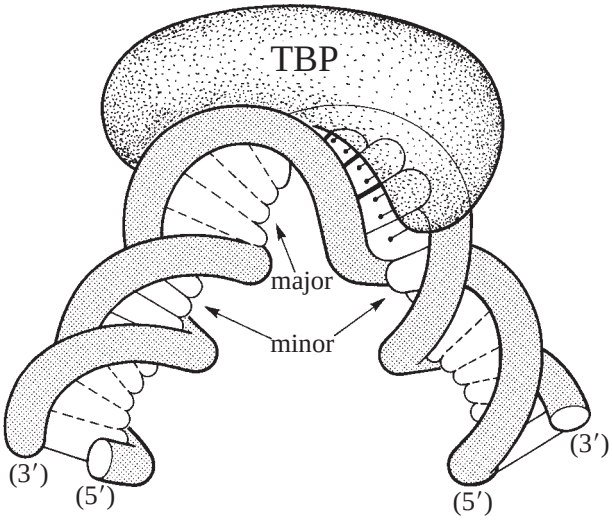


Figure 4.14 Schematic drawing of the TBP protein bound to a TATA-containing sequence in DNA, which it unwinds and bends sharply.

Studies of these zinc-finger proteins with DNA have provided some of our clearest insights into protein-DNA recognition. For example, N. Pavletich and C. Pabo have analysed by X-ray crystallography not only the three-finger complex of Zif268 with DNA as described above, but also a complex with five zinc-fingers from a human cancer-causing gene GLI, and two full turns of DNA to which it binds specifically. They find substantial variations in the slide and twist of this DNA which is recognized by the zinc-finger protein, that are also present in crystals of the same DNA without the protein. Thus, the GLI protein seems to recognize an inherent structural variation in slide and twist, as induced by the particular base sequence to which it binds.

In other studies, many workers have found systematic relations between the DNA bases recognized by any zinc-finger module, and the identities of amino acids in that finger module. These relations can be understood in terms of local patterns of hydrogen bonding, such as that shown in Fig. 4.12 and others of the same kind.

Much progress has also been made recently at determining the precise geometry by which the TATA sequence and other related sequences unwind. Three different crystal structures of the TBP protein (see above) with TATA-containing DNA show a highly untwisted double helix, as drawn schematically in Fig. 4.14. The base-step twist T decreases from about 34° outside the TATA region, to 20° within it; while roll R increases from about 0° outside the TATA region, to +20° within it. These changes untwist and open the

minor groove by a large amount, without much change in slide *S* from near 0 Å.

Such DNA when bound to TBP is thus untwisted and bent somewhat, in the fashion of an accordion that has been bent into a right-angle shape. The TBP protein induces such a highly distorted DNA structure, by placing a large hydrophobic surface in the minor groove, to which the minor groove edges of A–T base-pairs of the sequence TATAAA can bind. There is little doubt that this untwisted structure facilitates transcription, probably by helping to orient the DNA into the active site of the polymerase, where the strands will be separated; and to serve as a 'scaffold' for the recruitment of accessory proteins (see Chapter 8). The specific untwisting of a TATA sequence by TBP has been measured in solution as one-half turn of DNA, by S. Hirose and colleagues or as one-third turn by J. Kahn, in good accord with the X-ray structure. The TATA sequence also unwinds in its complexes with various enzymes that carry out DNA strand-switching or recombination, often known as 'resolvase' proteins; but no one is yet certain how that slight untwisting leads to full separation of the two strands, in preparation for switching between different DNA molecules.

A similar untwisted helix has been suggested for the complex of a human sex-determining protein SRY with DNA, where the DNA has a related sequence CACAAA in contact with the protein (instead of TATAAA). The SRY protein also untwists the double helix, by protruding its surfaces into the minor groove; and in addition, it untwists by pushing or intercalating a hydrophobic isoleucine amino acid between two adenine bases of the sequence CAA, somewhat as for the intercalation of ethidium bromide described in Chapter 2. Therefore, such untwisted structural forms of DNA may be very common in living cells.

In summary, this chapter has been about the twisting and curving of DNA, and how DNA is recognized by proteins according to its twisting and curving and other features. In order for the bases of DNA to be exposed for copying into more DNA or into RNA, the double helix has to untwist. The places where DNA unwinds in Nature, to start the copying process, are often the weakest parts of the double helix. In order for DNA to wrap itself tightly into chromosomes, the double helix has to curve. The curvature spreads itself out over as many base-pairs as possible, in order to keep the roll angle at individual base-pair steps small, so that the bases will stack well onto each other. The sequence of the DNA influences how easily DNA can adopt any given curved shape, by the preferences of different sequences for different roll angles. These preferences apply not only to protein–DNA complexes as part of a chromosome, but also to protein–DNA complexes generally, such as those formed between the

DNA and 'repressor' proteins. In general, a protein recognizes certain sequences in DNA by first recognising some large-scale feature of the molecule such as the sugar–phosphate chains (in other words, by 'docking'), and then by probing the details of the bases.

In these first four chapters we have talked about curving DNA in a plane, but not in three-dimensional space; we have treated twisting and curving as separate subjects, whereas often they are related by the shape and thermal vibration of the DNA. We have not yet discussed how a naked DNA molecule is assembled into a mixture of protein and DNA to make a chromosome, nor how genes are activated within chromosomes by removal of some of the proteins. Nor have we shown how proteins recognize specific sequences in DNA. Finally, we have not yet explained any of the experimental techniques which scientists use to probe DNA structure and function, such as X-ray diffraction, nuclear magnetic resonance, and gel electrophoresis. The following chapters will deal with these and other topics.

Note

1. See Appendix 1.

Further Reading

Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987) Structure of the repressor–operator complex of bacteriophage 434. *Nature* **326**, 846–52. X-ray crystal structure of a complex between the 434 protein and DNA, showing both 'docking' and 'probing' modes of binding.

Azizov, M., Ulyanov, A., Kuprash, D., Shakov, A., Gavin, I., and Nedospasov, S. (1992) Production and characterization of a library of mononucleosomal DNA from the chromatin of human cells. *Doklady Akademii Nauk* **322**, 415–20. Independent confirmation of DNA bending periodicities in the chromatin of human blood cells.

Bailey, K.A., Pereira, S.L., Widom, J., and Reeve, J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the prokaryotic origin of histone-dependent genome evolution. *Journal of Molecular Biology* **303**, 25–34. The AA/TT or GC sequence preferences seen for DNA curvature about eukaryotic histone proteins are seen also for DNA curvature about histone proteins from archaea.

Calladine, C.R. and Drew, H.R. (1986) Principles of sequence-dependent flexure of DNA. *Journal of Molecular Biology* **192**, 907–18. How roll angles in the DNA can change, to let DNA bend tightly about a protein.

Champoux, J.J. (2001) DNA topoisomerases: structure, function and mechanism. *Annual Review of Biochemistry* **70**, 369–413. A good review on various topoisomerase enzymes and their biological activities.

Chen, D., Bowater, R., Dorman, C., and Lilley, D.M.J. (1992) Activity of a plasmid borne *leu500* promoter depends on the transcription and translation of an adjacent gene. *Proceedings of the National Academy of Sciences, USA* **89**, 8784–8. Evidence for the local supercoiling of DNA in bacteria as induced by its motion through RNA polymerase.

Choo, Y., Sánchez-García, I., and Klug, A. (1994) *In vivo* repression by a site-specific DNA-binding protein designed against an oncongenic sequence. *Nature* **372**, 642–5. Specific contacts between single DNA bases and protein amino acids from zinc fingers are used to design new proteins, which bind to cancer-causing DNA sequences.

Drew, H.R., Weeks J.R., and Travers, A.A. (1985) Negative supercoiling induces spontaneous unwinding of a bacterial promoter. *EMBO Journal* **4**, 1025–32. The −10 region of a bacterial promoter, having a sequence of the kind TATA, unwinds more easily than other regions of the DNA at room temperature under torsional stress.

Fox, K.R. and Brown, P.M. (1996) Minor-groove binding ligands alter the rotational positioning of DNA fragments on nucleosome core particles. *Journal of Molecular Biology* **262**, 671–85. Small molecules which bind in the minor groove of AT-rich regions, can change the rotational setting with which DNA wraps around a histone octamer.

Kahn, J.D. (2000) Topological effects of the TATA-box binding-protein on minicircle DNA, and a possible thermodynamic linkage to chromatin remodeling. *Biochemistry* **39**, 3520–4. Unwinding of the TATA sequence by one-third of a double-helical turn upon addition of TBP protein, similar to that seen in the crystal.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993) Co-crystal structure of TBP recognizing the minor groove at a TATA element. *Nature* **365**, 520–7. Unwound DNA at TATA sequence as induced by binding of the TBP protein in the minor groove.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–20. More unwound DNA as induced by the binding of the TBP protein.

King, C.-Y. and Weiss, M.A. (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proceedings of the National Academy of Sciences, USA* **90**, 11990–4. Description of how the sex-determining SRY protein inserts a hydrophobic isoleucine side-chain between bases in the minor groove.

Koudelka, G.B., Harrison, S.C., and Ptashne, M. (1987) Effect of noncontacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature* **326**, 886–8. Quantitative measurements on how tightly the 434 and Cro repressors dock to different sequences in DNA.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505. The first crystal structure analysis of a hormone-receptor protein bound to DNA.

Malhotra, A., Severinova, E., and Darst, S.A. (1996) Crystal structure of a sigma-70 submit fragment from *E. coli* RNA polymerase. *Cell* **87**,

127–36. RNA polymerase from bacteria contains a protein structure that helps unwind DNA at TATAAT, which is different from that used by TBP.

Miller, J., McLachlan, A.D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO Journal* **4**, 1609–14. The earliest correct proposal of a repeating structure for the zinc-finger proteins.

Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA Polymerases: the wholo story. *Current Opinion in Structural Biology* **13**, 31–9. Summarizes current understanding of initiation of transcription by bacterial RNA polymerase.

Murakami, K.S., Matsuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285–90. How the sigma subunit of bacterial polymerase contacts the promoter region, and the transcription bubble is stabilized.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995) Crystal structure of a TFIIB-TBP-TATA element ternary complex. *Nature* **377**, 119–28. Both TBP and another protein, TFIIB, visualized as they bind to the DNA in order to start transcription.

Ohndorf, U.-M., Rould, M.A., He, Q., Pabo, C., and Lippard, S.J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708–12. The specific binding of cis-dichloro-diamino-platinum to one of its biological targets.

Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a *Zif268*–DNA complex at 2.1 Å. *Science* **252**, 809–17. Three-dimensional structure of a transcription factor somewhat like TFIIIA bound to DNA: the structure shows the probing of guanine bases by arginine amino acids.

Pruss, D., Reeves, R., Bushman, F.D., and Wolffe, A.P. (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *Journal of Biological Chemistry* **269**, 25031–41. The insertion enzyme for the HIV virus recognizes the rotational setting of the DNA on histone spools, as it inserts into chromosomes.

Pryciak, P.M., Sil, A., and Varmus, H.E. (1992) Retroviral integration into mini-chromosomes in vitro. *EMBO Journal* **11**, 291–303. A recombination enzyme also recognizes GC *versus* AT bending preferences of DNA in chromatin.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *Journal of Molecular Biology* **191**, 659–75. Different base sequences occupy different locations within DNA as it curves, owing to their preferences for different kinds of roll angle (see Fig. 4.10).

Werner, M.M., Gronenborn, A.M., and Clore, G.M. (1996) Intercalation, DNA kinking, and the control of transcription. *Science* **271**, 778–84. An excellent review of how proteins such as TBP and SRY bind to DNA.

Yang, W. and Steitz, T.A. (1995) Crystal structure of a site-specific recombinase or resolvase complexed with a 34 bp cleavage site. *Cell* **82**, 193–207. First vizualisation of a specific recombinase enzyme as bound to DNA.

Bibliography

Liu, L.F. and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences, USA* **84**, 7024–7. A classic paper describing what might happen if RNA polymerase could not rotate freely about the DNA during transcription.

McClellan, J.A., Palacek, E., and Lilley, D.M.J. (1986) $(A–T)_n$ tracts embedded in random-sequence DNA: formation of a structure which is chemically reactive and torsionally deformable. *Nucleic Acids Research* **14**, 9291–309. The unwinding of TATA-like sequences to a twist of 12 base-pairs per turn before a cruciform appears.

Pavletich, N.P. and Pabo, C.O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701–7. Analysis of five zinc-fingers bound to two turns of DNA, with many structural variations in the DNA being recognized by the protein.

Rosenberg, B., VanCamp, L., Trosko, J.E., and Mansour, V.H. (1969) Platinum compounds: a new class of potent antitumour agents. *Nature* **222**, 385–6. Testing of platinum compounds against cancer in mice, with positive results.

Tabuchi, H., Handa, H., and Hirose, S. (1993) Underwinding of DNA on binding of yeast TFIID to the TATA element. *Biochemical and Biophysical Research Communications* **192**, 1432–7. First detection of DNA unwinding by the TBP (or TFIID) protein in solution.

Exercises

4.1 Make a physical model to show some of the problems that occur when a polymerase molecule runs along double-helical DNA, as in Fig. 4.3.

To do this, take two equal lengths of rubber tubing, and tie them together firmly at one end. Then, working from that end, arrange the two tubes in the form of a fairly loose right-handed double helix; and when you get to the other end, tie the tubes together there also. The two tubes represent the two sugar–phosphate chains of a piece of DNA.

Now insert a stick between the two tubes, near the mid-point of the double helix. Holding an end of the double helix in one hand and the stick in the other, push the stick along the double helix. Observe how the helices tighten up ahead of the stick, and relax behind it – unless, that is, the stick is allowed to rotate as it moves forward along the model.

(This demonstration was suggested by Maxim Frank-Kamenetskii.)

4.2 The base sequence of a single strand of double-helical DNA is given below:

(5′) ACTTAAGGCCCTATATACCTAGACTCGGCGGTAAATTT (3′)

- **a** Underneath it, write out the base sequence of the complementary strand.
- **b** Identify AT-rich and GC-rich regions of the molecule.
- **c** Identify pyrimidine–purine steps.
- **d** Hence identify strong, medium, weak, and very weak regions of cross-chain bonding, in the manner of Fig. 4.4.

4.3 The base sequence of a single strand of double-helical DNA is given below:

(5′) GCGCCTAGAAATAATACTAGTATTATTTCTAGCCGG (3′)

- **a** Underneath it, write out the base sequence of the complementary strand.
- **b** Find a region which can make a 'cruciform' formation, as in Fig. 4.5; and draw a picture to show the new pairing.

4.4 A semicircle of garden path around a tree (cf. Fig. 4.7) is made from 10 rectangular slabs, with an individual angle of 20° at each of the 9 'steps' between consecutive slabs.

- **a** How many slabs, arranged in the same pattern, are needed to make a complete circle? Given that the center-to-center spacing of the slabs is 0.6 m, what is the circumference of the circle? And what is the radius of the circle?
- **b** Now do a different calculation on the same curved path, in order to find its radius. First, convert the curvature of 20° per slab into units of radian per meter by the use of the conversions: 1 radian = $180°/\pi = 57.3°$; 1 slab = 0.6 m. Then calculate the radius of the circle from the formula (which we have not mentioned before, but which is easy to prove): radius of curvature = 1/(curvature).

4.5 Imagine that the four DNA molecules shown in Fig. 4.8 have been made with a curvature of 30° per helical turn, instead of 45° per turn.

Sketch out the corresponding version of each of the four plots (a) to (d) in Fig. 4.9, and list the maximum and minimum values of roll angle R in each case.

4.6 A particular DNA molecule has $T = 36°$ at every step.

- **a** The pattern of roll angles along the length of the molecule is somewhat similar to that shown in Fig. 4.8(c) and Fig.

4.9(c), except that the batches of non-zero roll are only three steps long, and the individual roll angles are smaller. Specifically

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

where $n = 0$ to 9, and the pattern repeats many times.
Compute the curvature of the molecule (in units of degrees per helical turn) by evaluating the sum of 10 consecutive terms $R_n \cos(36°n)$, $n = 0$ to 9.
To do this, make a table of 10 rows and 4 columns, with the following entries:
Column 1, values of n;
Column 2, values of $\cos(36°n)$;
Column 3, values of R_n;
Column 4, values of $R_n \cos(36°n)$.
Obtain the required answer by taking the sum of all entries in Column 4.

b Repeat the calculation for

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

c Repeat for

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

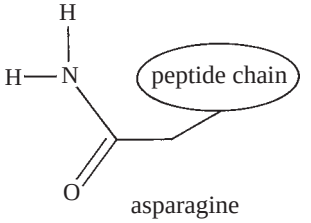
d Repeat for the case (which is somewhat similar to that of Figs 4.8(d) and 4.9(d), and to Exercise 4.5):

$$R_n = 6° \cos(36°n),\ n = 0 \text{ to } 9$$

e Repeat for

$$R_n = 6° \cos(36°n) + 4°$$

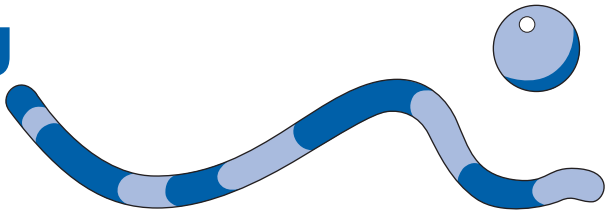
4.7 Fig. 4.12 shows the hydrogen bonding between a guanine base of DNA and an amino acid arginine on the peptide chain of a nearby protein. The diagram below shows an amino acid asparagine in a similar way. Would asparagine form hydrogen bonds better to guanine or to adenine, in the major groove of DNA, as in the style of Fig. 4.12? See Fig. 2.11 for chemical formulas.



asparagine

CHAPTER 5

Curving in Three Dimensions



So far in this book we have said a lot about the structure of DNA as a double helix, and the way in which different kinds of double helix can result from the different geometries with which base-pairs stack on one another. But in most chapters our DNA, seen as a long rod or thread, has been straight. Only in Chapter 4 did we consider a molecule of DNA which follows a curved path; and the curve there was in a plane, like a garden path. We have now reached the stage where we must explain the curvature of DNA in three dimensions.

A DNA helix often proceeds through space as a three-dimensional spiral, rather than as a straight line or as a plane curve. The spiral path of DNA is usually described as a 'superhelix' or 'supercoil'. This is because we know that the DNA thread itself has a local helical, twisted structure even when it is straight. The qualitative aspects of DNA curvature in three-dimensional space are not difficult to understand, although we shall need to be careful to distinguish between the twisting of the path of the DNA as a whole, and the twisting of sugar–phosphate chains on a local scale. But you may find that some of the quantitative and mathematical aspects of the subject are hard to grasp. Do not worry if you find the mathematics rather heavy going: be content to appreciate the qualitative aspects of three-dimensional curvature.

In Chapter 4, we explained how DNA can curve in a plane to make a circle. There, curvature k was defined as the angle turned for a given length of DNA. For example, k might be 1° per base-pair or, equivalently, 10° per double-helical turn of 10 base-pairs. When DNA curves in three dimensions, our simple definition of curvature k still holds for any small segment of its path, such as one helical turn or 10 base-pairs; but in practice the DNA often departs from the local plane

of curvature after two or more turns. In such cases, the plane of curvature twists by some angle t as the DNA advances through space, and so the DNA coils into three dimensions.

There is an easy way to understand this point. Consider a little man in an airplane (Fig. 5.1). In Fig. 5.1(a) he flies loop-the-loops, which are circular paths in a vertical plane, like the path of a car in a fairground or carnival ferris wheel; and he does so by pointing the nose of his 'plane either up or down, using his joystick to tilt the elevators on the tail of the 'plane. This is pure curvature k, since his path through the air curves by some angle k per unit length of loop. In Fig. 5.1(b), on the other hand, he flies straight ahead but in a narrow spiral, by pushing the joystick either to the left or to the right, and so activating the ailerons on the wings. The path of the airplane is now one of pure twist t, since the 'plane rotates about its long axis by some angle t per unit length of advance through the air. The twist t can be either plus or minus, depending on whether he turns the wheel to the right or to the left, respectively. After a while, the pilot gets dizzy from flying in a straight, twisted path, and he decides to combine the two motions of k and t. So he turns the nose of his 'plane both up and to the left at the same time; and he now flies in a broad spiral through the air, like the path of the wire in a coiled bed-spring. It is the *combination* of k and t that provides for a broad spiral path;

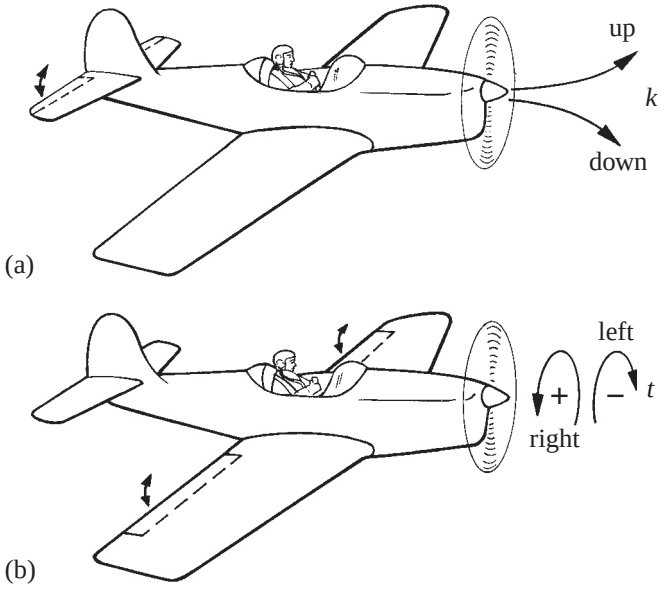


Figure 5.1 Our friend in the airplane explains to us about curvature k and twist t. Curvature k alone as in (a) makes the 'plane fly in vertical loop-the-loops. Twist t alone as in (b) makes the 'plane fly in a narrow spiral. Only the combination of curvature k and twist t produces a broad spiral.