DNA bending: the prevalence of kinkiness and the virtues of normality

Richard E. Dickerson*

Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

Received January 16, 1998; Revised and Accepted February 26, 1998

ABSTRACT

DNA bending in 86 complexes with sequence-specific proteins has been examined using normal vector plots, matrices of normal vector angles between all base pairs in the helix, and one-digit roll/slide/twist tables. FREEHELIX, a new program especially designed to analyze severely bent and kinked duplexes, generates the foregoing quantities plus local roll, tilt, twist, slide, shift and rise parameters that are completely free of any assumptions about an overall helix axis. In nearly every case, bending results from positive roll at pyrimidine-purine base pair steps: C–A (= T–G), T–A, or less frequently C–G, in a direction that compresses the major groove. Normal vector plots reveal three well-defined types of bending among the 86 examples: (i) localized kinks produced by positive roll at one or two discrete base pairs steps, (ii) three-dimensional writhe resulting from positive roll at a series of adjacent base pairs steps, or (iii) continuous curvature produced by alternations of positive and negative roll every 5 bp, with side-to-side zig-zag roll at intermediate position. In no case is tilt a significant component of the bending process. In sequences with two localized kinks, such as CAP and IHF, the dihedral angle formed by the three helix segments is a linear function of the number of base pair steps between kinks: dihedral angle = 36° × kink separation. Twenty-eight of the 86 examples can be described as major bends, and significant elements in the recognition of a given base sequence by protein. But even the minor bends play a role in fine-tuning protein/DNA interactions. Sequence-dependent helix deformability is an important component of protein/DNA recognition, alongside the more generally recognized patterns of hydrogen bonding. The combination of FREEHELIX, normal vector plots, full vector angle matrices, and one-digit roll/slide/twist tables affords a rapid and convenient method for assessing bending in DNA.

INTRODUCTION

Bending of the DNA duplex has proven to be a significant aspect of its interactions with many proteins. The first dramatic example of a 90° bend in the DNA helix was that produced by the catabolite activator protein, CAP (1,2), but comparable bends have since been seen with Lac operator (3), PurR (4), γδ-resolvase (5), integration host factor (IHF) (6), TATA-binding protein (TBP) (7–10) and others. Smaller DNA bends have been observed with prokaryotic helix–turn–helix (HTH) proteins such as the lambda and 434 repressors (11–15), and with the EcoRV repressor when bound specifically to its cognate sequence (16,17). In contrast, zinc-binding proteins, leucine zippers (bZIP) and basic helix–loop–helix (bHLH) proteins typically induce little or no bending in their bound DNA duplexes (summarized in table 2 of ref. 18 or table 4 of ref. 19).

In most of the original reports of protein–DNA complexes, primary attention has been given to the conformation of the protein. Lesser attention has been paid to the DNA, other than to note that it is indeed bent and to report the overall extent of bending. The author has recently concluded an analysis of bending in 86 DNA duplexes bound to sequence-specific proteins as deposited in the Nucleic Acid DataBase, employing the familiar but relatively under-used concept of normal vectors to base pairs. A new program has been developed, FREEHELIX, that facilitates the study of drastically bent or kinked helices. Comprehensive results will be reported in detail in a later review, but this paper focuses on the problem, the mode of attack, and the principal conclusions.

Four questions about bending in DNA are addressed in this paper.

(i) Through what local structural variations can a DNA duplex be induced to bend?
(ii) Are different types of bend induced by different classes of local structural variations?
(iii) To what extent are these structural variations dependent upon the base sequence of the DNA?
(iv) Are sequence-dependent local structure variations involved in the recognition of DNA by proteins?

Two complementary but quite different points of view are possible when examining and analyzing DNA structure: local and global. The former looks only at stacking contacts from one base pair to the next; the latter establishes an overall helical pathway and then analyzes stepwise behavior relative to this superimposed pathway. Both approaches have merits, and in fact the widely used program CURVES by Lavery and coworkers offers both options (20). But the present FREEHELIX analysis focuses on the local approach for phenomenological reasons. To anthropomorphise the
issue, one base pair ‘sees’ only its neighbors to either side, plus a secondary restraint imposed by a finite backbone length. One base pair says to its neighbors, ‘Let us stack.’ It does not say, ‘Let us build a helix.’ The helix is a secondary phenomenon that results from the cumulative stacking of many individual base pairs. If this stacking is reasonably uniform from one base pair to the next, then the result can be described by an outside observer as a helix, and this is a useful and aesthetically pleasing construct. But to understand the phenomenological basis of DNA bending, one must focus on the stacking and not on the helix. This is what FREEHELIX is specifically designed to do.

MATERIALS AND METHODS

The essence of the FREEHELIX program is that it calculates the familiar base step parameters; roll, tilt, twist, slide, shift and rise, relative to local axes defined between adjacent base pairs, without recourse to assumed helix axes, and then presents the results in a tabular form that facilitates a study of DNA bending. The name of the program was chosen to reflect the fact that the analysis is now free of all assumptions about a helix axis. Local helix parameters, of course, have been calculated by other helix parameter programs (e.g., 20,21). But a special feature of FREEHELIX is its utilization of normal vectors to individual base pairs as a device for following bending.

Establishing the viewing axis

The FREEHELIX program can handle up to 50 bp and 2000 non-hydrogen atoms in one continuous duplex. Missing phosphate groups at nicks in the backbone chain can be filled in with dummy coordinates, but all bases along each chain must be present. (Unpaired bases at the ends of a duplex are simply deleted.) The first act of FREEHELIX is to define an overall viewing axis, which in a straight helix would be the helix axis, but which in a badly bent or deformed helix is only the optimal direction from which to examine the deformation. A set of orthogonal working coordinates for the structure is then generated, with the viewing axis along z and the first strand of the helix rising to increasing z values.

Establishment of the viewing axis is schematized in Figure 1. The user of the program selects a set of interatomic vectors within the helix with which to define the viewing axis. The choice of these defining vectors is entirely optional, although the easiest and most customary set are all vectors from a C1′ atom of one base to that of the following base along the same helix strand, and from the N1/N9 atom of a pyrimidine/purine to the corresponding atom in the following base. FREEHELIX brings all of these defining vectors to a common point, and passes the best least squares plane through the tips of the vectors. The viewing axis then is the vector through the origin, perpendicular to this least squares plane.

For a straight helix, the viewing axis as just defined, coincides with the overall helix axis, as shown in Figure 1a. For a helix with two sharp kinks as in Figure 1b, the use of all C1′–C1′ and N–N vectors automatically produces a viewing axis that follows the overall direction of the duplex, and this in practice has proven to be the most informative as well as the simplest choice. If defining vectors are chosen only along the bottom segment as in Figure 1c, then the viewing axis follows that segment. If the top segment is shorter as in Figure 1d, the viewing axis obtained by using all C1′–C1′ and N–N vectors will be inclined in favor of the bottom segment. In such a case the viewing axis can be tilted more like that in Figure 1b by using only defining vectors C1′–C1′ and N–N from within the middle segment. In NEWHELIX, the predecessor of FREEHELIX, this viewing axis was critical in that it was the helix axis relative to which all helical parameters were calculated. In FREEHELIX the exact axis is of lesser significance, giving only the direction from which the deformed duplex is to be observed. Hence it probably suffices to use only C1′ atoms in the axis definition, omitting N1 and N9 atoms. Figure 2 shows the viewing axis obtained for CAP using all C1′–C1′ and N1–N9 vectors.

Normal vector plots

After establishing a principal or viewing axis for the DNA duplex as described above, FREEHELIX then expresses coordinates of the structure in terms of an orthogonal set of working axes (x, y, z), with the viewing axis along z. Unit vectors along the working axes (x, y, z) are (i, j, k). (Bold face will be used to denote vectors.) A best least squares plane is fitted to each base pair, and a unit vector is erected perpendicular to this plane. The unit normal vector \( \mathbf{P}_n \) for base pair, \( n \) is defined in terms of the working axes by:

\[
\mathbf{P}_n = a_n \mathbf{i} + b_n \mathbf{j} + c_n \mathbf{k} = \cos X_n \mathbf{i} + \cos Y_n \mathbf{j} + \cos Z_n \mathbf{k}
\]

where the coefficients \( a_n, b_n, \) and \( c_n \) are in fact the cosines of the angles between the normal vector and each of the three working axes.

A normal vector plot is obtained by plotting any two of the three direction cosines against one another. It amounts to bringing the
normal vectors for all base pairs of the helix to a common origin and then viewing the distribution of the tips of these vectors, in the manner schematized in Figure 3. The most informative normal plot is that in which the cluster of vectors is observed down the principal viewing axis of the helix, or a plot of \( \cos X \) versus \( \cos Y \).

Figure 4 shows this \( \cos X/\cos Y \) plot for DNA bound to CAP (1). Three clusters of vector points are positioned from right to left across the normal plot, marking the normal vectors associated with the three straight segments of helix in Figure 2. The \( \cos X/\cos Z \) plot of Figure 5a is a view of the normal vectors up from the bottom of Figure 4. Now the points representing tips of the vectors lie on an arc that traces a unit-radius hemisphere about the origin. A better idea of the physical meaning of this plot is gained if the vectors themselves are drawn, as in Figure 5b. Now it can be appreciated that looking down on these three clusters of vectors from the top of this drawing yields Figure 4, already seen. The third normal vector plot, \( \cos Y/\cos Z \) (Fig. 5c) is relatively uninformative for CAP because normal vectors from the three helix segments are superimposed.

The FREEHELIX program emits a table of \( \cos X/\cos Y/\cos Z \) values which can be visualized by any convenient plotting program. (Illustrations in this paper were prepared using Cricketgraph on a Macintosh.) It also emits a table as given in the Appendix, showing the angles between all pairs of base pair normal vectors from one end of the helix to the other. A quick scan of this table often is diagnostic in revealing the kind of bending present in the helix under study. Appendix Table 1, for CAP, has three blocks of low angle values along the matrix diagonal, with larger angles in off-diagonal blocks. This is a natural consequence of a helix that is broken into three successive segments, each of which is relatively straight internally, but is inclined relative to its neighbors. Other patterns in vector matrices will be noted under Results.

**Local or vector parameters**

FREEHELIX calculates and prints out all of the familiar parameters that had been calculated by its predecessor NEWHELIX, defined relative to an assumed overall helix axis (the viewing axis as defined earlier). But it also emits seven new
Figure 3. Generation of a normal vector plot. (a) (Side view.) A unit-length normal vector is first established perpendicular to each base pair of the duplex. (b) (Side view.) All normal vectors are then brought to a common origin. (c) (Top view.) The ends of the normal vectors are plotted as points and viewed down the z axis, which for a straight helix coincides with the helix axis. Motion of points 3–4–5 from left to right across the normal vector plot indicates a swinging of vectors 3–4–5 in a clockwise direction in (b), and arises from the bend in helix visible in (a). In practice, if the normal vector to base pair \( n \), relative to the working axes (x, y, z), is:

\[
P_n = a_n \mathbf{i} + b_n \mathbf{j} + c_n \mathbf{k} = \cos X_n \mathbf{i} + \cos Y_n \mathbf{j} + \cos Z_n \mathbf{k},
\]

then the normal vector plot is obtained simply by plotting \( \cos X_n \) versus \( \cos Y_n \).

Figure 4. \( \cos X/\cos Y \) normal vector plot for DNA of the CAP complex, viewed down the long arrow of Figure 2. Vectors from the segment containing G1–T9 are tilted to the right, those of the middle segment G10–C19 are oriented nearly along the viewing direction, while those of the third segment A20–G29 are tilted to the left, with angles of roughly 40° between segments. Kinks between segments occur at pyrimidine/purine steps: C–A = T–G.

Figure 4. CosX/CosY normal vector plot for DNA of the CAP complex, viewed down the long arrow of Figure 2. Vectors from the segment containing G1–T9 are tilted to the right, those of the middle segment G10–C19 are oriented nearly along the viewing direction, while those of the third segment A20–G29 are tilted to the left, with angles of roughly 40° between segments. Kinks between segments occur at pyrimidine/purine steps: C–A = T–G.

quantities, which are identified with a leading V because they are obtained by vector algebra. These are VALL (the total angle between two successive base pair normal vectors), three rotations: VROL (roll), VTIL (tilt), VTWI (twist) and three translations: VSLI (slide), VSHF (shift), VRIS (rise). These rotation and translation parameters are diagrammed in Figure 6. The vector quantities VROL, VTIL, VTWI, VSLI and VRIS all become identical to the old NEWHELIX variables roll, tilt, twist, slide and rise for the special case of a straight helix axis, as will be illustrated later for roll in Figure 21b. The two sets of variables can be monitored and their differences compared in other situations where the helix is not straight.

The local axes relative to which these six parameters are calculated are defined in Figure 7a. Unit vector \( \mathbf{L} \) lies along the long axis of the base pair, which extends between purine atom C8 and pyrimidine atom C6. Vector \( \mathbf{P} \) points from strand 2 to strand 1. Unit vector \( \mathbf{S} \), perpendicular to it, is simply the base pair normal vector itself. The third unit vector along the short axis of the base pair, \( \mathbf{S} \), is defined again as the cross product: \( \mathbf{S} = \mathbf{L} \times \mathbf{P} \). Hence \( \mathbf{S} \) constitutes an orthogonal set of median unit vectors, relative to which one can calculate parameters relating base pair \( n \) to base pair \( n+1 \). A different set of median vectors is generated for comparing base pair \( n+1 \) with \( n+2 \), and so forth along the helix.

Calculation of local base step parameters then is straightforward. VALL, the total angle between base pairs \( n \) and \( n+1 \), is centered about the \( \mathbf{P} \) axis by virtue of the way in which that axis was defined. VROL (roll) is the projection of VALL onto a plane perpendicular to \( \mathbf{L} \), as in Figure 8, and VTIL (tilt) is the
The six local helix parameters calculated by FREEHELIX for every pair of adjacent base pairs: rotation about \( L_m \) (roll, VROL), \( S_m \) (tilt, VTIL) and \( P_m \) (twist, VTWI), and translation along \( L_m \) (slide, VSLI), \( S_m \) (shift, VSHI) and \( P_m \) (rise, VRIS).

Projection onto a plane perpendicular to \( S_m \). VTWI (twist) is the angle between \( L_n \) and \( L_{n+1} \), projected onto a plane perpendicular to \( P_m \). Defining the relative translation of base pairs \( n \) and \( n+1 \) by movement of the midpoint between pyrimidine C6 and purine C8, VSLI (slide) is the component of this translation along \( L_m \), VSHI (shift) is the component along \( S_m \), and VRIS (rise) is the component along \( P_m \).

The full FREEHELIX output is extensive, but an auxiliary program SELECT picks out those parameters most useful for examining bending of a duplex, and emits them as a separate file containing: (i) the CosX/CosY/CosZ normal vector components, (ii) the matrix of angles between all pairs of normal vectors (as in the Appendix), (iii) the seven vector parameters VALL, VROL, VTIL, VTWI, VSLI, VSHI and VRIS, (iv) a selection of standard NEWHELIX helix-axis-based parameters for comparison purposes and (v) a table listing base sequence and VROL, VSLI, VTWI in compact single-digit form (as in Table 1). The latter is particularly useful in providing an initial overview of

**Figure 5.** The other two normal vectors plots for CAP: CosX versus CosZ, and CosY versus CosZ. (a) The tips of normal vectors in the CosX/CosZ plot lie along an arc of a hemispherical dome around the origin. The view is up from the bottom of Figure 4. (b) Same plot, with normal vectors drawn in specifically. (c) CosY/CosZ plot, looking along the direction of the curve, or from the right in Figure 4. Now vector points are superimposed, and little information about bending is learned.

**Figure 6.** The six local helix parameters calculated by FREEHELIX for every pair of adjacent base pairs: rotation about \( L_m \) (roll, VROL), \( S_m \) (tilt, VTIL) and \( P_m \) (twist, VTWI), and translation along \( L_m \) (slide, VSLI), \( S_m \) (shift, VSHI) and \( P_m \) (rise, VRIS).
bending behavior in a new analysis, or for comparing several different analyses. Both FREEHELIX and SELECT programs are available for distribution through the Nucleic Acid DataBase, and sample input/output programs as illustrations may be obtained from the author by email if desired.

RESULTS

With FREEHELIX as a tool, and with normal vector plots as diagnostic displays, analyses have been carried out of 86 different DNA helices bound to sequence-specific proteins, using coordinates deposited in the Nucleic Acid DataBase. From these there has emerged a clearer understanding of the various types of bending that the DNA duplex can undergo, and the extent to which this bending is governed by local base sequence. A full report will follow later, but this paper is intended to present many of these conclusions, illustrated with key examples. The structures to be discussed in this paper are listed in Table 1, along with literature citations and one-digit summaries of their roll, slide and twist behavior.

Table 1. Roll/slide/twist behaviour for DNA duplexes discussed in this paper

<table>
<thead>
<tr>
<th>Major Kinks</th>
<th>Cytosine Active Protein (1) [pdb006 (A08)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Major Kinks</td>
<td>Integration Host Factor (6) [pdb080 (F11)]</td>
</tr>
<tr>
<td>2. Major Kinks</td>
<td>Human TATA-Bending Protein (10) [pdb031 (F04)]</td>
</tr>
<tr>
<td>3. Minor Kinks</td>
<td>Lambda repressor (11) [pdb010 (A09)]</td>
</tr>
<tr>
<td>4. Minor Kinks</td>
<td>MAT alpha homodomain (38) [pdb075 (B02)]</td>
</tr>
</tbody>
</table>

Key to One Digit Roll/Side/Twist Codes:

Roll (R) is expressed in 0.25 A intervals centered around zero, and twist (T) is expressed in 0.25 A intervals centered around 35°, as shown below. Indices greater in magnitude than 29 are listed as 29.

<table>
<thead>
<tr>
<th>Code</th>
<th>R (°)</th>
<th>T (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roll</td>
<td>-10</td>
<td>75</td>
</tr>
<tr>
<td>Slide</td>
<td>10</td>
<td>-75</td>
</tr>
</tbody>
</table>

Nucleic Acid DataBase serial numbers are in square brackets following references, along with identification numbers for this study in parentheses.

Downloaded from https://academic.oup.com/nar/article-abstract/26/8/1906/2902066 by guest on 19 February 2019
Major bending

Nearly all examples of bending of the DNA duplex can be classified into one of three categories: (i) smooth, continuous planar bending, (ii) continuous writhe or (iii) abrupt kinking at one or two discrete loci. For fundamental structural reasons that will be examined later, continuous bending is less common than either writhe or kinking. In cases such as CAP the bending is major and is immediately apparent upon inspection of the helix; in other cases the deformation is minor, and may easily be overlooked in a casual inspection of what appears to be an ideal B-DNA duplex.

Major kinks: CAP and IHF. Catabolite activator protein, already seen in Figures 2 and 4, is a prime example of kinking of DNA. An otherwise straight helix is deformed at two loci, the deformation involving base pair roll that is large and positive (i.e. in a direction that compresses the major groove), at two C–A = T–G steps separated by 10 bp. As the one-digit representations of roll, slide and twist in Table 1 show, this large positive roll of 9 (i.e. >22.5°) is accompanied by a reduction in twist of –5 (i.e. twist angle between 22.5 and 25°), and in one case, by a large positive slide of 5 (i.e. between 1.25 and 1.50 Å). This correlation between large positive roll, large slide, and diminished twist is in part a consequence of the finite length of the backbone chain connecting bases, as diagrammed in figure 14 of reference 18.

It will be one of the leitmotifs of this paper that pyrimidine/purine steps are particularly susceptible to roll bending because of their relative limited overlap between base pairs (see fig. 4 of reference 18, or Fig. 2 of ref. 19). This is not a new observation; it goes back to the very first B-DNA single crystal structure analysis (22), and has become part of the canon of DNA structure (23–27). Because the two kinks occur nearly one helical turn apart, they are roughly additive, and the three segments of the helix are approximately coplanar.

The internal regularity of each of the three segments of CAP DNA is shown by the fact that the vector matrix, Appendix Table 1, divides naturally into 3 × 3 blocks, with angles of 12° or less between base pair normals within one segment, 37–54° between base pairs in neighboring segments, and 60–80° between base pairs in the two outer segments. Figure 9 compares VALL, VTIL
Figure 11. Human TATA-binding protein bound to the DNA sequence: C–G–T–A–T–A–T–A–T–A–C–G (10). One full turn of ideal Arnott B-DNA helix has been added with proper registration to each end of the TATA box, to demonstrate how TBP changes the direction of the duplex. TBP itself is shown by a flat ribbon, bending at alpha carbon positions. C marks the C-terminus of the TBP chain. (a) View in the plane of the 112° bend. (b) View from the right, demonstrating that the bend is actually a three-dimensional writhe, with sidewise dislocation of the helix axis, rather than a simple planar kink. The writhe makes the central two base pairs of the TATA box nearly vertical, rather than horizontal as they would be in a simple kink.
compressing the major groove. The one-digit roll/slide/twist values in Table 1 again show that large positive roll correlates with reduced twist, and to a lesser extent with positive slide. The CosX/CosY normal vector plot in Figure 10a shows three straight segments, with breaks between base pairs 11 and 12, and between 20 and 21. But the approximation that the three clusters of vectors progress linearly across the normal vector plot is even poorer for IHF than for CAP. This means that the three helix segments of IHF are even less coplanar, and the bends between them are even less additive. For example, the normal vector angle matrix in Appendix Table 2 shows that the angle between normal vectors to the base pairs containing A8 in segment 1 and T15 in segment 2 is 73°, that between T15 and T22 in segments 2 and 3 is 80°, but the overall angle between A8 and T22 is only 141°, not 153°.

The increased non-planarity in IHF compared with CAP arises because roll kinks in IHF are spaced only 9 bp apart rather than 10.Were the kinks to be spaced 8 bp apart, or 7, the out-of-plane three-dimensional writhe of the three segments would be even more pronounced. Spacing the two kinks 5 bp or one-half turn apart would again yield a planar molecule, but with segments in an S-shape rather than the C-shape of Figure 1b.

Rice et al. (6) propose that this non-coplanarity in IHF may indeed have a biological purpose: ‘The DNA lies largely in a single plane, making a dihedral angle of only ~10–15°. While small, the handedness of this angle is consistent with the placement of IHF at a node of a negatively supercoiled plasmid…’ In sum, positioning of roll-bend elements 9 bp apart in the IHF target sequence, instead of 10, creates a small but significant writhe that may be utilized to localize the binding of IHF; in short, an element of sequence recognition.

Figure 10b plots VALL, VTIL and VROL for DNA bound to IHF. As with CAP, kinking again is seen to result from large positive localized roll, with tilt making no appreciable contribution. But in IHF the roll kink sequence is C–A–A–T = A–T–T–G instead of C–A = T–G. The angle matrix in Appendix Table 2 again segregates naturally into 3 × 3 blocks, with boundaries where expected from the normal vector plot, at base steps 11/12 and 20/21.

Major writhe: TATA-binding protein. DNA bound to the TATA-binding protein, TBP (7–10), exhibits quite different behavior: a broad writhe rather than localized kinking. This writhe can be seen clearly in the two stereo views of Figure 11. As the base pairs writhe, their normal vectors sweep in a wide half-circle across the conformation sphere (Fig. 12a), with especially large changes of direction at pyrimidine/purine steps T3–A4 and T9–A10. Normals to base pairs T3 and A10 at the two ends of the TATA box make an angle of 81°.

The single-digit roll/slide/twist data of Table 1 show that TBP DNA exhibits massive roll at nearly every step within the TATA-box sequence. The helix must be untwisted by 10–25° at every step (twist codes from –4 to –9) to accommodate this distortion. The VALL/VTIL/VROL plot of Figure 12b demonstrates the complete irrelevance of tilt compared with roll, even more so than for CAP and IHF. Hence in the normal vector plot the VALL steps are for all practical purposes pure VROL steps. The occurrence of large roll at successive steps along the helix, rather than at steps separated by one helical turn, leads directly to writhing rather than simple planar bending. The side view of the TBP vectors in Figure 13 illustrates an important difference between writhing and bending. By analogy with the terrestrial...
Figure 13. CosY/CosZ plot for TBP, showing the normal vectors of Figure 12a viewed from the right. The open half-circle of writhe now is seen to lie at constant latitude around the z viewing axis (which is vertical in this plot).

Figure 14. Illustration of the difference between a planar bend or curvature and writhe. In a plane curve, the tips of the normal vectors describe a great circle across the conformation hemisphere. In a writhe, the vector points sit on a constant-latitude circle around the z axis. For an idealized writhe, VRol is constant and Vtil is negligible. In an idealized plane curve, both VRol and Vtil show sinusoidal variations, 90° out of phase. The physical difficulty of inducing appreciable tilt between base pairs makes simple planar curvature rare in DNA, but writhe is easily produced.

In a DNA duplex with 10 bp per turn is to form a simple plane curve, then the major contribution to bending must come successively from positive roll, positive tilt, negative roll and negative tilt, at intervals of a quarter turn of helix, or approximately every 2.5 bp. That is, roll and tilt contributions must be sinusoidal in nature, with a period of 10 bp, and must be 90° out of phase. This fact was recognized more than a decade ago by Calladine and Drew (34,35). Normal vectors from a smooth, planar bend march along a great circle of the conformation sphere, as in Figure 14. In contrast, perfectly symmetrical writhe requires a constant roll and/or tilt contribution from one base step to the next; most commonly this is encountered as a continuously large roll combined with zero tilt, as with TBP. Normal vectors in a writhed helix describe, not a great circle, but a circle of constant latitude about the axis of writhe. A finite writhe such as the half-circle of TBP can lead to drastic realignment of flanking DNA segments. But if the writhe continues long enough to describe a closed loop on the CosX/CosY vector plot, then the overall direction of the writhed helix axis is unaltered. This point will be reconsidered in a later discussion of alternative interpretations of A-DNA.
**Table 2. Distribution of types of bends among protein–DNA complexes**

<table>
<thead>
<tr>
<th>Types of bends</th>
<th>Helix–turn–helix</th>
<th>Zinc finger</th>
<th>Leucine zipper</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major bends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curved</td>
<td>2</td>
<td>–</td>
<td></td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Kinked</td>
<td>6</td>
<td>–</td>
<td></td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Writhe</td>
<td>–</td>
<td>–</td>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Kinked and writhe</td>
<td>1</td>
<td>–</td>
<td></td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Total major</td>
<td>9</td>
<td>–</td>
<td></td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>Minor bending</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straight</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Minor kink</td>
<td>9</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>10</td>
</tr>
<tr>
<td>Minor writhe</td>
<td>6</td>
<td>10</td>
<td>9</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Minor curve</td>
<td>4</td>
<td>2</td>
<td>–</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Minor, compound</td>
<td>1</td>
<td>2</td>
<td>–</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Total minor</td>
<td>21</td>
<td>16</td>
<td>10</td>
<td>11</td>
<td>58</td>
</tr>
<tr>
<td>Total, all forms</td>
<td>30</td>
<td>16</td>
<td>10</td>
<td>30</td>
<td>86</td>
</tr>
</tbody>
</table>

**Major curvature: MATα1/α2 homeodomain.** Figure 1 of Li et al. (36) shows a MATα1/MATα2 homeodomain heterodimer bound on the concave side of a smoothly curved DNA duplex of two helical turns. The total bend from one end of the duplex to the other is 60°. In the authors’ words: ‘The bend in the α1/α2 binding site occurs without dramatic local distortion or kinking of the B-DNA helix. Rather, the DNA helix is smoothly bent, most noticeably at the center of the DNA fragment and in the α1 half of the binding site. The bend is largely the result of a variation in base roll, which adopts negative values near the center of the DNA site and positive values in flanking base pairs.’ As has been mentioned, smooth, planar curvature ideally requires a sinusoidal alternation of roll and tilt. But tilt is an inherently difficult deformation for the DNA duplex, because it requires that bases at one end of the stacked base pairs be pulled apart. Rolling 2 bp about their long axes is much easier, and for this reason VROL is always much more important than VTIL. But, if tilt is disfavored, how can a planar bend in the DNA duplex be accomplished? Figure 15, for DNA bound to the MATα1/α2 heterodimer, provides an answer. The normal vector plot and Appendix Table 4 show that the DNA curves through 58° from C1 to C18. The bend is a continuous curve rather than a localized kink, as it is made up of a series of small steps across the plot, rather than one or two major steps between clusters of points. Figure 15a suggests that bending is somewhat less in the first half of the helix (the α2-binding site) than the second (the α1-binding site), as Li et al. have noted. This is borne out by the angle matrix: 24° bending between C1 and T10, versus 34° between T10 and C18. But bending is truly additive, as would be expected in a plane curve: a 58° angle between C1 and C18.

The normal vector trace is linear overall, but with substantial local excursions from side to side. These local excursions hold the key to how a DNA duplex can curve smoothly without involving tilt deformations. The VALL/VTIL/VROL plot of Figure 15b shows two important features of VROL: alternating local maxima and minima, superimposed on a generally sinusoidal roll curve with a period of roughly one helical turn. No such periodicity is observed for VTIL, which hovers around zero. The strategy by which a helix can bend smoothly without a tilt contribution is for large roll values, positive or negative, to occur every 5 bp, and for roll to ‘mark time’ halfway between by zig-zagging from side to side. For example, the normal vector plot shows that, for steps 3 (T3–G4), 4 (G4–T5), 9 (T9–T10), 14 (A14–C15) and 15 (C15–A16), roll angles lie approximately parallel to the overall bend direction. Here roll is large in magnitude, positive or negative (Fig. 15b). But halfway between these regions, near base steps 6 and 7, or 14 and 15, one does not find comparably large tilt contributions. Instead, tilt remains near zero while roll oscillates from side to side in a way that does not change the overall course of the bend. See, for example, the trace of base pairs A6–A7–T8 on the normal vector plot. A11–T12–T13 half a turn later, and A16–T17–C18 still another half-turn farther along, Figure 15b shows that the alternation of high and low roll is produced by alternation of pyrimidines and purines along the sequence. Nearly all of the roll maxima occur at pyrimidine/purine steps: T–A and C–A = T–G.

The angle matrix, Appendix Table 4, shows what would be expected for a curve which is reasonably continuous, without turning back upon itself. Angle values increase smoothly from the trough along the matrix diagonal, to maxima near the upper right and lower left corners of the matrix, without the block structure encountered in kinked helix segments. This same smooth behavior is seen with TBP (Appendix Table 3) because the writhe is only a half-turn long. Later writhes of more than a full turn will exhibit a different angle matrix behavior.

**Types of bends.** In summary, three types of major bends have been encountered in the 86 helices analyzed to date from specific protein–DNA complexes: isolated roll kinks, writhe, and continuous curvature. Examples of each have just been presented. Each bend has its characteristic features. Isolated kinks and continuous writhe are brought about by roll angles between base pairs. Single kinks as with Lac operator (3) or PurR (4) change the direction of the helix. Pairs of kinks located a full turn of helix apart (as CAP) are additive, and produce a larger planar bend. Brining the two kinks closer together (as IHF) introduces non-planarity between the bent segments which can be described as incipient writhe, and in the limit of a constant large roll at every step, the result is an ideal writhe (as TBP). One can view the normal vector...
Minor bending

Of the 86 sequence-specific protein/DNA associations examined, 28 exhibit major kinks, writhes or curves of the type just described, as summarized in the first half of Table 2. (The statistics are skewed by the presence of 10 different TBP complexes, all identically writhed. Note also the absence of major bending in zinc and leucine zipper proteins, for reasons to be discussed later.) The remaining 58 DNA helices in the second half of the table possess smaller deformations that in some cases would not be recognized as bending at all upon casual inspection of the protein–DNA complex. But even in these cases, the minor bending that is present can be classified into kinking, writhe or curvature, and its origin is predominately the same: positive roll (compressing the major groove), most frequently at pyrimidine/purine steps. Tilt becomes relatively more significant in minor bending, not because it is larger, but merely because the smaller roll values sinks closer to the noise level of tilt seen with major bending.

Figure 16. Normal vector plot for a straight helix, DNA bound to the even-skipped homeodomain (37). (a) Full plot, for comparison with Figures 4, 10a and 11a. (b) Enlargement of center, with limits ±0.3 in CosX and CosY. The tips of the normal vectors describe what is essentially a random walk about the viewing axis.

Figure 17. Minor kinks or curvature in the DNA bound to lambda repressor (11). Note that vector points are confined to the ±0.3 square in the center of the plot, as in Figure 16b. Two small non-coplanar bends of 16° and 19° result in an overall bending of 25° from one end of the helix to the other. Pyrimidine/purine steps are marked by heavier lines. An essentially identical normal vector plot is seen in figure 12 of reference 11.

Figure 18. Planar curvature is a different matter. In default of significant tilt contributions, curvature is produced by quarter-turn alternations of positive roll, zig-zag roll oscillation, negative roll and zig-zag oscillation again. Kinking and writhe are easy operations to achieve with DNA and are encountered frequently; planar curvature is more difficult to achieve, and is correspondingly less common.

Minor bending

Of the 86 sequence-specific protein/DNA associations examined, 28 exhibit major kinks, writhes or curves of the type just described, as summarized in the first half of Table 2. (The statistics are skewed by the presence of 10 different TBP complexes, all identically writhed. Note also the absence of major bending in zinc and leucine zipper proteins, for reasons to be discussed later.) The remaining 58 DNA helices in the second half of the table possess smaller deformations that in some cases would not be recognized as bending at all upon casual inspection of the protein–DNA complex. But even in these cases, the minor bending that is present can be classified into kinking, writhe or curvature, and its origin is predominately the same: positive roll (compressing the major groove), most frequently at pyrimidine/purine steps. Tilt becomes relatively more significant in minor bending, not because it is larger, but merely because the smaller roll values sinks closer to the noise level of tilt seen with major bending.

Straight helix: even-skipped homeodomain. The even-skipped homeodomain (37) is a good example of a straight, unbent DNA helix in a protein complex. Two HTH homeodomains bind to the major groove one-half turn apart, on opposite sides of the helix. Hence any curvature induced by one homeodomain would be canceled by binding of the other. This is most probably the reason why bending is absent from multiple zinc finger complexes and from leucine zipper and bHLH complexes.

The normal vector plot for the even-skipped homeodomain, Figure 16a, shows only a dense cluster of vector points around the origin. Even when the central region of the plot is enlarged as in Figure 16b, no systematic pattern is seen, but merely an apparent random walk of base pair orientations from one step to the next. This will be the standard enlargement of the center of the normal vector plots to display minor bending, with limits between ±0.3 and ±0.3 in both CosX and CosY. Helices whose normal vectors fall entirely within this inner 10% of the plot will be classified, by definition, as exhibiting only minor bending. It is important, when
Figure 18. Schematic of the lambda operator duplex, as viewed into the unrolled major groove. Numbering on the left is that of Beamer and Pabo (11); that on the right is as used in this paper. The helix in this schematic is oriented like that in stereo figure 4 of reference 11. Roll indicates positions of appreciable roll kinking, compressing the major groove. \( \tau_a, \tau_b, \text{etc.} \) mark phosphates that are protected from ethylation by hydrogen-bonding to the repressor, primed in the non-consensus sequence (below), and unprimed in the consensus sequence (above).

Looking at such plots, to recall the difference in scale between Figure 16a and 16b, and to remember that one is seeing only the central 10% of the entire normal plot domain.

**Minor kinks: lambda repressor.** Ten of the 58 minor-bend helices exhibit, on a reduced scale, what clearly are local kinks. A good example is the lambda repressor-operator complex (11), shown in Figures 17 and 18, example 6 of Table 1, and Appendix Table 6. Base pairs 1–5, 6–13 and 15–19 constitute three straight segments, with angles of no more than 9° between base pairs within each segment. Adjacent segments are reoriented by ∼16–23°. As with major kinks, the deformation is brought about primarily by positive roll at discrete pyrimidine/purine steps. Table 1 exhibits single-digit roll codes of +3 (i.e. 7.5–10°) at C5–A6 and T14–G15. T14 is intermediate between the second and third segments, as is shown clearly by the normal vector plot. In fact, the total bend between segments 2 and 3 is produced by a combination of T14–G15 and the preceding G13–T14. The kinks are sufficiently sharp that the vector angle matrix, Appendix Table 6, again falls naturally into 3 × 3 blocks just as for CAP and IHF, even though the overall bending involved is far smaller, ∼25° rather than 80–150°.

The two minor kinks in lambda operator are almost exactly three-quarters of a helical turn apart: at step 5 and conjointly at steps 13 and 14 (Fig. 18). Hence, if these were major kinks, the dihedral angle between the outer two segments would be 90°. This dihedral angle is clearly visible in the L-shaped trace on the normal vector plot. But all of these bends fall within ±0.3 limits in the center of the normal vector plot. Do such small roll kinks have any real structural significance?

The answer is that they do have a subtle effect in fitting the lambda operator duplex to its repressor. As discussed in Beamer and Pabo (11) and illustrated in their stereo Figure 4, the repressor
Figure 20. Idealized A- and B-DNA, analyzed with FREEHELIX, and compared with the human TATA-box bound to TBP. Non-zero roll angles in B-DNA contribute a minor writhe with amplitude of $5^\circ$. The $21^\circ$ inclination of base pairs in this fiber-derived A-DNA helix, acted upon by helix rotation, results in a constant base pair roll angle $VROL$ of $11^\circ$, and a writhe of amplitude $22^\circ$. Note that a continuous writhe does not change the overall direction of the helix axis. The human TATA box has a writhe of amplitude double that of A-DNA, and a larger mean Inclination of $\sim 40$–$60^\circ$. (TATA box points from Fig. 12a.)

Minor writhe: MATα2 homeodomain. The 19 bp DNA duplex bound to the MATα2 homeodomain (38) exhibits two full turns of writhe, as seen in its normal vector plot, Figure 19a. The amplitude of this writhe, however, is only $\sim 16^\circ$, or one-fifth that of the TBP TATA-box. As Figure 19b shows, this writhe is produced, not by continuously high roll, but by an alternation of high positive roll at pyrimidine–purine steps, alternating with zero roll at purine–pyrimidine steps. In other words, pairs of parallel purine–pyrimidine base pairs roll as a unit.

Figure 21. Comparison of (a) base pair inclination, (b) roll, (c) X displacement and (d) twist for ideal A-DNA (triangles), ideal B-DNA (circles) and the human TATA-box (squares). (b) Large open symbols indicate the vector quantity $VROL$, while small solid symbols indicate conventional Roll as calculated by NEWHELIX (and also by FREEHELIX).
Figure 22. Illustration of the ability of the EcoRV restriction enzyme to discriminate among different DNA sequences. (a) The non-cognate sequence C–G–A–G–C–T–G–C binds to the enzyme as a straight and unbent helix. (b) When the cognate sequence A–A–G–A–T–A–T–C–T–T binds to EcoRV, the enzyme produces a large 55° kink at the central pyrimidine/purine T–A step. As usual, this kink is a roll bend compressing the major groove.

The result of these roll bends is a gentle writhe of base pairs inclined slightly away from perpendicularity to the overall helix axis, without producing a net bend in that axis. Indeed, Wolberger et al. (38) describe their DNA as ‘essentially B-form DNA, with only minor distortions in the double helix’. [Note: they use the term ‘tilt’ for what the 1988 Cambridge accords (39–42) define as ‘inclination’, or the angle that an individual base pair makes away from perpendicularity to the helix axis by virtue of rotation about its short axis.] They conclude, ‘The net effect of the tilt (sic) at each end of the operator is to bend the DNA slightly around the recognition helix of α2.’ In agreement with this, Appendix Table 7 shows that the normals to base pairs make an angle of 20° to one another.

Appendix Table 7 shows one further consequence of writhe that extends over more than one turn of helix. The principal diagonal of the matrix is a trough of low angle values, because adjacent base pairs tend to be oriented in the same direction. Two other diagonal troughs of low angle values run parallel to this, one turn of helix away to the upper right and lower left of the matrix. These subsidiary minima mark similar orientations of base planes one turn of helix apart, where the writhe brings base pairs back into something like their original orientation.

Continuous writhe in A-DNA. To emphasize the fact that continuous minor writhe need not change the overall direction of a DNA helix, consider the comparison of ideal, fiber-derived A-DNA and B-DNA in Figures 20 and 21. In the fiber A-DNA model, base pairs have an average inclination of 21° (double that found in most A-DNA oligonucleotide crystal structures), and this coupled with helix rotation leads to a 11° roll angle between successive base pairs. Base pair normal vectors move around the normal vector plot as shown in Figure 20, describing a circle of radius 0.35. By comparison, the fiber B-DNA helix has essentially zero inclination (–6°) and roll (–3°), and its vector points are clustered tightly around the origin. It is well to keep in mind that what we have been calling ‘minor bending’, with vector points confined to the ±0.3 center of the normal plot, is comparable to the level of deformation exhibited by A-DNA in comparison with B.

B-DNA is properly described as a stacking of base pairs essentially perpendicular to the helix axis. But A-DNA can be described in two different ways: (i) base pairs writhed around a straight and unbent helix axis, or (ii) base pairs stacked perpendicular to their local helix axis, which itself is writhed around a straight global axis. Both descriptions are correct and are only matters of preference. Indeed, Guzikevich-Guerstein and Shakked (43) have even taken the reverse route and interpreted the central writhed segment of the TATA box as a straight A-like helix, joined at each end to normal B-DNA.

Figures 20 and 21 show the degree to which this latter characterization is just. The normal vector plot of Figure 20 reveals immediately that the degree of writhe found in the TATA box is double that of A-DNA. Correspondingly, the inclination angle of base pairs in Figure 21a varies between 40° and 60°, again double the 21° of fiber A-DNA, which itself in turn is larger than the inclination usually seen in crystal structures of A-DNA oligonucleotides. X-displacement of base pairs (Fig. 21c) is in the same direction as A-DNA (helix axis running parallel to the major groove edge of each base pair), but the magnitude of displacement again is larger. Roll and twist angles (Fig. 21b and d) are sequence-dependent, and vary so much that one is scarcely justified in drawing any conclusions from them as to helix type.

In summary, if one prefers the straight-axis picture, one must say that the TATA-box is not A-DNA; it is a kind of hyper-A helix, with double the parameter excursions found in normal A-DNA.

DISCUSSION

Of the 86 examples of DNA bound to sequence-specific protein molecules, 28 can be described as major bending (Table 2), and the other 58 as minor bending comparable to the deformations observed in A-DNA. They permit three major generalizations about DNA bending.
Table 3. Dihedral angles versus kink spacing in doubly-kinked DNA

<table>
<thead>
<tr>
<th>Protein/DNA Complex</th>
<th>DNA Sequence</th>
<th>No. Steps Between Kinks</th>
<th>Dihedral Angle</th>
</tr>
</thead>
</table>
Figure 23. Relationship between the dihedral angles in a doubly-kinked DNA chain and the appearance of the normal vector plot. (Top row) Each of the internally straight DNA segments is represented by an arrow. Dihedral angles at top are angles between planes defined by segments 1 and 2, and 2 and 3. (Bottom row) All of the normal vectors from base pairs within one straight segment form a cluster of points, here symbolized by a large dot labeled 1, 2 or 3. The normal vector plot may be thought of as resulting from bringing all three of the segment axis arrows to the origin of coordinates, with labeled dots marking the arrowheads. Examples approximating each of the indicated dihedral angles are given below.

But does this sequence-dependent bending behavior have anything to do with the recognition of DNA by proteins? Sequence-dependent differential bendability obviously is a factor in the correct positioning of proteins that bend the DNA duplex. As the most simplistic possible examples, CAP, IHF and TBP could not bind DNA in the same manner if their hinge points were replaced by A-tracts. Even the rather subtle minor kinking of the lambda repressor assists in bringing the phosphate backbones closer to the repressor protein. No one imagines however that a CAP site, or an IHF site, wanders through an aqueous solution with pre-formed bends, searching for its target protein. CAP and IHF target DNAs are not intrinsically bent; they are inherently bendable.

Sequence discrimination by relative bendability: EcoRV

A particularly elegant example of differential binding of a protein to target and non-target DNA is provided by the EcoRV restriction endonuclease. When bound to its cognate or target sequence: A–A–G–A–T–A–T–C–G–T, it induces a major 55° kink at the central T–A step (Fig. 22b). Yet substitution of the non-cognate sequence: C–G–A–G–C–T–C–G–C–G–A–G–C–T–C–G produces no such kinked bending (Fig. 5 of ref. 16), even though the central G/C step is the unbounded stacking junction of two separate octamers. A pyrimidine–purine step such as T–A is bendable, whereas a purine–pyrimidine G–C step resists bending even when phosphate backbone connectors are absent. This illustrates once again the overwhelming importance of base pair stacking in determining B-DNA helix structure, compared with the sugar–phosphate backbone.

Kink positioning as a geometric tool

Brief mention has been made that, in doubly kinked DNA, the dihedral angle between the outermost helical segments depends on the spacing between kinks. This is quantified in Figure 23 and Table 3. Figure 23 schematizes a helix made up of three helical segments. The dihedral angle, viewed around the central segment 2, is the angle between the plane defined by segments 1 and 2, and that defined by segments 2 and 3. A dihedral angle of 0° means that the three segments are coplanar in a C-shape. A 180° dihedral angle means that the segments are again coplanar, but this time arranged in an S-shape. At intermediate dihedral angles the segments are not coplanar, but are as drawn in Figure 23. The bottom diagrams in that figure demonstrate the appearance of a normal vector plot for each configuration, with the three numbered dots representing clusters of vector points for base pairs in each of the three segments (compare Fig. 4 or 10a).

If B-DNA is assumed for purposes of discussion to have 10 bp per turn, then this dihedral angle behavior should have a period of 10 bp steps. Roll kinks 10 base steps apart, or 20 steps, should exhibit 0° dihedral angle, whereas kinks 5 or 15 steps apart should show 180° angle behavior. Table 3 lists 10 examples of protein–DNA complexes that display roll kinks located 2–10 base steps apart, and Figure 24 shows that these spacings correlate...
strongly with dihedral angles as measured from normal vector plots.

In sum, facultative bending of B-DNA helices by control proteins may be achieved by selection of the correct dinucleotide steps, and by positioning those steps at suitable intervals along the helix. This differential bendability of the duplex is an important component of protein recognition, to be set alongside the formation of key patterns of hydrogen bonding. These are two elements of the grammar of recognition, by which a DNA duplex "talks" to its protein. It remains to be seen whether other comparably important grammatical elements exist.

ACKNOWLEDGEMENTS

I thank Sir Edward Abraham and the Rector and Fellows of Lincoln College, Oxford, for their hospitality in welcoming me as the Newton-Abraham Visiting Fellow during the 1997–1998 academic year in which this work was carried out. I also thank Prof. Louise N. Johnson, Director of the Molecular Biophysics Laboratory, for providing a home for a wandering scientist and for welcoming me into the life of the Laboratory. Special thanks are due Dr Anke Gelbin of the Nucleic Acid DataBase, for yeoman service over the past two years in providing endless data files upon request by email, and to Thang Chiu for help in preparing illustrations. Finally, I thank Drs Keith R. Fox and Andrew Lane for an invitation to a British Biophysical Society meeting that provided the first public forum for these ideas.

REFERENCES


APPENDIX. ANGLES BETWEEN ALL NORMAL VECTOR PAIRS

1. Catabolite Activator Protein


APPENDIX. ANGLES BETWEEN ALL NORMAL VECTOR PAIRS

Downloaded from https://academic.oup.com/nar/article-abstract/26/8/1906/2902066 by guest on 19 February 2019