Table 5.1 Calculated superhelical parameters α, N^*, and r (see Fig. 5.4) for DNA of various repeating sequences

Sequence	Repeat	k (°/bp)	t (°/bp)	α (°)	N^* (bp)	r (bp)
A_6N_2	8	1.00	−10.38	−84.5	34.5	0.5
A_6N_3	9	1.06	−5.44	−79.0	65.0	2.0
A_6N_4	10	1.07	−1.50	−54.5	195.4	18.1
A_6N_5	11	1.06	+1.73	+58.5	177.4	14.8
A_6N_6	12	1.03	+4.42	+76.9	79.3	2.9
A_6N_7	13	0.99	+6.69	+81.6	53.2	1.2

The curvature and twist (k, t) were calculated from the base sequence according to angles of roll R and local twist T given in the text.

already mentioned, namely A_6N_4 and A_6N_5. These calculated values of k and t have then been used to compute the superhelical parameters α, N^* and r for each sequence. Looking first at the two central columns of the table for k and t, we can see clearly that if we had made a 9-base-pair repeat such as A_6N_3, or a 12-base-pair repeat such as A_6N_6, then twist t would have become much larger in magnitude than curvature k. In the case of A_6N_3, you would get $t = (5 \times 35° + 4 \times 34°) - 360° = 311° - 360° = -49°$ per repeat, which gives $t = -49°/9 = -5.44°$ per base-pair. But the curvature k is still only 1.06° per base-pair, and so the pitch angle $\alpha = \arctan(t/k)$ becomes $\arctan(-5.44/1.06) = -79.0°$. This means that A_6N_3 would form a highly elongated, left-handed superhelix of contour length $N^* = 65$ base-pairs and radius $r = 2$ base-pairs.

Those calculations lead us to the conclusion that we can only expect to find DNA molecules in the form of a *broad* spiral when the DNA sequence-repeat is close to 10 or 11 bases, so that the twist t is low. Of course, if twist $t = 0$ exactly, then the DNA forms a plane circle of contour length $N^* = 360°/k$; and indeed many naturally occurring DNA molecules form plane circles.

How could we calculate k and t if the base sequence was more complicated than A_6N_4 or the related sequences discussed above? Previously we assigned roll R and twist T values only to AA/TT and 'other' steps. Thus, we set $T = 35°$ for AA/TT steps but $T = 34°$ for 'others'. This model is clearly too simple to account for all of the possible arrangements of base-pairs in DNA, but it serves as a first approximation for the case of free DNA in solution, or on an electron-microscope grid. When more data have been obtained about the precise shapes of DNA in solution or on a microscope grid, it should be possible to assign more accurate values of R and T to every kind of step. For example, experiments by many workers have shown

that the sequences GGC/GCC and AGC/GCT have much higher roll angles *R* than the average, perhaps +10° to +15° over the two steps. Those high roll angles for certain GC steps seem especially pronounced in solutions that contain magnesium ions, at a concentration similar to that found in living cells. Recall that the GC step was also implicated earlier, in Chapter 4, to assist in the curvature of DNA about various proteins.

So far, we have been investigating the geometry of supercoiled DNA as if the molecule were a long, rigid object. That is satisfactory for short pieces of DNA, but not if the DNA is long. In general, a short piece of DNA will behave like a rigid body, whereas a long piece of DNA will behave rather flexibly, as it is buffeted in 'Brownian movement' by fast-moving water molecules in solution. Now the terms 'short' and 'long' are meaningless unless we have a standard length for purposes of comparison. Scientists have supplied exactly such a comparison length, and they call it the 'persistence length' of DNA. They define it as the length at which the time-averaged angle made between the two ends of an intrinsically straight DNA molecule is equal to one radian, or 57°. The persistence length, as so defined, may then be deduced by several different experimental protocols such as electron microscopy, viscosity or light-scattering.

For many years, the accepted value for the persistence length of mixed-sequence DNA was near 140 base-pairs, or 450 Å. Thus, a piece of DNA 100 base-pairs long was expected to behave more-or-less as if it were rigid, whereas a piece 200 base-pairs long was expected to behave more-or-less as if it were a flexible string. Recent studies, however, have shown that previous estimates for the persistence length of DNA reach only to about one half of its true value, which is near 240 base-pairs or 800 Å. The mistake in previous assays was to assume that mixed-sequence DNA would be perfectly straight, whereas in fact it has a small but significant local curvature on account of its base sequence, near 1° to 3° per helical turn. So now the persistence length of DNA is taken as about 240 base-pairs or 800 Å, for a thermally-induced deflection of 57°. Thus, the contour length of a single turn of DNA supercoil, of size near 200 base-pairs, may be regarded as rigid to a first approximation – as seems plain, in fact, from electron micrographs such as those shown in Fig. 5.5.

It should be emphasized that free DNA in solution has, in general, little intrinsic curvature. Such curvature probably amounts at most to 15° or 20° per helix turn; and this level of curvature seems to be reached only for certain base sequences such as those listed in Table 5.1. In close association with positively charged proteins, however, DNA normally curves much more strongly than this, often by as much as 40 to 50° per helix turn, as described in Chapter 4. The

DNA has to deform to fit into the shape required by the protein, since this high level of curvature requires a systematic variation in roll angles of about plus or minus 9°, as we found earlier. In such cases of enforced curvature, a typical dinucleotide step which adopts $R = +3°$ under stress-free conditions in solution might be required to change to $R = +9°$ or $-9°$, when the DNA wraps tightly around a protein.

Yet in principle, if we are given a DNA sequence, and we have a table of allowable ranges of roll R for each type of dinucleotide step, we can still test for the ability of DNA to wrap tightly around a protein, by trying to establish a best-fit cosine wave of amplitude 9° through the allowable range. Figure 5.7(a) shows a hypothetical

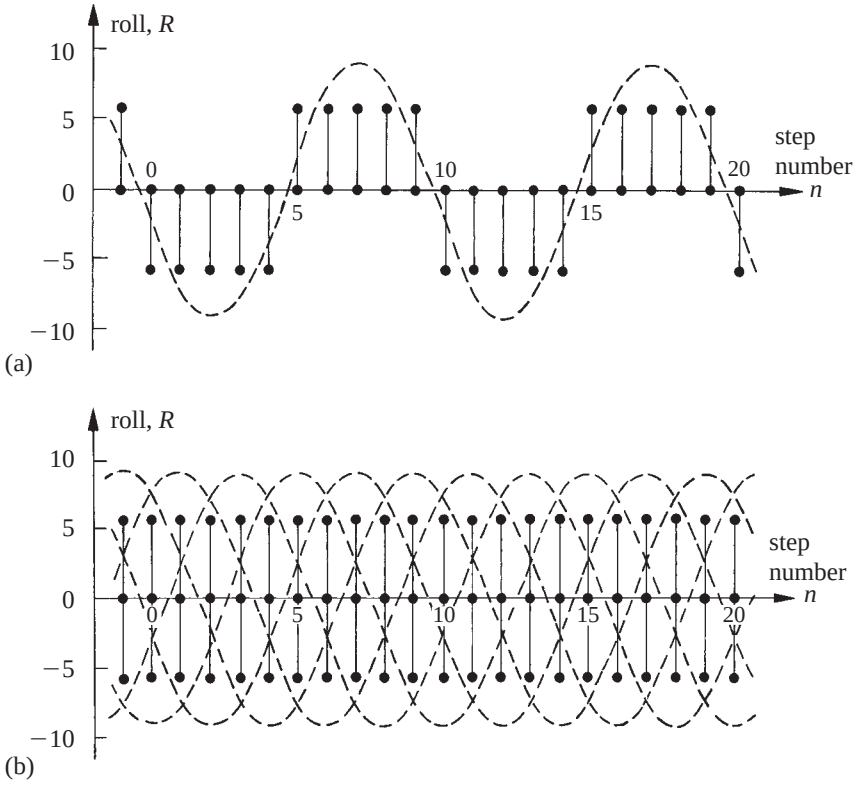

Figure 5.7 The ability of DNA to curve tightly about a protein depends on the allowable ranges of roll R at its base-pair steps. In (a), allowable ranges are $R = 0°$ to $-6°$ for steps 0, 1, 2, 3, 4, but $R = 0°$ to $+6°$ for steps 5, 6, 7, 8, 9. These roll angles may all remain equal to 0° for DNA free in solution; but when the DNA curves about a protein, many of them must switch to $-6°$ or $+6°$, in order to match the cosine wave of amplitude 9°, shown as a broken line, which coincides with the Fourier term of period 10. This DNA can bend only in one direction, since the best-fit cosine wave has a unique phase. In (b), allowable ranges of R are $-6°$ to $+6°$ for all steps. Now the DNA can bend in any direction, since the same cosine wave can fit the new roll angles with any left-to-right phase.

example of this sort, where a DNA molecule is perfectly straight in solution, but can curve around a protein when it is required to do so. The DNA in this case contains two kinds of dinucleotide step; one with an allowed range of roll $R = 0°$ to $-6°$, and the other with an allowed range of $R = 0°$ to $+6°$. In free DNA, the roll angles are all equal to zero for both kinds of step; and so a horizontal line can be drawn through these points, which indicates that the curvature k must be zero. But when such DNA is forced to bend around a protein, its sequence of roll values will approximate a cosine wave of amplitude 9°. In Fig. 5.7(a), a best-fit cosine wave of amplitude 9° has been drawn as a broken curve; it is clear that most of the steps must adopt their extreme roll values of $\pm 6°$ in order to fit the curve. The DNA steps whose minor-groove edges lie along the inside of the curve are those in the middle of each batch (at steps 2 and 12) where $R = -6°$. Any other phasing of this cosine wave would not fit the allowable roll values so well.

In summary, this particular kind of DNA, which is typical of many DNA molecules found in biology, can either be straight, or else it can curve in one specific direction when wrapping around a protein. In this respect, a typical DNA molecule is just like your finger or your knee: it can only bend in one direction and in one plane. On the other hand, if we want to construct a piece of DNA that can bend with equal ease in all directions, we need to include an allowable range of R values from $-6°$ to $+6°$ at every step (or, for example, from $-2°$ to $+10°$). Then the required cosine wave can be fitted at any phase, as shown in Fig. 5.7(b), and the DNA can bend in any direction quite easily.

How can we find out which kinds of dinucleotide step have high or low allowable ranges of roll R, under the bending stress applied by a protein? Some experimental data on this subject have already been presented in Fig. 4.10. Steps such as AA or TT, which prefer to occupy the low-roll locations of DNA when it is wrapped around a histone spool, presumably find it difficult to adopt a high-roll configuration. The preference for low roll is seen to be even stronger for sequences such as AAA; and this fits in with the idea that such sequences are relatively rigid, with R near 0°. Indeed, it is well-known that DNA molecules containing more than about 40 AA or TT steps in a row cannot easily wrap around a histone spool, because they cannot curve to the degree required. Similar considerations apply also to the step GC, which prefers to occupy the high-roll locations of DNA wrapped around a histone spool; it might be assigned an allowable range of R from $+5°$ to $+10°$. Finally, it is thought that steps TA and CA, among others, are relatively flexible to changes in roll, and that they can adopt the full range of R from $-9°$ to $+9°$.

Thus, sequences such as TATATATA can wrap with any phase (or angular setting) around a histone spool.

Although we have used the same geometrical ideas in relation both to the intrinsic curvature of repeating-sequence DNA in solution, and also to the enforced curvature of DNA around a protein spool, it cannot be stated too strongly that the curvature of DNA about a protein is very different in magnitude from its curvature when it is free in solution. Most DNA sequences have only small intrinsic curvature, when they are free in solution; but they often adopt preferred, highly curved shapes when they bend around a protein, owing to the different kinds of roll-angle flexibility at different sequences.

In Fig. 5.7 we were discussing the simple two-dimensional curvature of DNA in a plane, which is well-understood. But how does the three-dimensional curvature of DNA into superhelical shapes relate to its role in biology? In principle, there are two ways by which this could happen. First, the three-dimensional shape of the DNA might affect how well it binds to different proteins in the cell. For example, one protein may prefer that the DNA should wrap around it as a left-handed superhelix, while another might prefer a right-handed superhelix. Again, one protein may prefer to bind the DNA in the form of an extended superhelix, while another might prefer a flat superhelix. Furthermore, the three-dimensional shape of DNA might affect how it vibrates in solution, in response to the thermal motion of water molecules. It is well-known that things vibrate differently according to their shapes. For example, small molecules such as carbon dioxide (CO_2) and nitrogen dioxide (NO_2) vibrate differently because one is linear whereas the other is bent; and so they show different infra-red spectra. Also, the nature of these vibrations or fluctuations may be influenced by any local patches of high flexibility within the DNA, which are present on account of particular base sequences, as shown in Fig. 5.7(b).

Unfortunately, although the theoretical principles are straightforward, chemists and biologists have not yet collected many clear experimental data on how the three-dimensional shape of DNA, over a large scale, affects its binding to proteins or its thermal fluctuations. The two-dimensional curvature of DNA into a plane has been implicated in many processes, but not yet the three-dimensional curvature into superhelices. All that we know today about the three-dimensional curvature of DNA is what we have learned from physical studies such as electron microscopy and gel electrophoresis. Here is a field where there is plenty of room for improved understanding of biological function.

Further Reading

Bednar, J., Furrer, P., Katritch, V., Stasiak, A.Z., Dubochet, J., and Stasiak, A. (1995) Determination of DNA persistence length by cryo-electron microscopy: separation of the static and dynamic contributions to the apparent persistence length of DNA. *Journal of Molecular Biology* **254**, 579–94. Half of the bending of DNA as seen by electron microscopy is due to its base sequence and half is due to its thermal motion.

Brukner, I., Susic, S., Dlakic, M., Savic, A., and Pongor, S. (1994) Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *Journal of Molecular Biology* **236**, 26–32. Metal ions strongly influence DNA curvature in solution at GGGCCC-type sequences.

Brukner, I., Balmaaza, A., and Chartrand, P. (1997) Differential behaviour of curved DNA upon untwisting. *Proceedings of the National Academy of Sciences, USA* **94**, 403–6. Influence of the intercalator ethidium bromide on DNA supercoiled structures in solution.

Calladine, C.R., Drew, H.R., and McCall, M.J. (1988) The intrinsic curvature of DNA in solution. *Journal of Molecular Biology* **201**, 127–37. All relevant equations for the shape of curved DNA in three dimensions, in terms of its curvature and twist.

Calladine, C.R., Collis, C.M., Drew, H.R., and Mott, M.R. (1991) A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *Journal of Molecular Biology* **221**, 981–1005. Pictures of long, superhelically curved DNA such as those shown in Fig. 5.5, by electron microscopy.

Dlakic, M. and Harrington, R.E. (1995) Bending and torsional flexibility of GC-rich sequences as determined by cyclization assays. *Journal of Biological Chemistry* **270**, 29945–52. The sequence GGGCCC helps DNA to curve into small circles, as much as does AAAAA but in an opposite helical phase, with high roll rather than low or zero roll.

Goodsell, D.S., Kopka, M.L., Cascio, D., and Dickerson, R.E. (1993) Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proceedings of the National Academy of Sciences, USA* **90**, 2930–4. The curved structure of a GGCC sequence as seen in a crystal with magnesium ions.

Lavigne, M., Kolb, A., Yeramian, E., and Buc, H. (1994) CRP fixes the rotational orientation of covalently closed DNA molecules. *EMBO Journal* **13**, 4983–90. Even when a DNA molecule is not much bent by its base sequence, the binding of a protein such as CRP that curves DNA locally, can extend the phase of curvature for hundreds of base-pairs in either direction.

Murphy, C.J. (2001) Photophysical probes of DNA sequence-directed structure and dynamics. *Advances in Photochemistry* **26**, 145–217. A good summary of DNA structure and its measurement by light-based molecular probes.

Nishikawa, J., Amano, M., Fukue, Y., Takana, S., Kishi, H. *et al*. (2003) Left-handedly curved DNA regulates accessibility to *cis*-DNA elements in chromatin. *Nucleic Acids Research* **31**, 6651–62. The artificial placement of left-handed supercoiled DNA, upstream from a gene in human cells, attracts a nucleosome to that curved location; and thereby places the downstream binding-site for TBP protein in an exposed location between nucleosomes, where it can act readily to initiate transcription.

Revet, B., Brahms, S., and Brahms, G. (1995) Binding of the transcription activator NRI to a supercoiled DNA segment imitates association with the natural enhancer: an electron microscopic investigation. *Proceedings of the National Academy of Sciences, USA* **92**, 7535–9. A left-handed supercoil with a sequence repeat of 10 base-pairs binds specifically to a protein that activates transcription, and may be seen clearly as a well-defined superhelix on the grid.

Rivetti, C., Walker, C., and Bustamente, C. (1998) Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *Journal of Molecular Biology* **280**, 41–59. The local bend-angle induced by d(AAAAAA) was estimated as 13.5° by atomic force microscopy (see Chapter 9).

Roychoudhury, M., Sitlani, A., Lapham, J., and Crothers, D.M. (2000) Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proceedings of the National Academy of Sciences, USA* **97**, 13608–13, The local curvature of d(TATAAACGCC) was estimated as 13° by rates of ligase-mediated cyclization.

Sampaolese, B., Bergia, A., Scipioni, A., Zuccheri, G., Savino, M., Samori, B., and De Santis, P. (2002) Recognition of the DNA sequence by an inorganic crystal surface. *Proceedings of the National Academy of Sciences, USA* **99**, 13566–70. C-shaped or S-shaped DNA molecules will adhere to a mica surface in precise ways.

Tchernaenko, V., Radlinska, M., Drabik, C., Bujnicki, J., Halvorson, H.R., and Lutter, L.C. (2003) Topological measurement of an A-tract bend angle: comparison of bent and straight states. *Journal of Molecular Biology* **326**, 737–60. The local bend-angle induced by d(AAAAAA) was estimated by topological methods to be 26° at 4°C, or 17° at 37°C, by comparison with a high-temperature straight state. (Differences between these values and those of Rivetti *et al*. are probably attributable to the different concentrations of magnesium ions.)

Exercises

5.1 Any ribbon with uniform curvature k and twist t will generate a spiral or helix. The characteristic shape of the helix may be calculated from k and t by use of the formulas given on p. 100 for pitch angle α, radius r, pitch p, and contour length N^* of one complete turn (see also Fig. 5.4). By convention, k, r and N^* are always positive; while t, p, and α are all positive for a right-handed helix, but negative for a left-handed helix.

helix	k (°/bp)	t (°/bp)
a	2	−2
b	2	2
c	2	0.2
d	0.2	2

Compute values of α (in degrees) and r, p, N^* (in base-pairs) for the four sets of k and t specified above. Which helix has the smallest diameter, and which has the largest? Which helices are left-handed, and which are right-handed?

5.2 The helical geometry of a uniformly curved and twisted ribbon (Fig. 5.4) may be used in a different context to provide an approximate calculation for the situation shown in Fig. 3.14(c), where the addition of uniform roll at every base-pair step tilts the base-pairs with respect to an overall helix axis, which is vertical in the picture.

The key to the calculation is to regard the central 'wire', which is shown connecting base-pairs in models (a) and (c) of Fig. 3.14, as a 'twisted ribbon'. Since the blocks are attached locally perpendicular to the wire, tilt = 90° − α.

In this application, t becomes the local twist T (°/bp) between successive base-pairs, while k becomes the roll angle R (°/bp). N^* is the number of steps per complete helical turn; and so $360°/N^*$ corresponds to the global twist T_G (°/bp) measured with respect to an overall helix axis, i.e. an imaginary vertical line in Fig. 3.14(c). Formulas from p. 100 may thus be re-written in the present context:

$$\tan\alpha = T/R, \text{ so } \tan(\text{tilt}) = R/T; \quad T_G^2 = T^2 + R^2.$$

a Given the T_G and tilt values shown below for idealized 'A,' 'B', and 'C' forms of DNA from fibers (see Chapter 9), use the formulas above to obtain the local twist T and the roll R in each case.

Fiber model	T_G (°/step)	Tilt (°)
'A'	32.7	+20
'B'	36.0	0
'C'	40.0	−10

b Inspection of Fig. 3.14(c) shows that it has global twist $T_G = 36°$. Use the formulas above to obtain the local twist T, for roll $R = 12°$. (Note that in Chapter 3 it was stated that $T \approx 36°$ for this model: thus the slight difference between values of global twist T_G and local twist T was overlooked there.)

5.3a In Exercise 4.6 you were asked to compute the quantity which is now described on p. 104 as the 'second sum', for five different repeating sequences R_n of roll angles. Confirm by direct calculation (suggestion: extend the table of Exercise 4.6 to five columns) that the 'first sum' is precisely zero in all of these five cases. For ease of calculation, we fixed the roll angles in Exercise 4.6 always to give a first sum of zero; but this will not generally be the case.

b Perform a calculation as set out on p. 104 for the case

$$R_n = (0°, 10°, 10°, 10°, 0°, 0°, 0°, 0°, 0°, 0°)$$

in order to find the curvature k and its phase. Is the first sum equal to zero in this example? Which step has the maximum roll R in the Fourier wave?

Note on the computation of $\arctan(f/s)$, where f = first sum and s = second sum. Suppose, for example, $f = 1.2$, $s = -1.6$. Then $f/s = -0.75$. If you ask your calculator for $\arctan(-0.75)$ it will probably give $-36.9°$, although another equally valid answer is $+143.1°$; for there are always two angles in the 360° circle, separated by 180°, whose tangents are identical. In the present case we must choose only one of these two angles. The rule which we need here is that if θ is the correct answer, $\sin\theta$ and $\cos\theta$ have the same signs as f and s, respectively. Here, $\sin -36.9° = -0.6$ and $\cos -36.9° = +0.8$, while $\sin 143.1° = +0.6$ and $\cos 143.1° = -0.8$; and so it is the second answer, 143.1°, which is correct.

5.4 Consider a long DNA molecule having a repeating sequence of the kind $A_4N_6A_4N_6A_4N_6$ …, or $(A_4N_6)_m$, i.e. a 10-base-pair repeat of the kind A_4N_6, where N stands for any base other than A.

a Taking $R = 0°$ for steps AA, and $R = 3.3°$ for all other steps, convert this sequence repeat into the corresponding roll angle repeat R_n for $n = 0$ to 9, where $n = 0$ corresponds to the first AA step in the repeat. Compute the first and second sums as on p. 104, and hence evaluate k; and finally express this in units of °/bp. Which step has the largest value of roll R in the Fourier wave?

b Taking twist values $T = 35°$ for step AA, and $T = 34°$ for all other steps, calculate the overall twist t for a 10-step repeat (i.e. sum the T values and then subtract 360°); and express this also in units of °/bp.

c Using the values of k and t as calculated above, find the radius r of the superhelical curve which is made by this repeated-sequence DNA, and the number N^* of base-pair steps in a

complete superhelical turn. Use the formulas on p. 100; and note that r is given here in base-pair units: 1 bp $\approx$ 3.3 Å.

5.5 Repeat Exercise 5.4, but using instead sequence $(A_6N_4)_m$.

5.6 Repeat Exercise 5.4, but using the sequence $(A_6N_5)_m$. Here, in working out k it will be necessary to use multiples of $360°/11 = 32.7°$ for $n = 0$ to $n = 10$;

$$\text{i.e. first sum} = \sum_{n=0,10} R_n \sin(32.7°n), \text{ etc.}$$

The value of k will be in degrees per 11-step repeat; but this should be expressed in °/bp for use in the formulas on p. 100.

Which of the sequences $(A_4N_6)_m$, $(A_6N_4)_m$, and $(A_6N_5)_m$ form left-handed superhelices, and which form right-handed superhelices?