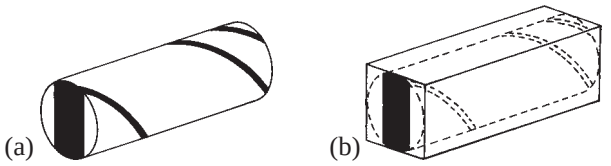

Figure 5.2 (a) One complete turn of double-helical DNA, showing two sugar–phosphate chains along the outside, and a black rectangle as the first base-pair: the whole structure has been drawn as a cylinder. (b) The same as in (a), but now the DNA has been encased in a semitransparent solid block.

k alone gives a plane circle, while t alone gives a narrow spiral like a twisted ribbon.

This story gives us a good initial understanding of how curvature k and twist t influence the path of DNA through space. Let us now ask, first, how k and t influence the shape of any small segment of DNA; and second, how these small segments can be joined together to make different paths through space, like those flown by our friend in the airplane.

We can represent any small segment of DNA as in Fig. 5.2(a) by a short cylinder with the sugar–phosphate chains along its outside. This particular segment contains exactly one double-helical turn of 360°, or 10 base-pair steps. The first and last base-pairs, numbers 0 and 10, can be drawn as black rectangles on the two ends of the cylinder, to demonstrate that they are parallel to one another, and are vertical in the drawing. Only one of these two black rectangles can be seen in the perspective of Fig. 5.2(a), so you have to imagine that there is another one on the far end of the cylinder.

An even simpler version of the same thing is shown in Fig. 5.2(b), where the single helical turn of DNA is now represented by a solid block. The two ends of the block are perfectly aligned, because the DNA has twisted by 360° along its axis in going from one end to the other; and the block is straight as in (a). You can imagine that the cylinder of (a) is converted into the block of (b) by first glueing squares to the ends of the cylinder, and then filling in the space between the squares with some sort of semi-transparent jelly which enables us to see, darkly, the sugar–phosphate chains buried inside. From now on we shall mainly disregard the twisting of these strands within the block, and will usually draw the block as in Fig. 5.3(a), without any hint of what is actually hidden inside it. In this figure both the curvature k and the twist t of the block are precisely zero, and all six faces of the block are plane squares or rectangles.

Figure 5.3(b) shows a block that has curved by 10° to the left, or equivalently by $k = 1°$ per step over 10 steps. It looks something like a banana. The upper surface is obviously curved, while the shaded

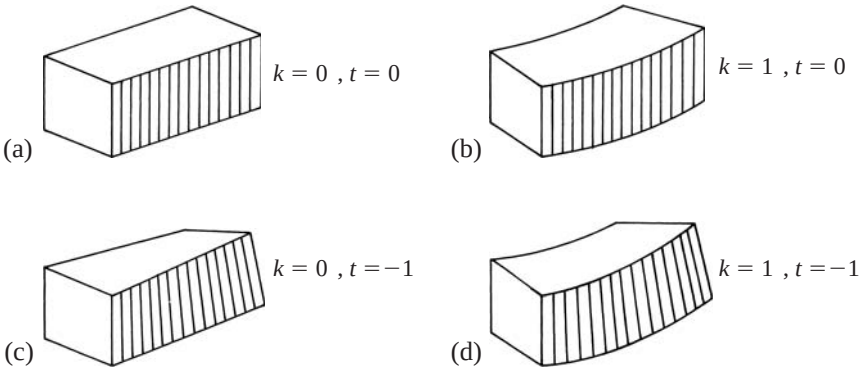

Figure 5.3 Three possible ways by which the solid block shown in Fig. 5.2(b) can change its shape, by the application of uniform curvature k and twist t. In this figure, curvature k and twist t are given in units of degrees per base-pair. Since there are 10 base-pair steps, the total curvature is 10° in (b) and (d), while the total twist is −10° in (c) and (d).

surface might form part of a cylinder. Presumably, the roll angles R of the DNA within the block have changed on going from (a) to (b), so that the best-fit cosine wave to these roll angles, as described in Chapter 4, now sweeps out an angle of 10°. The twist t is still zero, because the two ends of the block remain aligned: note that the closely spaced lines on the shaded surface are all strictly parallel to one another.

Figure 5.3(c) shows a block that has twisted in a negative or left-handed sense by 10° from end to end, or equivalently by $t = -1°$ per step over 10 steps. Presumably, the local twist angles T between successive base-pairs in the DNA have decreased from $T = 36°$ to $T = 35°$ on average, in going from (a) to (c). Thus, the 10 steps of hidden DNA now twist only by a total of 350° rather than 360°, leaving a deficit of $350° - 360° = -10°$; and this is what shows in the picture. In (c) there is no curvature: the top surface, though now twisted, is still rectangular.

Lastly, Fig. 5.3(d) shows a block that is both twisted and curved; it looks like a slightly twisted banana. Both roll R and twist T have changed at the base-pair level to produce this result, in the same ways as described for parts (b) and (c) previously; but now R and T have both changed simultaneously.

We have used the word 'twist' in two senses here, with two symbols t and T. Lower-case twist t is the difference in total twist from 360° after one unit-length of helix. It amounts to −10° after 10 steps in (c), as explained above. Upper-case twist T is the twist of any base-pair step locally. For example, $T = 36°$ in both (a) and (b), but

$T = 35°$ in (c) and (d). It follows, therefore, for any unit of n base-pair steps with uniform twist T:

$$t = nT - 360°.$$

Here $t = (10 \times 35°) - 360° = -10°$ in both (c) and (d) of Fig. 5.3. What do we mean by a 'unit of n base-pair steps'? Basically, we mean that the internal structure of the DNA is identical in each successive set of n steps – as it is likely to be, for example, if the *sequence* of the DNA repeats every n steps.

In summary, curvature k and twist t can describe the shape of any small segment of DNA on a local scale. These values of k and t come from variations in base-step roll R and twist T, respectively. We explain below how to calculate k and t from given values of roll R and twist T, for the base-pair steps of practically any segment of DNA.

Now you must use your imagination. What happens if we take many identical blocks of the kind shown in any one part of Fig. 5.3, and join them together, end-to-end, over a long distance? Using blocks all of type (a), we would get a straight, untwisted rod; using blocks of type (b), we would get a plane circle; while the blocks of type (c) would build a straight but twisted rod. Finally, the blocks of type (d) when joined together would make a broad spiral that we might call a 'superhelix', if we remembered the DNA hidden within it. Such a spiral will rotate counterclockwise as it goes forward, and so be left-handed, because the sign of twist t is negative. If t were positive, we would get a right-handed superhelix. This would happen, for example, if the local twist T became 37°, so that $t = nT - 360° = (10 \times 37°) - 360° = +10°$.

In fact, we do not have to rely entirely on our imagination for these constructions. We can *calculate* the path of the DNA through space for different values of k and t, by using certain geometrical formulas. Therefore, let us now derive these simple formulas.

A typical left-handed superhelix is shown in Fig. 5.4(a). This superhelix might be made by joining together many blocks like the one shown in Fig. 5.3(d). It winds uniformly like a ribbon about a vertical cylinder of radius r, at an angle α with respect to the horizontal at any point. It makes a left-handed spiral as it winds round the cylinder. It turns by 360° around the vertical axis of the cylinder after a vertical distance p, which is known as the superhelical 'pitch'. Finally, it has a contour length, not indicated in the picture, of N^* base-pair steps in one 360° turn. If an ant were to crawl along the path of the superhelix for any full 360° turn around the cylinder, and then measure or count over how many steps it had crawled, that would be N^*.

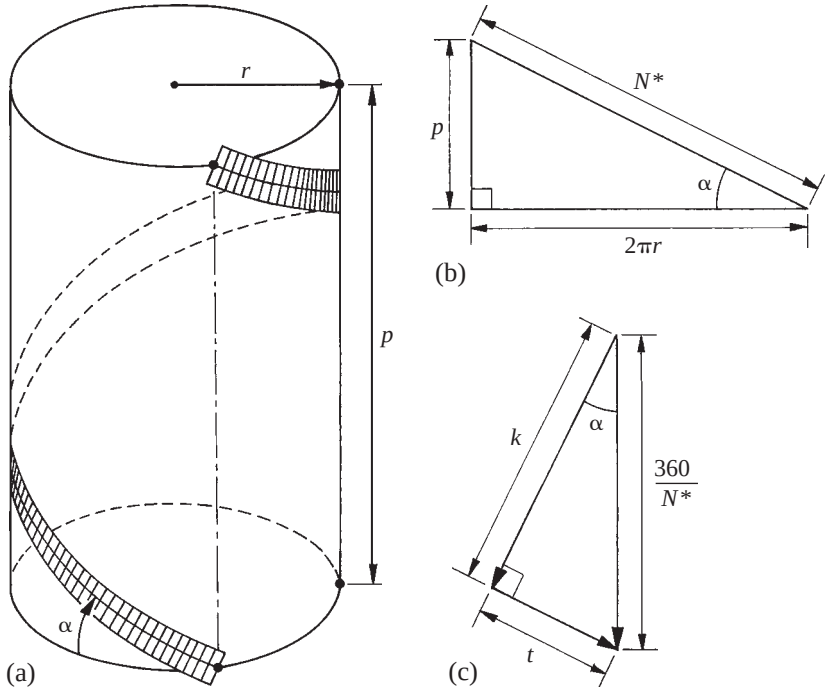

Figure 5.4 Curvature and twist in a spiral ribbon. In (a), a left-handed spiral ribbon of pitch angle α and radius *r* goes through one turn every *p* units of distance along the cylinder. In (b), the surface of the cylinder has been unwrapped on a smaller scale, to show geometrical relations among α, *r*, *p* and contour length *N**. In (c), the vector sum of curvature *k* and twist *t*, on a very small scale, gives the amount by which each step of the ribbon rotates around the axis of the cylinder. Note that triangle (c) is similar to triangle (b).

The usual way to analyse these parameters r, α, p, and N^* is to imagine that we can 'unroll' the cylinder onto a piece of paper, in the same way that you can unroll the cardboard cylinder on which paper towels are wrapped. This is called a 'cylindrical projection', and it is shown, to a smaller scale, in Fig. 5.4(b). The path of the superhelix becomes the diagonal of a right-angled triangle, while the pitch p and the circumference $2\pi r$ make up the other two sides; they are related by the angle α. It is a simple matter to find all of N^*, p, r, and α from any two of these four parameters; but we shall not do so here.

All of this is very well-known geometry; but the next part is not so well-known. Consider any small part of the superhelical path, say just 2° of the total 360° for each turn around the cylinder. For example, when $N^* = 180$ base-pairs, any single base-pair makes up $360°/N^* = 2°$ of rotation about the circumference, as seen in a view along the axis of the cylinder. In this case, the values of curvature k and twist t for the spiral ribbon are related to $360°/N^*$ as shown in

Fig. 5.4(c). Here, the total rotation of 2° about the axis is represented by the vertical arrow: so curvature k is the cosine component of it, while twist t is the sine component. This result follows from a branch of mathematics known as 'differential geometry', for the special case of a ribbon wrapped about a cylinder. In simple terms, the curvature k tells how far the superhelix should *curve* around the cylinder in the manner of Fig. 5.3(b), and t tells us how much the ribbon should *twist* up the cylinder in the manner of Fig. 5.3(c). The sum of the two effects (it is a vector sum, so is represented properly by the triangle) gives the total rotation about the axis of the cylinder.

Several interesting equations follow from the triangle of Fig. 5.4(c):

$$\tan \alpha = t/k$$

$$k = (360°/N^*) \cos \alpha$$

$$t = (360°/N^*) \sin \alpha$$

$$(360°/N^*)^2 = k^2 + t^2$$

We can now relate the diagrams in Fig. 5.4(b) and (c) to each other, and thereby get the overall shape of the superhelix from any given values of k and t. By similar triangles we have

$$p/N^* = t/(360°/N^*)$$

Substituting for N^* from the last of the four equations above, we obtain

$$p = 360°t/(k^2 + t^2).$$

Similarly,

$$2\pi r = 360°k/(k^2 + t^2).$$

Let us consider as a specific example a case where $k = 1°$ and $t = -1°$ per base-pair. Then $\alpha = -45°$, meaning that the superhelix is left-handed as in Fig. 5.4(a), since t is negative, and climbing at 45°. Also, contour length $N^* = 360°/\sqrt{2} = 255$ base-pairs, while pitch $p = 180$ base-pairs. Finally, the circumference of the cylinder $2\pi r = 180$ base-pairs, so radius $r = 29$ base-pairs. The final form of the superhelix is similar to that shown in Fig. 5.4(a), except that the angle α is larger.

In this particular example k and t both had the same magnitude. In a case where the magnitude of t is small compared with k, the formulas give us a low value of α; this means simply that the superhelical coil is almost flat, like a circle going round and round the same path. However, when k is small compared with t, the formulas give $\alpha \approx 90°$. This means that the superhelix takes the form of

a highly extended spiral, almost like a straight line but with a small superhelical 'wobble'.

All of this seems very straightforward, once we have learned how to use the formulas. But how can we get k and t from the sequence of the DNA? Unless the base sequence is very regular, k and t will vary from one double-helical turn to the next. In that case, the path of the DNA will not describe a regular superhelix, and so the formulas which we have just derived will be useless. But if the sequence of the DNA does repeat exactly – or almost exactly – once every double-helical turn, then the DNA will form a regular superhelix and our formulas will hold good.

Fortunately, scientists have done many experiments on the structure of 'repeating-sequence' DNA, because they can make it rather easily. They synthesize chemically one small part of the DNA, say 10 base-pairs of a defined sequence, and then they join these units together to make a long polymer, just as we imagined when we were studying Fig. 5.3. Having made such repeating-sequence DNA, they can study its structure by the methods of electron microscopy and gel electrophoresis. Electron microscopy shows the shape of the DNA directly, although at low resolution, while gel electrophoresis measures indirectly the 'apparent volume' of the DNA cylinder. We shall discuss this method of analysis in Chapter 9, but for the present we may say that the curved DNA shown in Fig. 5.4(a) could be enclosed in a cylinder of larger volume than for straight DNA of the same length; and so in a gel, curved DNA would come into contact with more gel fibers, and hence go more slowly. That is precisely what is observed in experiments with gels.

Still, it is nice to see the DNA directly, without having to worry about indirect measurements of its volume by gel electrophoresis. For that reason, it is sensible to study repeating-sequence polymers by electron microscopy. Some pictures of curved, repeating-sequence DNA as obtained by this method are shown in Fig. 5.5. The first five frames in Fig. 5.5(a) show a collection of similarly curved DNA molecules of size about 1000 base-pairs. They lie in various shapes on the support 'grid', having been flattened onto two dimensions in preparation for microscopy. The last frame in Fig. 5.5(a) shows two DNA rings or 'plasmids' of length 3000 base-pairs, of ordinary (i.e. not repeating-sequence) DNA, as controls for the appearance of DNA with little or no curvature; they were prepared for microscopy under identical conditions. Figure 5.5(b) shows the outlines of curved DNA molecules from (a). Each of the curved DNA molecules has a wiggly, snake-like appearance compared with the two DNA plasmids. The 'wiggle' of the curves on the photographs can be fitted by a sine-wave of contour length $N^* = 533$ Å and pitch $p = 385$ Å, as an

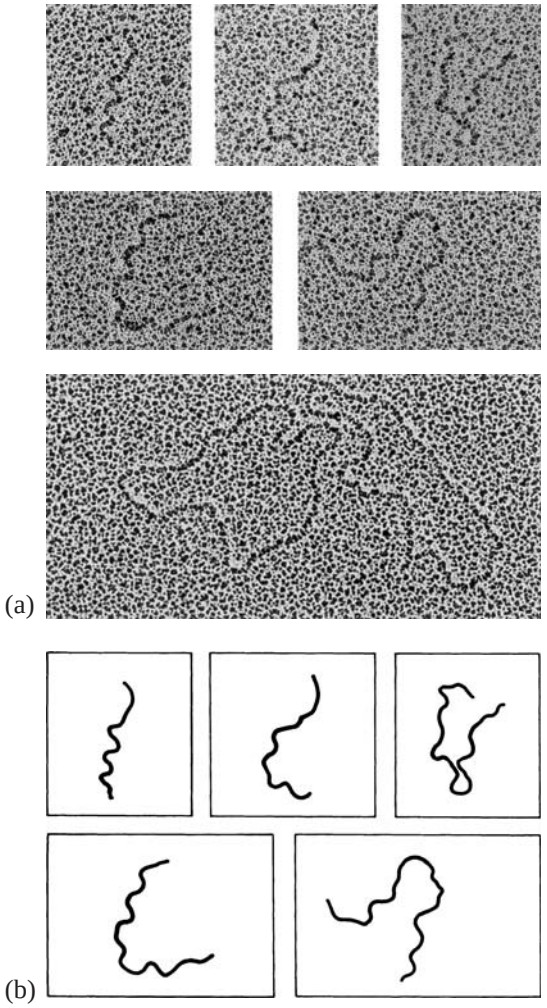

Figure 5.5 Electron micrographs of repeating-sequence DNA. The first five frames in (a) show an assortment of pictures of repeating-sequence, curved DNA of size about 1000 base-pairs. The last frame in (a) shows two DNA plasmid or ring molecules of size 3000 base-pairs, that have been included as controls. In (b), approximate tracings of the curved DNA molecules from (a) are presented. Courtesy of Margaret Mott.

average over many molecules. The amplitude of the 'wiggle' is obviously unreliable as a measure of the original three-dimensional structure, because the DNA superhelix was flattened significantly when placed onto the grid for microscopy; but the values of N^* and p should be representative of the original conditions in solution. From the values for N^* and p, and the triangle shown in Fig. 5.4(b), we can calculate that $\alpha = \pm 46°$, while radius $r = 59$ Å, before flattening onto the grid. The length per base-pair in these pictures has

been measured as 3.0 Å, so we can express the dimensions of the DNA superhelix in units of base-pairs as $N^* = 178$ base-pairs, $p = 128$ base-pairs, and $r = 20$ base-pairs.

Figure 5.5 is a good piece of experimental evidence about the shape of curved, superhelical DNA. But can we calculate the values of N^*, p, and r independently, directly from the base sequence of the DNA, by using a suitable theory? It is easy to determine from the superhelical parameters listed above that $k = 1.40°$ per base-pair and $t = \pm 1.45°$ per base-pair; but how can we derive k and t values from the base sequence? Note that the electron microscope pictures do not indicate whether the original superhelix was left-handed or right-handed, and so we do not know from the experiment whether t is positive or negative.

The sequence which was used to make these polymers was of the kind …AAAAAANNNNAAAAAANNNN…, where N = C or G, mainly. This sequence repeats once every 10 base-pairs, as AAAAAANNNN. Now, it is known that the average local twist T for an AA step is close to $T = 35°$, while $T = 34°$ for other steps such as NN, NA, or AN. The slight difference in twist possibly comes from an extra hydrogen bond across the major-groove side of the AA step, as shown in Fig. 3.6. Therefore, we can calculate the overall twist t as $(5 \times 35° + 5 \times 34°) - 360° = 345° - 360° = -15°$ per 10 steps, or $t = -1.5°$ per base-pair. Not bad! We have already come close to the value for t of (plus or minus) 1.45° which was obtained by electron microscopy. Both of the values for T at particular steps, which were used in the calculation above, had been determined by gel electrophoresis or other techniques several years before the electron microscope pictures were taken.

But how can we calculate the curvature k from the base sequence? The roll angles R are thought to be close to 0° for an AA step, as against $R = +3.3°$ for the others as a broad average: again, these values come from gel electrophoresis, X-ray crystallography, or other techniques. Such approximate values of base-step roll are plotted in Fig. 5.6 against the step number, for this particular repeating sequence. As explained in Chapter 4, we must fit a cosine wave to the values of R plotted in Fig. 5.6, in order to determine the curvature k. We could draw a 'best-fit' cosine wave over the points in Fig. 5.6 by eye, in order to get a satisfactory approximate solution (see the broken line in the diagram), but really we would prefer to have a more systematic way of doing the calculation.

The most accurate method is to take what is known as the 'Fourier transform' of the roll angles over any period of steps 0 to 9 in the base sequence. This is just like what we did in Chapter 4, except that here we shall show the full mathematics. It doesn't matter

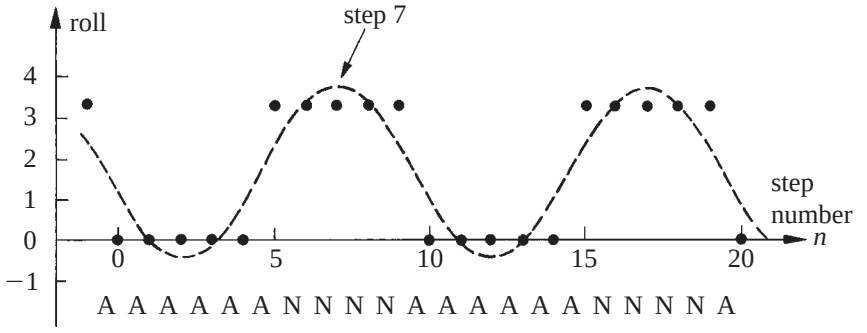

Figure 5.6 A plot of expected roll angles for the repeating-sequence, curved DNA studied by electron microscopy in Fig. 5.5. The curvature k may be calculated as 10.7° over any A_6N_4 repeat by taking a Fourier transform of these roll angles, as indicated by the broken line. Roll angles are most positive in the Fourier transform at step 7, where the minor groove of the DNA lies along the outside of the curve.

which step you assign to be step 0, or which step you assign as step 9, so long as all 10 steps of the repeated sequence (0, 1, 2, …, 9) are counted only once.

First, you must evaluate two sums, which are the correlations of the roll angles R with either a sine or a cosine wave; for this purpose it is convenient to write R_n for the roll R at any step n:

$$\text{first sum} = \sum_{n=0,9} R_n \sin(36\ n)$$

$$\text{second sum} = \sum_{n=0,9} R_n \cos(36\ n).$$

Thus, using the numbering of steps shown in Fig. 5.6, we find that:

$$\begin{aligned}\text{first sum} &= 0° + 0° + 0° + 0° + 0° + 3.3°(0.0) + 3.3°(-0.59) \\ &\quad + 3.3°(-0.95) + 3.3°(-0.95) + 3.3°(-0.59) \\ &= -10.2°\end{aligned}$$

$$\begin{aligned}\text{second sum} &= 0° + 0° + 0° + 0° + 0° + 3.3°(-1.0) + 3.3°(-0.81) \\ &\quad + 3.3°(-0.31) + 3.3°(0.31) + 3.3°(0.81) \\ &= -3.3°\end{aligned}$$

Then the curvature k is simply given by the square root of

$$\begin{aligned}k^2 &= (\text{first sum})^2 + (\text{second sum})^2 \\ &= (-10.2)^2 + (-3.3)^2 = 115; \quad \text{so } k = 10.7°.\end{aligned}$$

Our calculated value of $k = 10.7°$ for 10 steps, or 1.07° per base-pair, compares fairly well with the value of $k = 1.4°$ per base-pair as determined from the electron microscope pictures of the overall shape. We could also calculate from this theory that our repeating-sequence

DNA should make a left-handed superhelix of contour length $N^* = 195$ base-pairs and radius $r = 61$ Å, using the theoretical values of k and t. These are again fairly close to the experimental values of $N^* = 178$ base-pairs and $r = 59$ Å, given above.

Before we go on, let us consider two more features of the Fourier transform: its amplitude and its phase. First, the amplitude of the best-fit cosine wave in most of our examples is equal to the curvature k divided here by 5.0, which is the sum of $\cos^2\theta$ over steps 0 to 9. Thus, in Fig. 5.6, the amplitude or half-height of the dotted line is simply $10.7/5.0 = 2.1°$. We can say then that the total variation in roll R, for the best-fit wave, is $2 \times 2.1° = 4.2°$. This wave goes from a peak at $R = +3.75°$ to a trough at $-0.45°$ about a mean of $+1.65°$. Second, the phase of a Fourier transform tells us where the best-fit cosine wave is located in a left-to-right sense, relative to the numbering of steps in the sequence. It can be calculated as the

arctangent of the ratio (first sum/second sum).

That gives, in the case above, $\arctan(-10.2/-3.3) = 252°$. This shows us that the origin of the best-fit cosine wave, where its value is most positive, lies 252° to the right of step 0 in the sequence, or in Fig. 5.6 at step $(252°/36°) = 7$ (see the arrow). In other words, the base-step roll R is the most positive at step 7, and least positive (or most negative) at steps 2 and 12, which are 180° out of phase in either direction. In molecular terms, this means that the minor groove of the DNA lies along the outside of the curve at step 7, and along the inside of the curve at steps 2 and 12.

Now suppose we had made another, slightly different DNA molecule with repeating sequence, but this time with a repeat once every 11 base-pairs rather than every 10. For example, AAAAAANNNNN would have an 11-base-pair repeat, as compared with the 10-base-pair repeat for AAAAAANNNN studied above. Then we could calculate the overall twist t by adding up the local twist angles T over 11 steps, to yield $t = (5 \times 35° + 6 \times 34°) - 360° = 379° - 360° = +19°$. Our new superhelix would thus be right-handed, since t is now positive. We could calculate the curvature k by taking the Fourier transform of the roll angles over 11 steps, at intervals of $360°/11 = 32.7°$, using cosines and sines of angles $(32.7°n)$, to yield $k = 11.6°$. Here, we have calculated both t and k for a complete 11-step repeat. In smaller units of base-pairs, t and k would be $t = +19°/11 = +1.73°$ and $k = 11.6/11 = 1.06°$. Finally, from these values of t and k in degrees per base-pair, we could calculate the dimensions of the right-handed superhelix, by use of the formulas of the preceding section.

Table 5.1 lists values of k and t as calculated for a series of DNA molecules of repeating sequence, which includes the two sequences