precisely equivalent to the roll angle R in DNA, as described in Chapter 3. If we wish to make a circle from eight helical turns of DNA, we can do so by putting one roll angle of 45° in each double-helical turn: that will make a fine circle.

Such a scheme is shown in Fig. 4.8(a), which shows 20 base-pair steps that make up one-quarter of our 80-base-pair circle. In order to

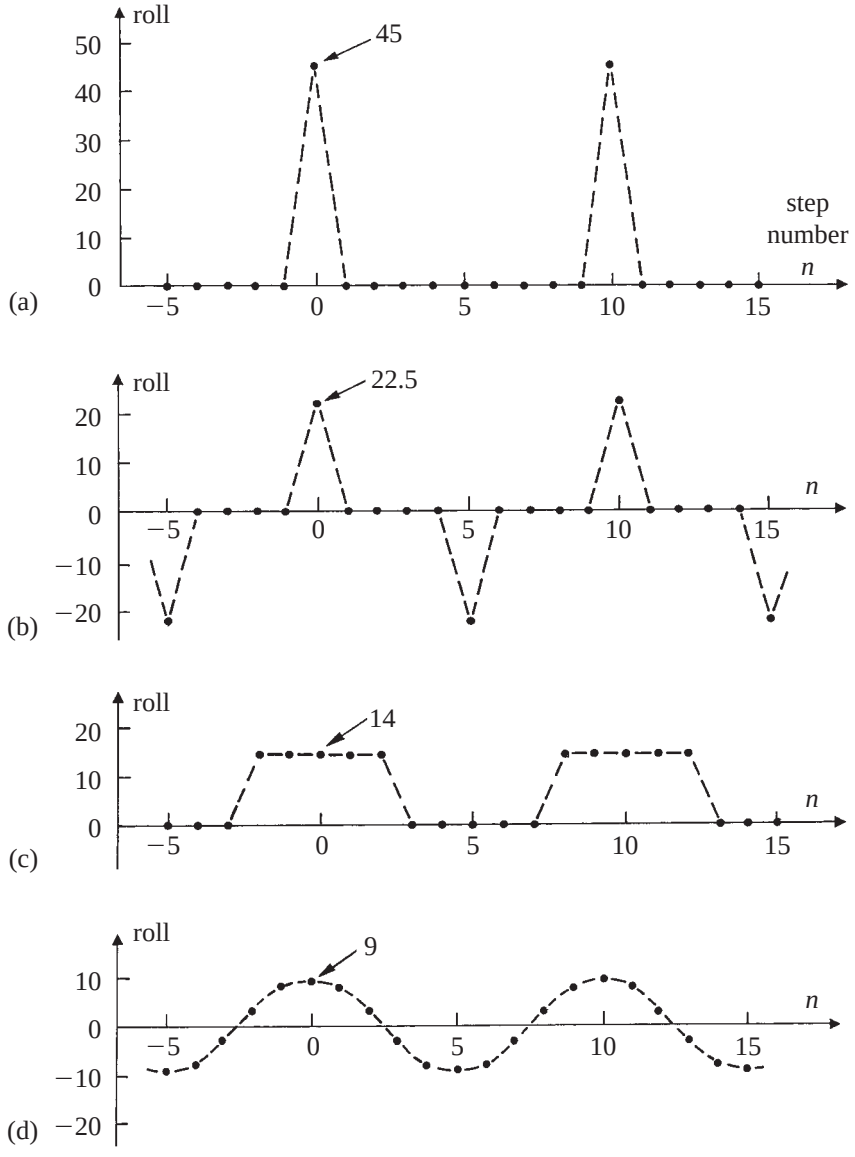


Figure 4.9 A plot of roll angle *versus* step number for the four cases shown in Fig. 4.8, each of which curves the DNA by 45° per helical turn.

convert straight DNA into curved DNA, we have introduced a roll angle of 45° at the two steps labeled 0 and 10, respectively, on the left-hand side of the drawing. These roll angles open up the minor-groove edges of the base-pairs, which are colored black, to yield a total curvature of $2 \times 45° = 90°$ over the 20 steps. The same type of curvature is shown schematically in Fig. 4.9(a), which gives a plot of roll angle