1906–1926 *Nucleic Acids Research, 1998, Vol. 26, No. 8*

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 *Eco*RV 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

*Tel: +44 1865 275372; Fax: +44 1865 275182; Email: red@biop.ox.ac.uk. Permanent address: Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90095-1570, USA. Tel: +1 310 825 5864; Fax: +1 310 267 1957; Email: red@ewald.mbi.ucla.edu

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

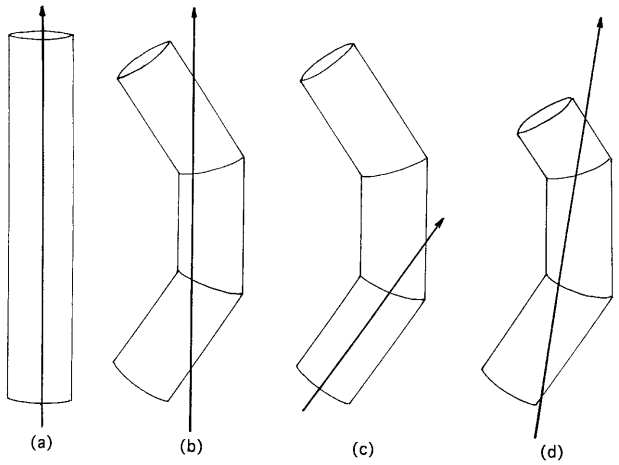


Figure 1. The FREEHELIX analysis program defines a viewing direction for a straight, curved or kinked helix, using vectors from one C1′ atom to the next along each strand of duplex, and from one N1/N9 (pyrimidine/purine) to the next (see Materials and Methods). (**a**) If the helix is straight, then the viewing direction coincides with the best overall helix axis. (**b**) If the helix has three approximately equal segments with kinks as shown (i.e. CAP or IHF), then the viewing direction is roughly along the helix axis of the central segment. (**c**) If only C1′/C1′ and N1/N9 vectors from the first segment are used, then the viewing direction coincides with the axis of that segment. (**d**) If the first segment is substantially larger than the third, then the viewing direction is inclined as shown. It can be realigned (b) by using only C1′/C1′ and N1/N9 vectors from the central segment to define the viewing direction. FREEHELIX generates a working set of coordinates in orthogonal axes in which the viewing direction lies along z, and strand 1 of the helix rises to increasing z values.

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:

$$\mathbf{P}_n = a_n\,\mathbf{i} + b_n\,\mathbf{j} + c_n\,\mathbf{k} = \mathrm{Cos}X_n\,\mathbf{i} + \mathrm{Cos}Y_n\,\mathbf{j} + \mathrm{Cos}Z_n\,\mathbf{k}$$

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

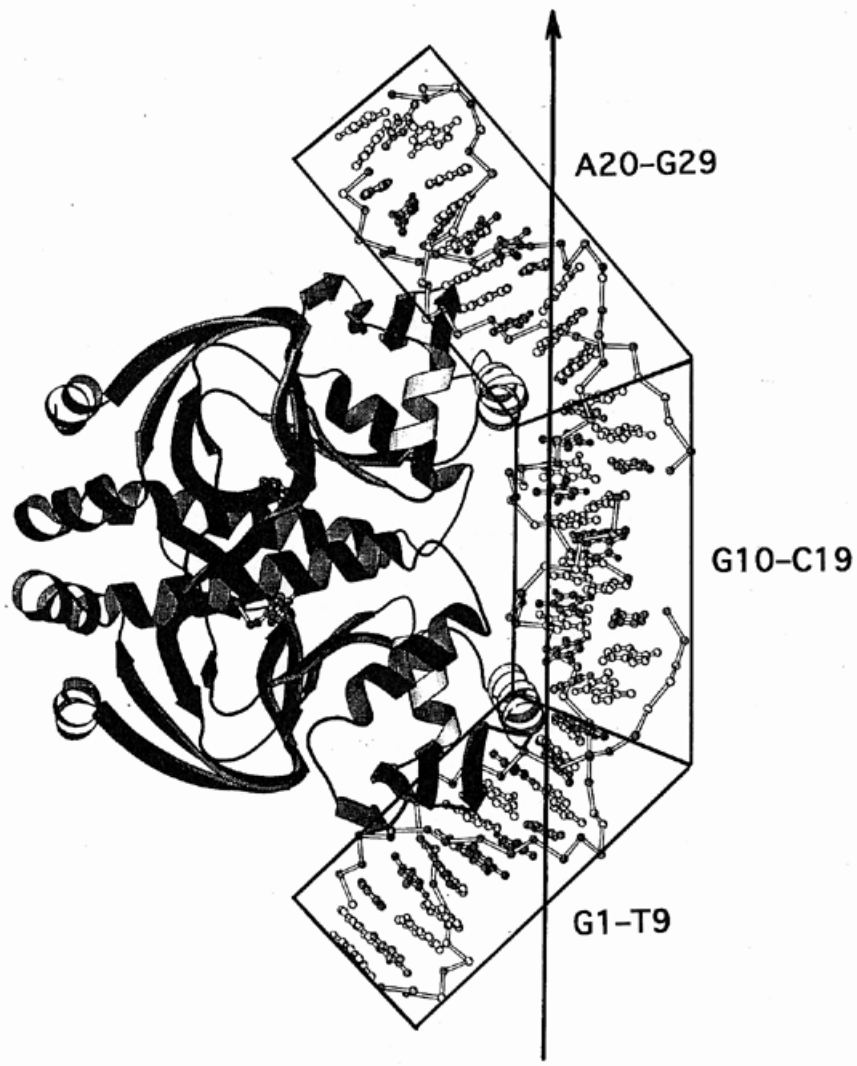


Figure 2. Construction of a viewing direction (long arrow) along the doubly kinked DNA duplex bound to CAP (1). The viewing direction was generated as discussed in Figure 1b. Base sequences of strand 1 of the duplex are marked. Drawing adapted from Parkinson *et al.* (2).

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 $CosX_n$ versus $CosY_n$.

Figure 4 shows this CosX/CosY 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 CosX/CosZ 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, CosY/CosZ (Fig. 5c) is relatively uninformative for CAP because normal vectors from the three helix segments are superimposed.

The FREEHELIX program emits a table of CosX/CosY/CosZ 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

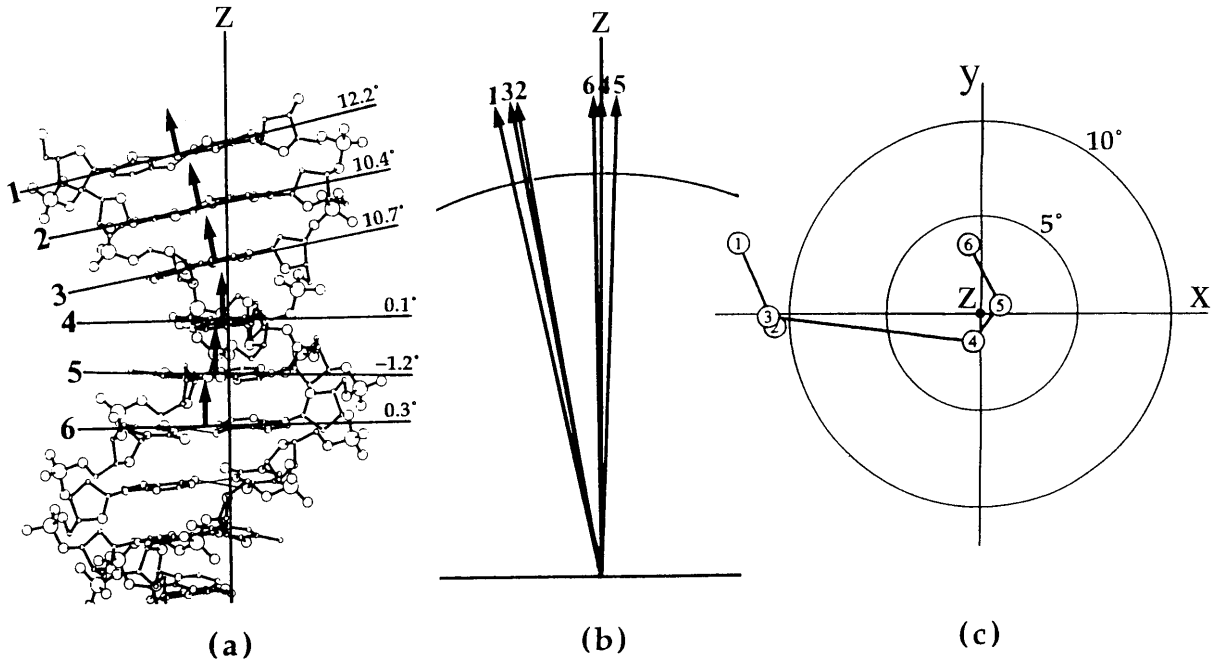


Figure 3. Generation of a normal vector plot. (**a**) (Side view.) A unit-length normal vector is first established perpendicular to each base pair of the duplex. (**b**) (Side view.) All normal vectors are then brought to a common origin. (**c**) (Top view.) The ends of the normal vectors are plotted 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: $\mathbf{P}_n = a_n\,\mathbf{i} + b_n\,\mathbf{j} + c_n\,\mathbf{k} = \mathrm{CosX}_n\,\mathbf{i} + \mathrm{CosY}_n\,\mathbf{j} + \mathrm{CosZ}_n\,\mathbf{k}$, then the normal vector plot is obtained simply by plotting CosX_n versus CosY_n.

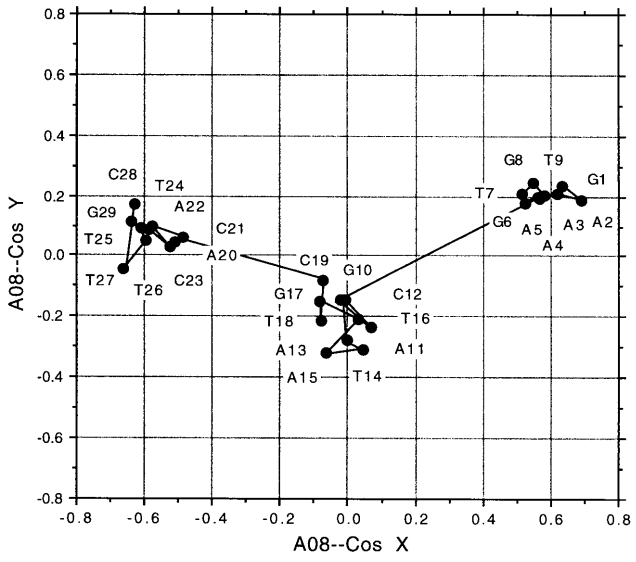


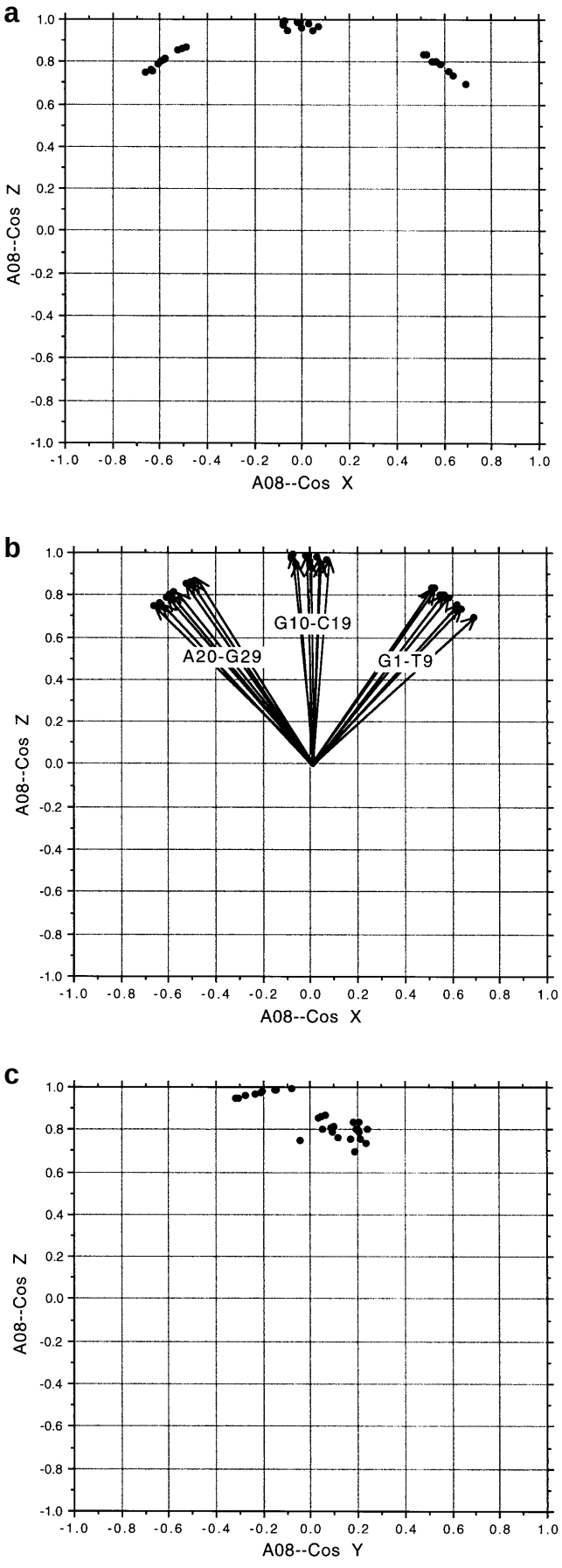
Figure 4. CosX/CosY normal vector plot for DNA of the CAP complex, viewed down the long arrow of Figure 2. Vectors from the segment containing G1–T9 are tilted to the right, those of the middle segment G10–C19 are oriented nearly along the viewing direction, while those of the third segment A20–G29 are 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 **L** lies along the long axis of the base pair, which extends between purine atom C8 and pyrimidine atom C6. Vector **L** points from strand 2 to strand 1. Unit vector **P**, perpendicular to it, is simply the base pair normal vector itself. The third unit vector along the short axis of the base pair, **S**, is defined by the cross product: $\mathbf{S} = \mathbf{L} \times \mathbf{P}$.

With unit vectors defined for two successive base pairs, ($\mathbf{L}_n$, $\mathbf{P}_n$, $\mathbf{S}_n$) and ($\mathbf{L}_{n+1}$, $\mathbf{P}_{n+1}$, $\mathbf{S}_{n+1}$), a median axis set ($\mathbf{L}_m$, $\mathbf{P}_m$, $\mathbf{S}_m$) is computed, relative to which the six vector parmeters will be calculated. As shown in Figure 7b, $\mathbf{L}_m$ and $\mathbf{P}_m$ are chosen halfway between the equivalent vectors of the individual base pairs, and $\mathbf{S}_m$ is defined again as the cross product: $\mathbf{S}_m = \mathbf{L}_m \times \mathbf{P}_m$. Hence ($\mathbf{L}_m$, $\mathbf{P}_m$, $\mathbf{S}_m$) constitute 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 $\mathbf{P}_m$ axis by virtue of the way in which that axis was defined. VROL (roll) is the projection of VALL onto a plane perpendicular to $\mathbf{L}_m$, as in Figure 8, and VTIL (tilt) is the



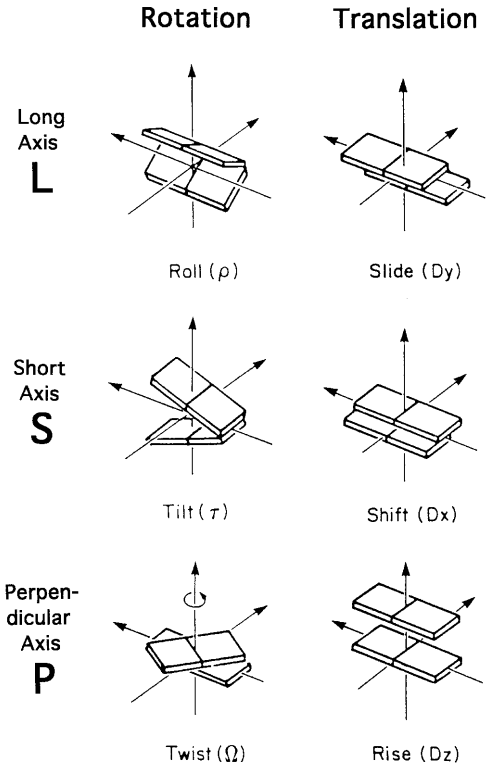


Figure 6. The six local helix parameters calculated by FREEHELIX for every pair of adjacent base pairs: rotation about $\mathbf{L}_m$ (roll, VROL), $\mathbf{S}_m$ (tilt, VTIL) and $\mathbf{P}_m$ (twist, VTWI), and translation along $\mathbf{L}_m$ (slide, VSLI), $\mathbf{S}_m$ (shift, VSHI) and $\mathbf{P}_m$ (rise, VRIS).

projection onto a plane perpendicular to $\mathbf{S}_m$. VTWI (twist) is the angle between $\mathbf{L}_n$ and $\mathbf{L}_{n+1}$, projected onto a plane perpendicular to $\mathbf{P}_m$. 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 $\mathbf{L}_m$, VSHF (shift) is the component along $\mathbf{S}_m$, and VRIS (rise) is the component along $\mathbf{P}_m$.

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, VSHF 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.

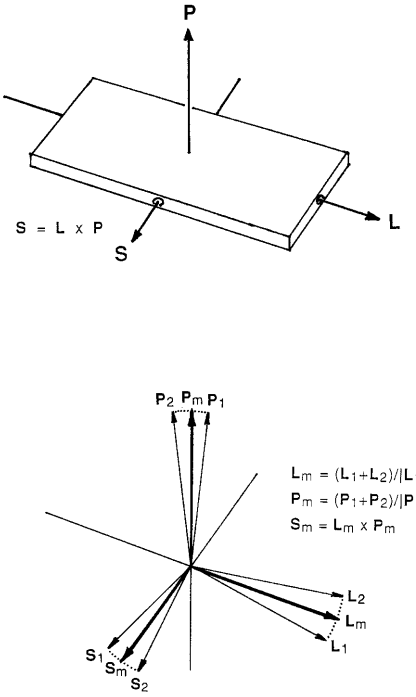


Figure 7. Definitions of local axes (**L**, **P**, **S**). (**a**) **L** is a unit vector along the long axis of the base pair between purine C8 and pyrimidine C6 atoms. **P** is the unit normal vector described earlier. **S** is the base pair short axis, defined by: **S** = **L** × **P** (bold face denotes vectors). (**b**) An orthonormal mean vector set ($\mathbf{L}_m$, $\mathbf{P}_m$, $\mathbf{S}_m$) is defined between each pair of adjacent base pairs. All local helix parameters between the 2 bp are calculated with reference to this mean vector set, and no assumptions about helix axes, global or local, are necessary.

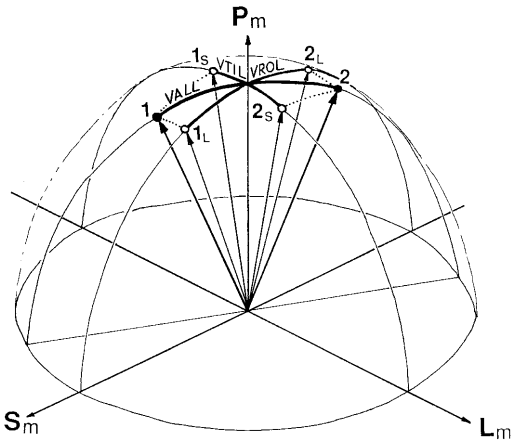


Figure 8. Factorization of the total angle between adjacent normal vectors, VALL, into a roll component VROL and a tilt component VTIL. VROL is the projection of VALL onto the plane perpendicular to $\mathbf{L}_m$, and VTIL is the projection of VALL onto the plane perpendiclar to $\mathbf{S}_m$. It is approximately true that: $(\text{VALL})^2 = (\text{VROL})^2 + (\text{VTIL})^2$.

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,

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

1. Major Kinks: Catabolite Activator Protein (1) [pdr006 (A08)]

```
                5         10        15        20        25
  G-A-A-A-A-G-T-G-T-G-A-C-A-T-A-T-G-T-C-A-C-A-C-T-T-T-T-C-G
R: 1-1-1 0-0 0-0-0 9 0-2 2 1-2 3 0-1 1 9-0 0 1-1-0 0-2 5-1
S: 5-0-1-1-1-2 4 2 2 2-3-0 0 7-1-1-4 3 5 0 3-3-3-1-1 1 4 2
T: 0 1 0 0 0-1 3-2-5-0-0-1-2 2-2 2-2 0-5-3 0 0-0 0 1 0-0 0
```

2. Major Kinks: Integration Host Factor (6) [pdt040 (F11)]

```
          5         10        15        20        25        30
  G-T-G-C-A-A-C-A-A-A-A-T-T-G-A-T-A-A-G-C-A-A-T-G-C-T-T-T-T-T-G-G-C
R: 0-1 2-3-0-1 5 2 2 5 9 3-0-1-1-0 0-0-2 9 3-0 0 0-0 0-1-1 0 3 4 1
S:-1 5 1 7-0-4-3-2-0-0 3 3 3-1-3-2-1 0 9 2 1 2-0-2-2-1-1-1-1 3-1-2
T:-1 2-4 5 0 0 0-2-2-3-6-2-1-0 1-0 1-0 9-9-4-0-0 1-0 0 1 0-0-0-0-1
```

3. Major Writhe: Human TATA-Binding Protein (10) [pdt024 (E04)]

```
            5         10
  C-G-T-A-T-A-T-A-T-A-C-G
R: 2-0 9 9 1 8 9 4 9 5 0
S: 1-2-1-1 7 2 6-1-5 0 4
T: 2-0-7-6-5-9-5-4-7-8 0
```

4. Major Curvature: MAT a1/alpha2 homeodomain (36) [pdt028 (B03)]

```
            5         10        15
  C-A-T-G-T-A-A-T-T-T-A-T-T-A-C-A-T-C-A
R: 2 0 4 0 2 2-2 1-5-0-2 1 5 2 5-0 3 2
S: 1-1-3-2-0-0-2-1-1 0-1-0 4-2 0-2-2 0
T:-0-0 1-4 2-0-2 2-0 3-1 0-1-2 0-1-0-0
```

5. Straight Helix: Even-skipped homeodomain (37) [pdt031 (B04)]

```
            5
  T-A-A-T-T-G-A-A-T-T
R: 2-0-0-0 1-1 3-2 0
S: 4-0-1-2 3 5-1-1-0
T: 1 0-0-0 1 1-0-0-1
```

6. Minor Kinks: Lambda repressor (11) [pdr010 (A01)]

```
            5         10        15
  A-T-A-C-C-A-C-T-G-G-C-G-G-T-G-A-T-A-T
R:-0 0 0 1 3 0 1-1 2-1-2 1 2 3-0-1 1-0
S:-2-1-2-1 3-3 0 9 3 3 3-5-4 0-0-2-1-2
T:-1 1-1-0-0-1-0 5-4 2 0-5-1-0 1-1 1-1
```

7. Minor Writhe: MAT alpha2 homeodomain (38) [pdt005 (B02)]

```
            5         10        15
  C-A-T-G-T-A-A-T-T-C-A-T-T-T-A-C-A-C-G-C
R: 3-0 4 0 3 1-1 0 2 1-1 2 2 1 1 4-1 4 3
S: 1-0-2-1-1-0-1-0 0 2-2-1 1-1-1-2 0-0-1
T: 0-1 1-1-0-0 1-0 0-2 0 0 0 0-2 1-2-0 0
```

8,9. Sequence Discrimination: Eco RV

(non-cognate sequence) (16) [pde002 (F06A)]

```
    5
  C-G-A-G-C-T-C-G
R: 2 0 0-0 0 1-0
S:-3-0 1 0 1 7 5
T: 1-0 1 0-0 1-0
```

(cognate sequence) (17) [pde015 (F09)]

```
    5         10
  A-A-G-A-T-T-C-T-T
R:-0 0-4 0 9-0-2 0 1
S: 2-0-1-3-6-3-3-0 2
T:-1-1 1-3-4-2 1-1-2
```

Key to One-Digit Roll/Slide/Twist Codes:

Roll (R) is expressed in 2.5° intervals centered around zero, slide (S) is expressed in 0.25 Å intervals centered around zero, and twist (T) is expressed in 2.5° intervals centered around 35°, as shown below. Indices greater in magnitude than ±9 are listed as ±9.

Code:	-4	-3	-2	-1	-0	0	1	2	3	4	5	6	7	8	9	
Roll:	-10	-7.5	-5	-2.5	0	2.5	5	7.5	10	12.5	15	17.5	20	22.5+		degr
Slide:	-1.00	-0.75	-0.50	-0.25	0	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25+		angs
Twist:	25	27.5	30	32.5	35	37.5	40	42.5	45	47.5	50	52.5	55	57.5+		degr

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

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.

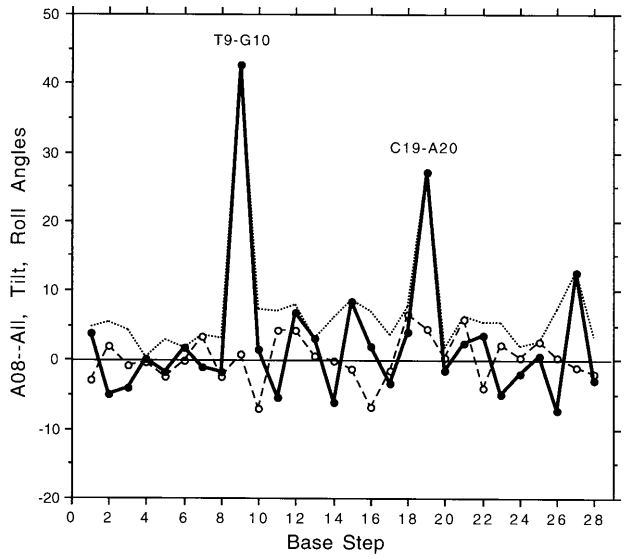


Figure 9. Plot of total angle (VALL, dotted lines), tilt (VTIL, dashed lines) and roll (VROL, solid lines) for the DNA duplex bound to CAP. VROL and VTIL are signed quantities, whereas VALL is only a magnitude. VROL is positive when the minor groove is opened and the major groove is compressed. VTIL is positive when base pairs separate at their strand 1 ends. Note that bending in CAP DNA occurs almost entirely at two kinked pyrimidine/purine steps: C–A = T–G, and involves positive roll. Because the two kinks are 10 bp or approximately one helix turn apart, they are cooperative, and the three helix segments are effectively coplanar. This is expressed by the nearly linear arrangement of the three clusters of normal vector points in Figure 4.

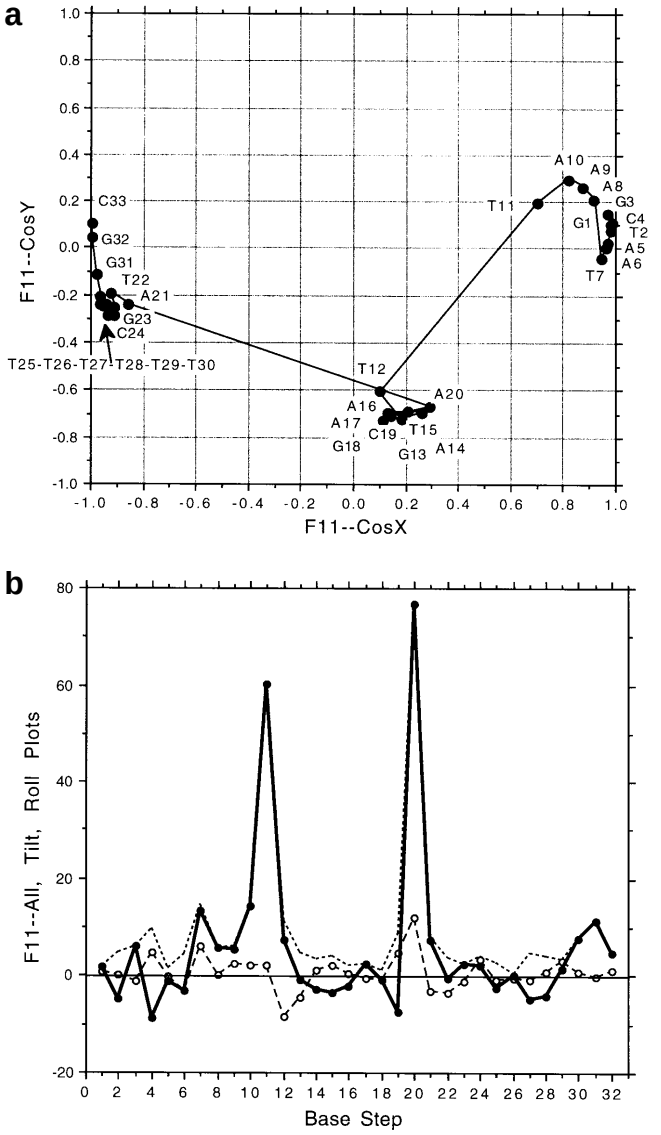


Figure 10. Major kinking in the DNA duplex bound to integration host factor, IHF (6). (**a**) CosX/CosY normal vector plot. Like CAP, IHF has three relatively straight helix segments separated by very large kinks. In IHF the kinks are only 9 bp apart, so the three segments do not share a common plane, but are twisted into an incipient writhe (of segments, not of individual base pairs). Note that the A-tract, T25→T30, is straight and unbent, with normal vectors that are virtually superimposed. (**b**) VALL/VTIL/VROL plot. As with CAP, the two kinks involve almost entirely positive roll, in a direction that compresses the broad major groove. Tilt makes only a minor contribution. Roll kinks appear at steps: C–A–A–T = A–T–T–G. Solid and dashed lines as in Figure 9. The A-tract at right, base steps 25–29, has essentially zero roll and tilt.

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 backbbone 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

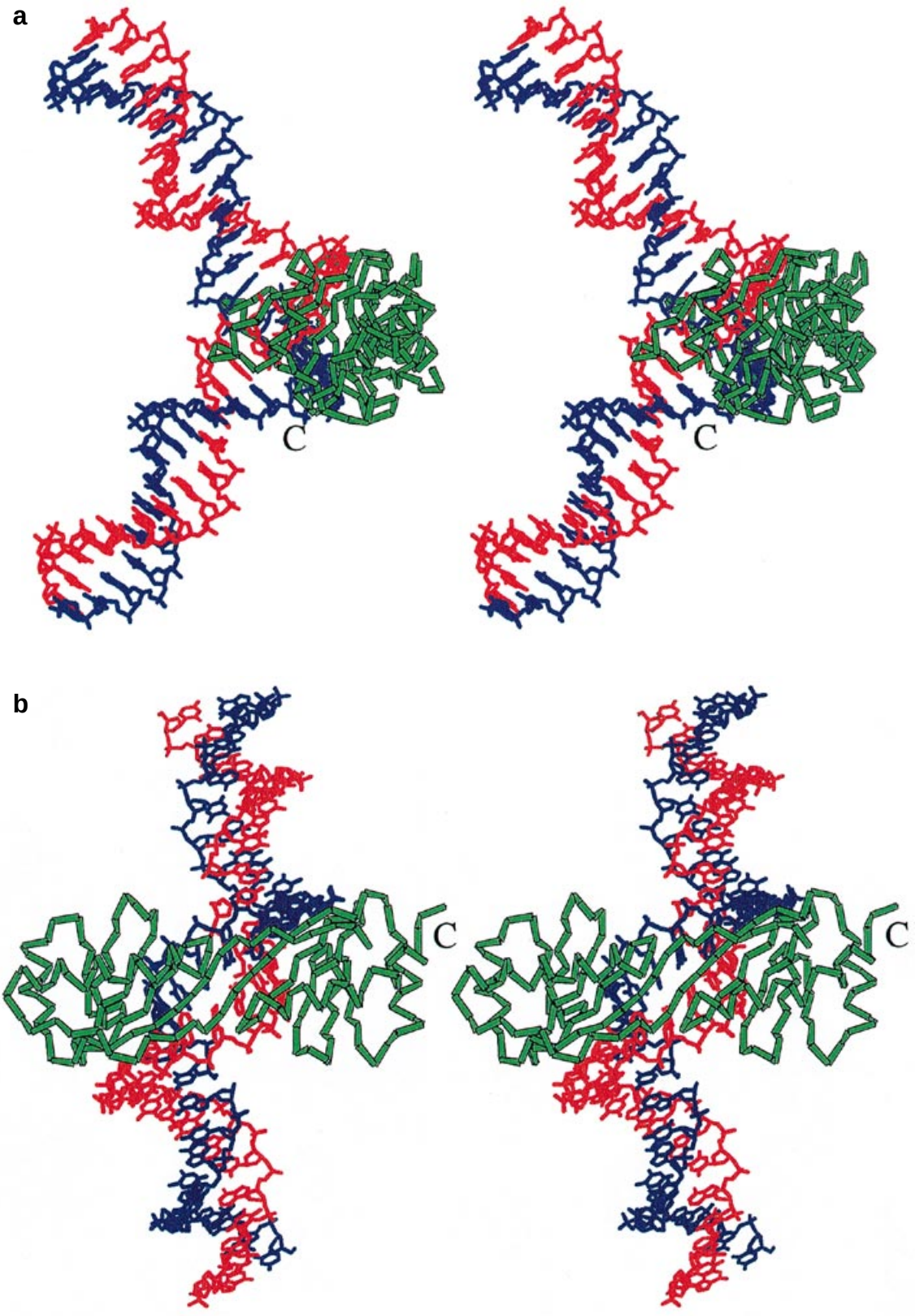


Figure 11. Human TATA-binding protein bound to the DNA sequence: C–G–T–A–T–A–T–A–T–A–C–G (10). One full turn of ideal Arnott B-DNA helix has been added with proper registration to each end of the TATA box, to demonstrate how TBP changes the direction of the duplex. TBP itself is shown by a flat ribbon, 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.

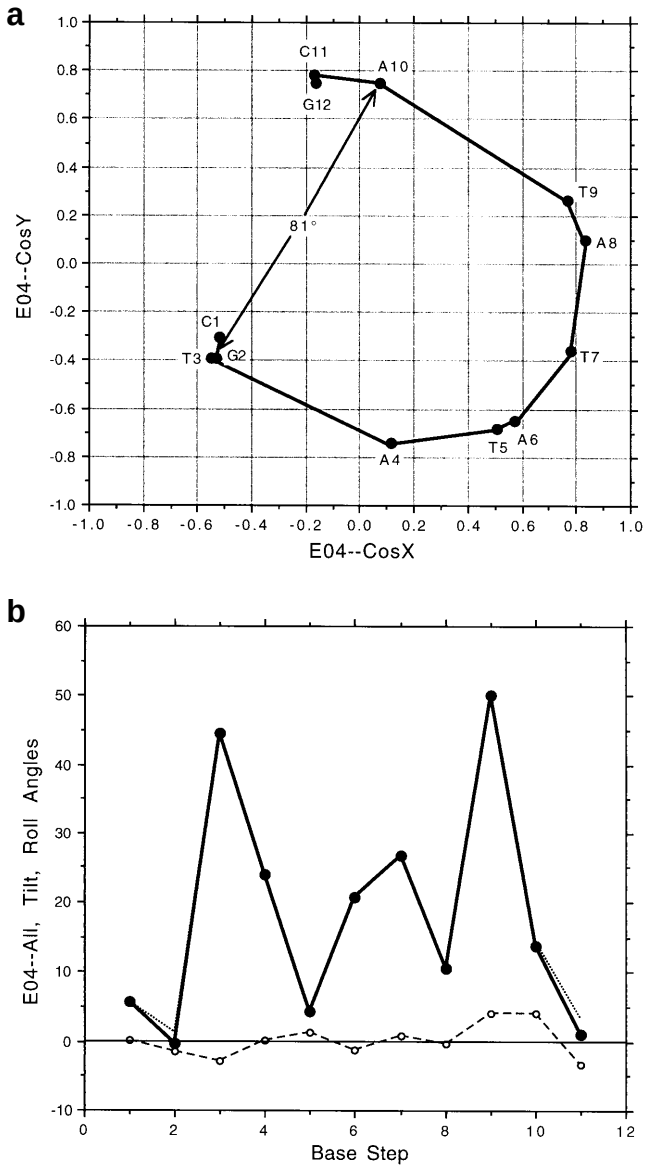


Figure 12. Major writhe in DNA bound to the human TATA-binding protein or TBP (10). (**a**) CosX/CosY normal vector plot for showing the broad half-circle writhe induced in the DNA duplex by binding to TBP. Major kinks occur at pyrimidine/purine steps: T–A. (**b**) VALL/VTIL/VROL plot for TBP, showing that the writhe arises from roll (heavy solid lines), with completely negligible contribution from tilt (dashed lines). The VALL curve (dotted lines) is effectively buried underneath the VROL curve except at the two ends of the helix.

and VROL at each step of the helix. The two kinks are seen to result from large positive roll at C–A = T–G steps.

Tilt plays little or no part in CAP bending, an observation that will prove to be valid for DNA in general. The theoretical calculations of Zhurkin *et al.* in 1979 (28) demonstrated the extraordinary energy cost entailed by a tilt wedge that lifts apart the stacked base pairs at one end. Subsequent analyses (29–33) showed a regular pattern of behavior: bending of the DNA duplex via roll, not tilt, with Y–R (pyrimidine–purine) steps favoring compression of the broad major groove, and R–Y steps (purine–pyrimidine) less strongly favoring compression of the minor groove.

The DNA bound to integration host factor (see fig. 2a of ref. 6), also exhibits two large kinks, again produced by local positive roll 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 succesive 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

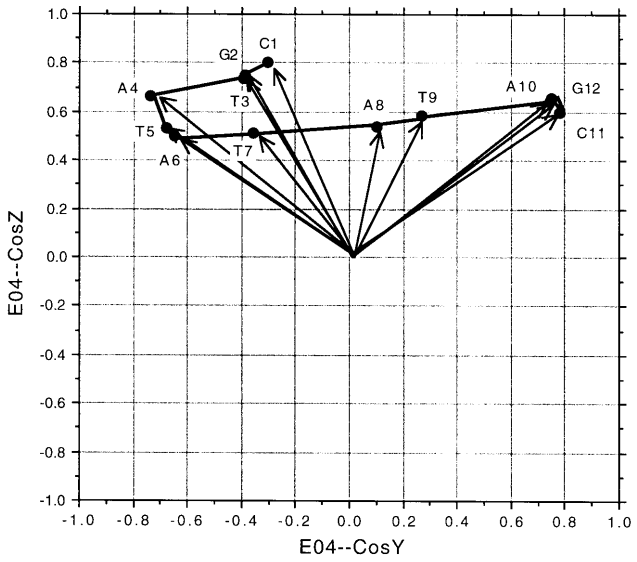


Figure 13. CosY/CosZ plot for TBP, showing the normal vectors of Figure 12a viewed from the right. The open half-circle of writhe now is seen to lie at constant latitude around the z viewing axis (which is vertical in this plot).

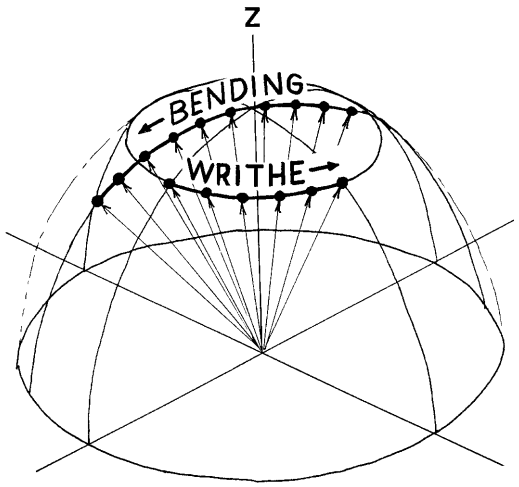


Figure 14. Illustration of the difference between a planar bend or curvature and writhe. In a plane curve, the tips of the normal vectors describe a great circle across the conformation hemisphere. In a writhe, the vector points sit on a constant-latitude circle around the z axis. For an idealized writhe, 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.

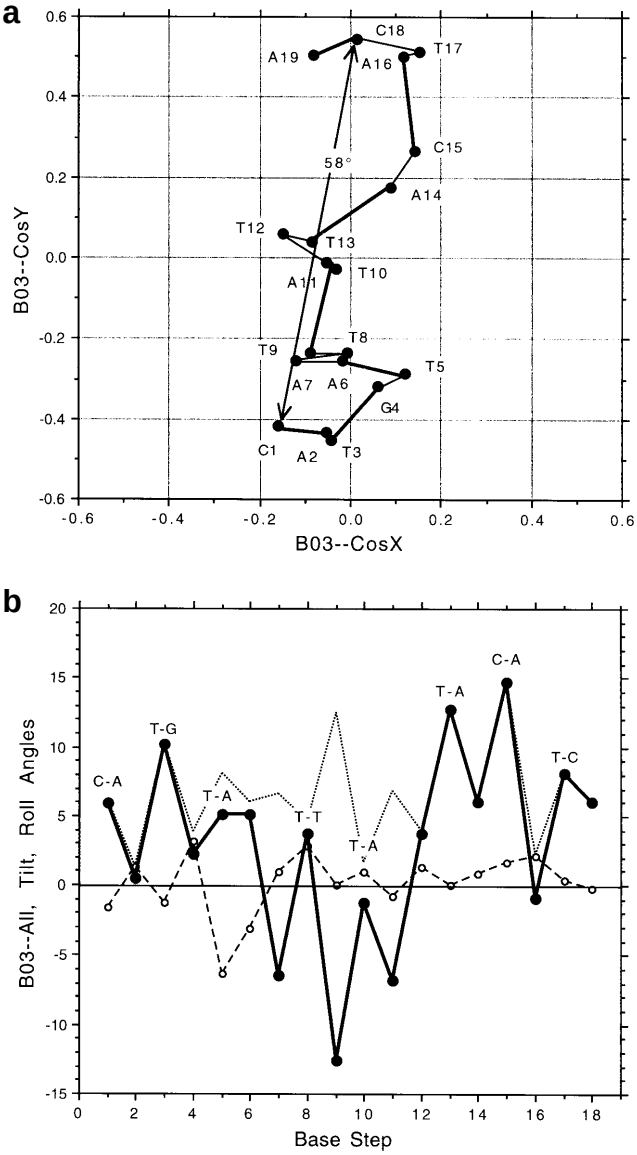


Figure 15. Effectively planar curvature in DNA bound to the MATa1/α2 homeodomain (36). (**a**) CosX/CosY normal vector plot, showing an overall bend of 58° that is produced by many small steps. Excursions to left and right of the plane of curvature arise from the reluctance of DNA to allow appreciable tilt angles between base pairs. Pyrimidine/purine steps are marked by heavier lines. (**b**) VALL/VTIL/VROL plot, showing negligible tilt contributions, and sinusoidal variation in the zig-zag plot of roll with a period of ~10 bp. Note that nearly all of the local roll maxima involve pyrimidine/purine steps: either T–A or C–A = T–G. Solid and dashed lines as in Figure 9.

globe, the normal vectors for T–A–T–A–T–A–T–A are seen to sweep around a circle of constant latitude relative to the z axis as the north pole. This is represented in idealized form in Figure 14.

If 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

	Helix–turn–helix	Zinc finger	Leucine zipper	Other	Total
Major bends					
Curved	2	–	–	2	4
Kinked	6	–	–	7	13
Writhed	–	–	–	10	10
Kinked and writhed	1	–	–	–	1
Total major	9	–	–	19	28
Minor bending					
Straight	1	1	1	4	7
Minor kink	9	1	–	–	10
Minor writhe	6	10	9	4	29
Minor curve	4	2	–	1	7
Minor, compound	1	2	–	2	5
Total minor	21	16	10	11	58
Total, all forms	30	16	10	30	86

Major curvature: MATa1/α2 homeodomain. Figure 1 of Li *et al*. (36) shows a MATa1/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 a1/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 a1 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 MATa1/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 a1-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 continous 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. Bringing 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

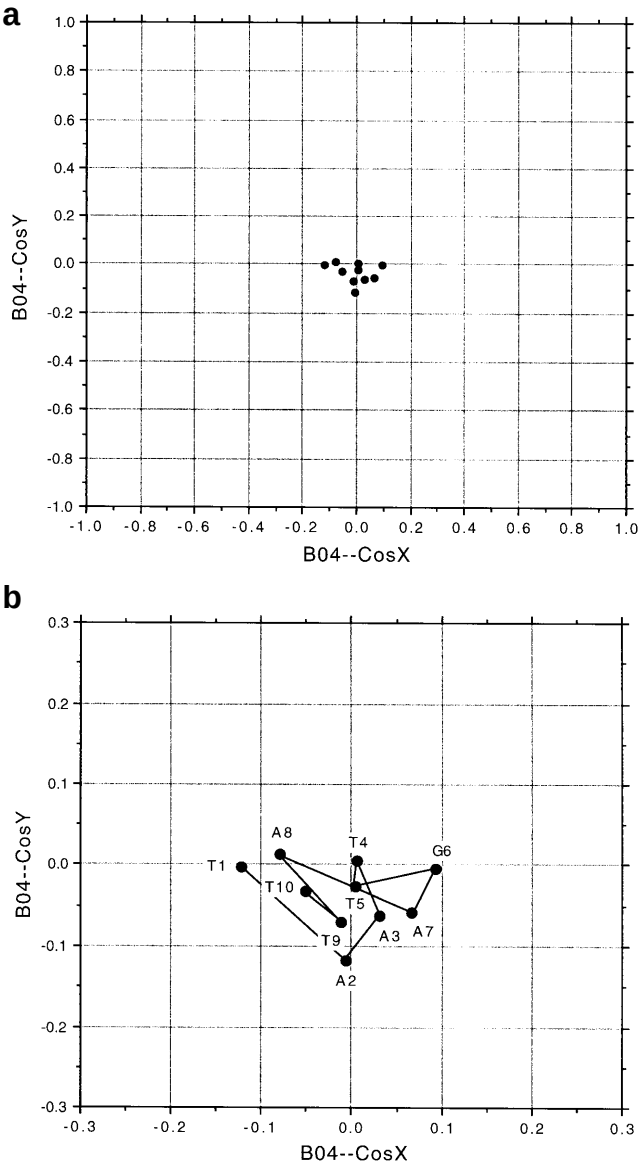


Figure 16. Normal vector plot for a straight helix, DNA bound to the even-skipped homeodomain (37). (**a**) Full plot, for comparison with Figures 4, 10a and 11a. (**b**) Enlargement of center, with limits ±0.3 in CosX and CosY. The tips of the normal vectors describe what is essentially a random walk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

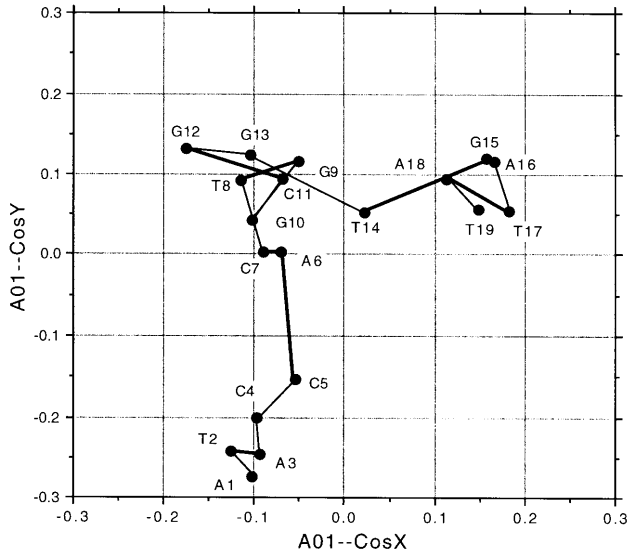


Figure 17. Minor kinks or curvature in the DNA bound to lambda repressor (11). Note that vector points are confined to the ±0.3 square in the center of the plot, as in Figure 16b. Two small non-coplanar bends of 16 and 19° result in an overall bending of 25° from one end of the helix to the other. Pyrimidine/purine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

plots for CAP, IHF and TBP (Figs 4, 10a and 12a) as steps along a conformational continuum.

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

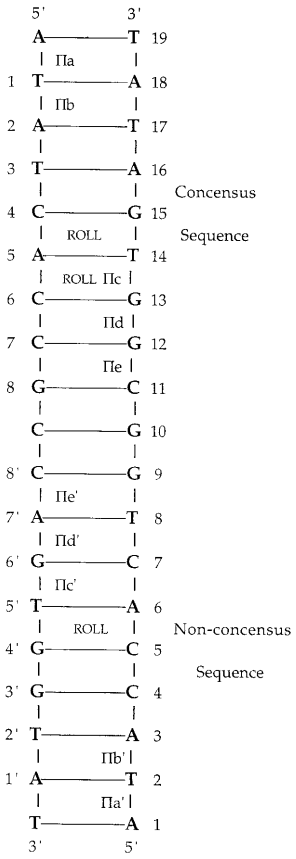


Figure 18. Schematic of the lambda operator duplex, as viewed into the unrolled major groove. Numbering on the left is that of Beamer and Pabo (11); that on the right is as used in this paper. The helix in this schematic is oriented like that in stereo figure 4 of reference 11. roll indicates positions of appreciable roll kinking, compressing the major groove. τa, τb, etc. mark phosphates that are protected from ethylation by hydrogen-bonding to the repressor, primed in the non-concensus sequence (below), and unprimed in the concensus seqence (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 betwen 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 preceeding 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°.

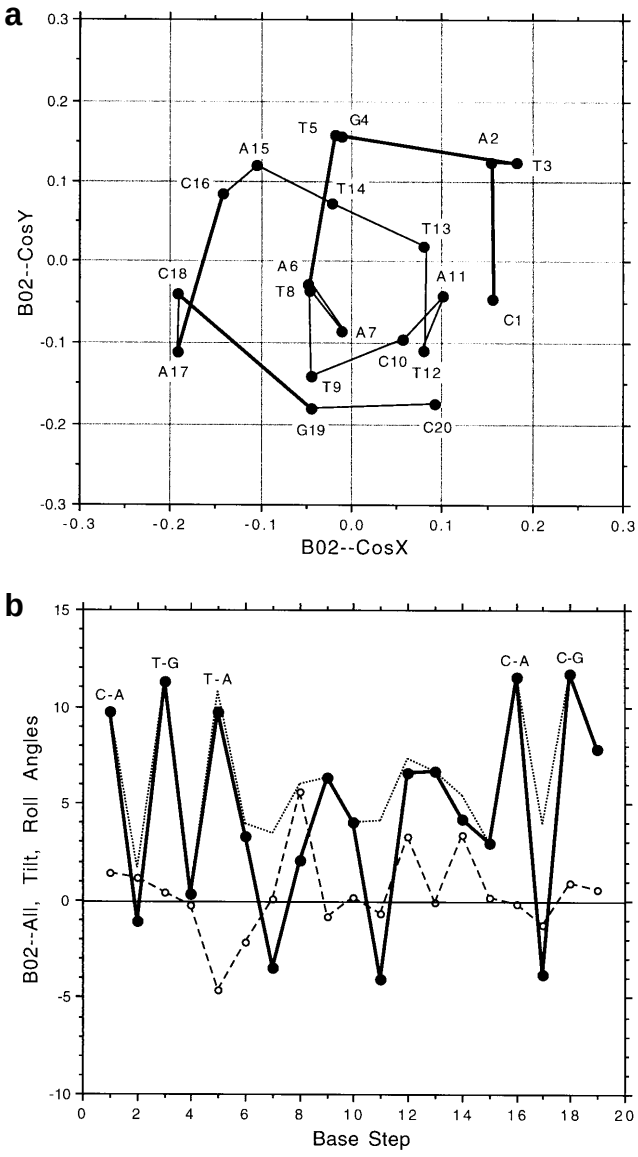


Figure 19. Minor writhe in DNA bound to the MATα2 homeodomain (38). (**a**) Normal vector plot. The two turns of helix (20 bp) describe two complete cycles of writhe, with a total spread of 16° or an amplitude of 8°. Pyrimidine/purine steps are marked with heavier lines. (**b**) VALL/VTIL/VROL plot. Note that at the two ends of the helix, tilt is negligible and maxima of positive roll occur only at pyrimidine/purine steps: T–A, C–G, and C–A = T–G. The center of the helix is more complex, with contributions from both roll and tilt. Solid and dashed lines as in Figure 9.

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

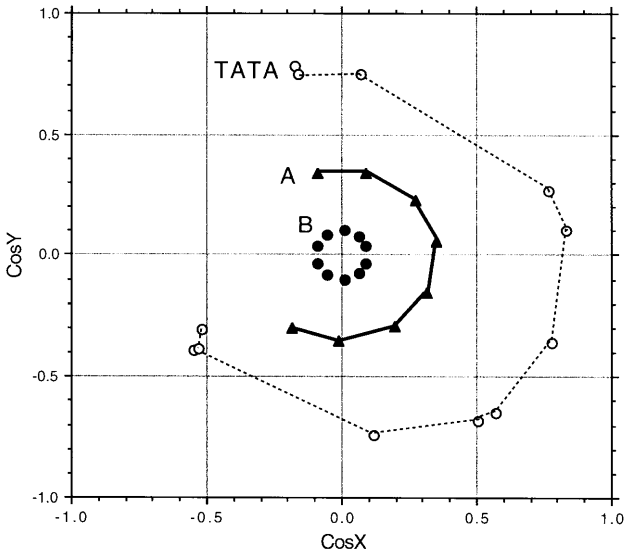


Figure 20. Idealized A- and B-DNA, analyzed with FREEHELIX, and compared with the human TATA-box bound to TBP. Non-zero roll angles in B-DNA contribute a minor writhe with amplitude of 5°. The 21° inclination of base pairs in this fiber-derived A-DNA helix, acted upon by helix rotation, results in a constant base pair roll angle VROL of 11°, and a writhe of amplitude 22°. 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 ~40–60°. (TATA box points from Fig. 12a.)

pushes up against the six phosphate groups lettered πc–πe and πc′–πe′ in Figure 18, protecting them from chemical ethylation. The two roll kinks at 5/6 and at 13/14/15 each have the effect of bringing two more phosphates from the other backbone chain into close contact with the repressor: πa and πb via kink 5/6, and πa′ and πb′ via kink 13/14/15. In the words of Beamer and Pabo: 'The bend is most obvious in our complex near the ends of the site where it allows repressor to make a number of contacts with the sugar–phosphate backbone, which could otherwise not occur.' If the operator duplex is viewed as seen by the approaching repressor (from the right in stereo Fig. 4 of ref. 11), then the top end of the operator bends down and to the right, to bring its phosphates πa and πb close to alpha helix 3 of one of the repressor monomers. The bottom end of the operator bends upward and to the left, to bring its phosphates πa′ and πb′ close to alpha helix 3 of the other repressor monomer in a similar manner. The role of the minor kinking visible in Figure 17 is to fine-tune the interactions between operator and repressor.

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 ~16°, 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.

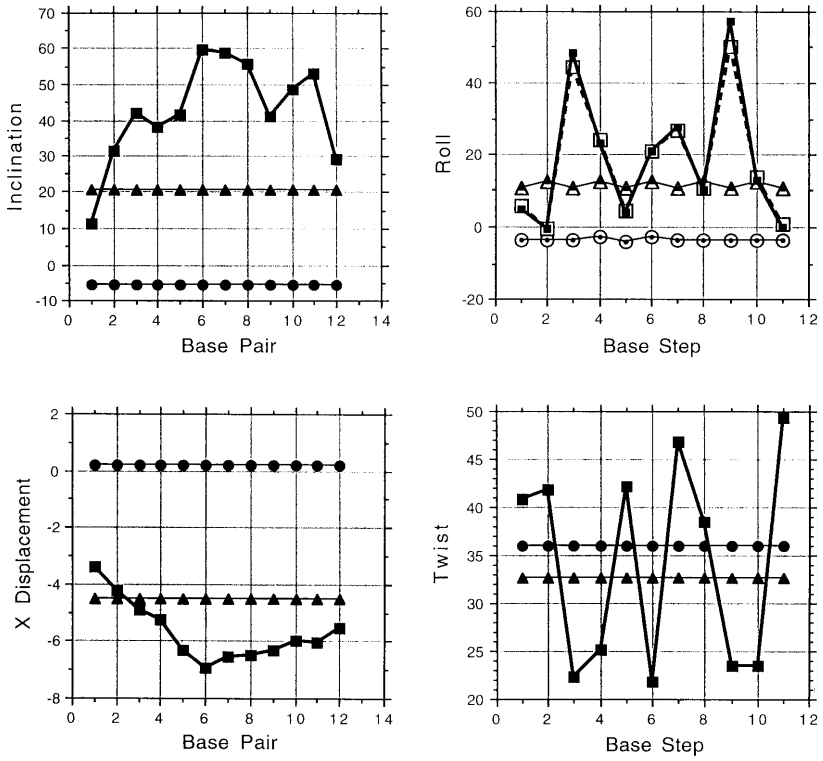


Figure 21. Comparison of (**a**) base pair inclination, (**b**) roll, (**c**) X displacement and (**d**) twist for ideal A-DNA (triangles), ideal B-DNA (circles) and the human TATA-box (squares). (b) Large open symbols indicate the vector quantity VROL, while small solid symbols indicate conventional Roll as calculated by NEWHELIX (and also by FREEHELIX).

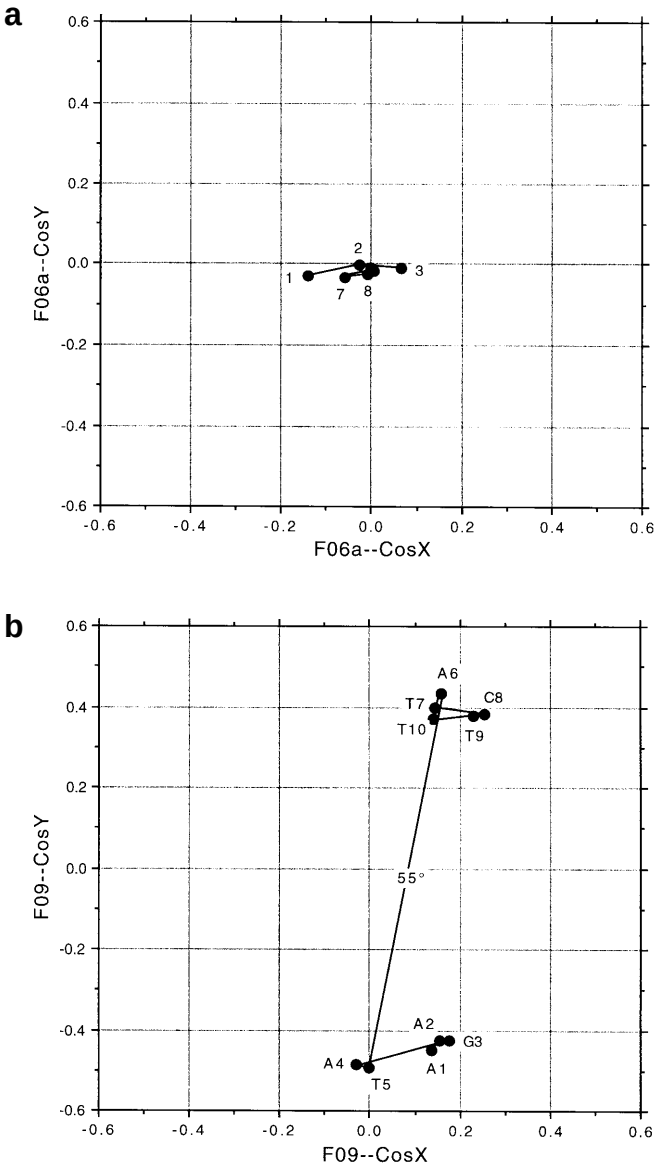


Figure 22. Illustration of the ability of the *Eco*RV restriction enzyme to discriminate among different DNA sequences. (**a**) The non-cognate sequence C–G–A–G–C–T–G–C binds to the enzyme as a straight and unbent helix. (**b**) When the cognate sequence A–A–G–A–T–A–T–C–T–T binds to *Eco*RV, the enzyme 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 past 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

Protein/DNA Complex	DNA Sequence	No. Steps Between Kinks	Dihedral Angle
γδ Resolvase [A15]	C-A-G-T-G-T-C-C-G-A-T-A-A-T-T-**T-A-T-A**-A-A-T-T-A-T-C-G-G-A-C-A-C-T-G 1 0-1 5-0 3 0 1 1-1 4 1-2 1 0 **9** 3 **7** 2 3-2 2 4-4 1 0-0 1 1 4 0 2 5	2	90
TBP [E04]	C-G-**T-A**-T-A-T-A-**T-A**-C-G 2-0 **9** 9 1 8 9 4 **9** 5 0	6	180
Trp Repressor [A12a]	G-T-A-C-**T-A**-G-T-T-A-A-C-**T-A**-G-T-A-C -0-0 0 1 **3** 2-0 1 2 0 0 2 **4** 1 0-0-0	8	286
Trp Repressor [A12b]	G-T-A-C-**T-A**-G-T-T-A-A-C-**T-A**-G-T-A-C -0 0-0 2 **4** 1-0 0 1-0 0 0 **5** 1-0-0 0	8	300
Lambda Repressor [A01]	A-T-A-C-**C-A**-C-T-G-G-C-G-**G-T-G**-A-T-A-T -0 0 0 1 **3** 0 1-1 2-1-2 1 **2** **3**-0-1 1-0	8.5	272
Lambda Repressor [A02]	A-T-A-C-**C-A**-C-T-G-G-C-G-**G-T-G**-A-T-A-T -1 2 0 0 **2** 1 1-0-0-0-1 1 **2** **2** 0-1 2-0	8.5	270
IHF [F11]	G-T-G-C-A-A-C-A-A-**A-T-T-G**-A-T-A-A-G-**C-A-A-T**-G-C-T-T-T-T-T-G-G-C 0-1 2-3-0-1 5 2 2 5 **9** 3-0-1-1-0 0-0-2 **9** 3-0 0 0-0 0-1-1 0 3 4 1	9	294
CAP [A08]	G-A-A-A-A-G-T-G-**T-G**-A-C-A-T-A-T-G-T-**C-A**-C-A-C-T-T-T-T-C-G 1-1-1 0-0 0-0-0 **9** 0-2 2 1-2 3 0-1 1 **9**-0 0 1-1-0 0-2 5-1	10	306
CAP [A09]	G-A-A-A-A-G-T-G-**T-G**-A-C-A-T-A-T-G-T-**C-A**-C-A-C-T-T-T-T-C-G 3 1-7 1-2 6-2 3 **9** 0-3-2 2-2-2 0 4 4 **9** 0-2-0-0 0-4-0 5 2	10	303
434 Cro [A07]	A-G-T-A-**C-A**-A-A-C-T-T-T-C-T-**T-G**-T-A-T 1 1-0 1 **3** 0 0-0-1 0 0-0-0-1 **6** 0-0-1	10	360

Structure analyses cited, in addition to those of Table 1, are: γ–δ resolvase (5) [pde0115 (A15)]; Trp repressor (51) [pdr009 (A12a,b)]; Lambda repressor (12) [pdr016 (A02)]; CAP (2) [pdr023 (A09)]; 434 Cro (52) [pdr001 (A07)].

Whether major or minor, bends are of three general classes.

(i) Localized *kinks*, produced by large positive roll at one or two discrete base pair steps.

(ii) Three-dimensional *writhe*, produced by positive roll at a series of adjacent steps.

(iii) Smooth *curvature*, produced by alternation of positive and negative roll every one-half helical turn, with side-to-side zig zagging of roll at intermediate points.

These bends are produced almost exclusively by rolling base pairs around their long axes. Tilt, or rotation about short axes, is insignificant. This agrees with past discussions of bending in DNA (28–35, 44–47), and also with common sense in terms of the energy required for the deformation: roll is easy; tilt is difficult. Most rolling is in a direction that compresses the large and relatively open major groove.

Roll bending occurs almost exclusively at pyrimidine–purine steps: C–A (= T–G), T–A or C–G. This is understandable in terms of the relatively small base pair overlap of pyrimidine–purine steps, by comparison with purine–pyrimidine or purine–purine steps. (See figs 4–6 of ref. 18, or fig. 2 of ref. 19).

Kinks and writhes are easy to achieve, and account for 23 and 39 examples of the 86 helices examined, respectively (Table 2). Smooth curvature is more difficult to bring about because of the absence of significant tilt, and is found in only 11 cases out of 86. Truly straight, unbent, unwrithed B-DNA with near-zero local roll is even rarer, occurring in only seven cases. The most common type of helix has a small but continuous writhe resulting from uniform local roll angles (similar to that of A-DNA but on a smaller scale), and an overall or global helix axis that need not deviate from linearity. But it would be stretching nomenclature to attempt to describe these as, in any sense, 'A-DNA'; they are simply minor variants of the B-form.

Bending in B-DNA is both sequence-dependent and facultative. For example, pyrimidine–purine sequences have a tendency toward positive roll bending, but in a given setting not every pyrimidine-purine step is necessarily bent. Examination of roll behavior in Table 1 shows many positions where a C–A step, or a T–G step, is not bent. The proper analogy with a pyrimidine–purine step is with a flexible hinge. If the necessary stress is applied, a hinge can bend; otherwise it can remain unbent. By contrast, an A-tract (defined as four or more successive As or Ts without a disruptive T–A step) is the quintessential rigid rod. A-tracts have been demonstrated to be stiff and unbent in DNA single crystals, in protein–DNA complexes, and are compatible with the solution data as well (18,19,27,48–50). Introduction of an A-tract in IHF target sequences that lack one, leads to tighter binding as the rigid A-tract packs against one side of the IHF

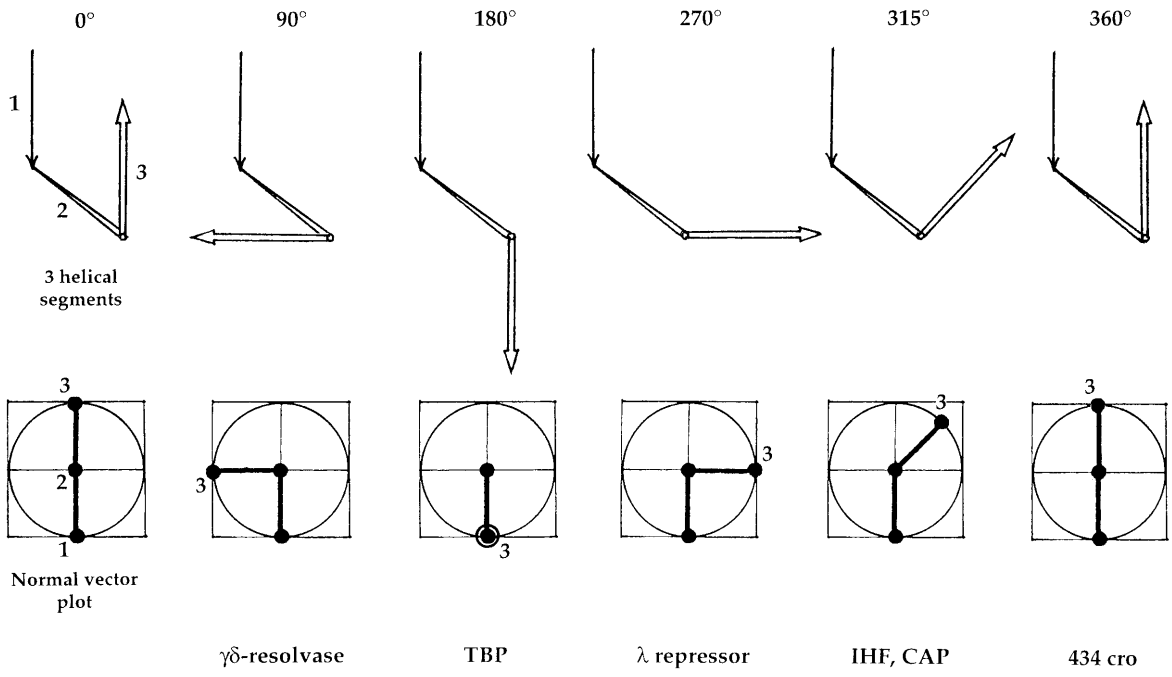


Figure 23. Relationship between the dihedral angles in a doubly-kinked DNA chain and the appearance of the normal vector plot. (Top row) Each of the internally straight DNA segments is represented by an arrow. Dihedral angles at top are angles between planes defined by segments 1 and 2, and 2 and 3. (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.

duplex (6). Conversely, improper positioning of an A-tract in the TATA box can damage binding by TBP (10).

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: *Eco*RV

A particularly elegant example of differential binding of a protein to target and non-target DNA is provided by the *Eco*RV 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 unbonded 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

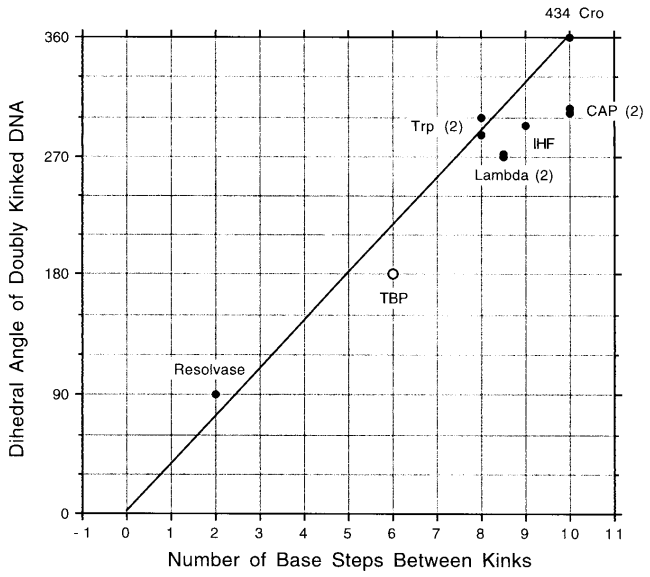


Figure 24. Plot of the dihedral angle induced by two successive roll kinks in DNA, versus the number of base pair steps between kinks, *n*, for sequences listed in Table 3. A spacing of 5*n*, where *n* is an even integer, leads to a C-shaped planar bend and a dihedral angle of 0° = 360°. A spacing of 5*n*, where *n* is an odd integer, produces an S-shaped planar bend, with dihedral angle 180°. All other spacings or values of *n* lead to non-planar, writhed segments. The TBP case is only approximate because the TATA box is writhed, not simply kinked at two places.

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

1 Schultz, S. C., Shields G. C. and Steitz T. A. (1991) *Science*, **253**, 1001–1007.
2 Parkinson, G., Gunasekera, A., Vojtechovsky, J., Zhang, X., Kunkel, T. A., Berman H. and Ebright R. H. (1996) *Nat. Struct. Biol.*, **3**, 837–841.
3 Lewis, M., Chang, G., Horton, N. C., Kercherl, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. and Lu, P. (1996) *Science*, **271**, 1247–1254.
4 Schumacher, M. A., Choi, K. Y., Zalkin H. and Brennan R. G. (1994) *Science*, **266**, 763–770.
5 Yang, W. and Steitz, T. A. (1995) *Cell*, **82**, 193–207.
6 Rice, P. A., Yang, S. -W., Mizuuchi, K. and Nash, H. A. (1996) *Cell*, **87**, 1295–1306.
7 Kim, Y., Geiger, J. H., Hahn, S. and Sigler, P. B. (1993) *Nature*, **365**, 512–520.
8 Kim J. L., Nikolov, D. B. and Burley, S. K. (1993) *Nature*, **365**, 520–527.
9 Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder R. G. and Burley S. K. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 4862–4867.
10 Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. and Dickerson, R. E. (1996) *J. Mol. Biol.*, **261**, 239–254.
11 Beamer, L. J. and Pabo, C. O. (1992) *J. Mol. Biol.*, **227**, 177–196.
12 Lim, W. A., Hodel, A., Sauer, R. T. and Richards, F. M. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 423–427.
13 Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne M. and Harrison S. C. (1988) *Science*, **242**, 899–907.
14 Shimon, L. J. W. and Harrison, S. C. (1993) *J. Mol. Biol.*, **232**, 826–838.
15 Rodgers, D. W. and Harrison, S. C. (1993) *Structure*, **1**, 227–240.
16 Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos K. and Wilson K. S. (1993) *EMBO J.*, **12**, 1781–1795.
17 Kostrewa, D. and Winkler, F. K. (1995) *Biochemistry*, **34**, 683–696.
18 Dickerson, R. E. (1998) In Neidle, S. (ed.), *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, Chapter 7.
19 Dickerson, R. E. (1998) In Sarma, R. H. and Sarma, M. H. (eds). *Structure, Motion, Interaction and Expression of Biological Macromolecules.* Proceedings of Tenth Conversation in Biomolecular Stereodynamics, Academic Press, Schenectady, NY, USA.
20 Lavery, R. and Sklenar, H. (1988) *J. Biomol. Struct. Dynam.*, **6**, 63–91.
21 Babcock, M. S., Pednault, E. P. D. and Olson, W. K. (1993) *J. Biomol. Struct. Dynam.*, **11**, 597–628.
22 Dickerson, R. E. and Drew, H. R. (1981) *J. Mol. Biol.*, **149**, 761–786.
23 Privé, G. G., Yanagi, K. and Dickerson, R. E. (1991) *J. Mol. Biol.*, **217**, 177–199.
24 Yanagi, K., Privé, G. G. and Dickerson, R. E. (1991) *J. Mol. Biol.*, **217**, 201–214.
25 Calladine, C. R. and Drew, H. R. (1992, 1997) *Understanding DNA: The Molecule and How it Works.* Academic Press, San Diego, etc., pp. 49–53.
26 Travers, A. (1993) *DNA–Protein Interactions.* Chapman and Hall, London, pp. 8–12.
27 Dickerson, R. E., Goodsell, D. S. and Kopka, M. L. (1996) *J. Mol. Biol.*, **256**, 108–125.
28 Zhurkin, V. B., Lysov, Y. P. and Ivanov, V. I. (1979) *Nucleic Acids Res.*, **6**, 1081–1096.
29 Zhurkin, V. B., Lysov, Y. P., Florentiev, V. L. and Ivanov, V. I. (1982) *Nucleic Acids Res.*, **10**, 1811–1830.
30 Zhurkin, V. B. (1983) *FEBS Lett.*, **158**, 293–297.
31 Ulyanov, N. B. and Zhurkin, V. B. (1984) *J. Biomol. Struct. Dyn.*, **2**, 361–385.
32 Zhurkin, V. B. (1985) *J. Biomol. Struct. Dyn.*, **2**, 785–804.
33 Sarai, A., Mazur, J., Nussinov, R. and Jernigan, R. L. (1989) *Biochemistry*, **28**, 7842–7849.
34 Calladine, C. R. and Drew, H. R. (1986) *J. Mol. Biol.*, **192**, 907–918.
35 Calladine, C. R., Drew, H. R. and McCall, M. J. (1988) *J. Mol. Biol.*, **201**, 127–137.
36 Li, T., Stark, M. R., Johnson, A. D. and Wolberger C. (1995) *Science*, **270**, 262–269.
37 Hirsch, J. A. and Aggarwal, A. K. (1995) *EMBO J.*, **14**, 6280–6291.
38 Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. and Pabo C. O. (1991) *Cell*, **67**, 517–528.

39 Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C. -S., Wang, A. H. -J. and Zhurkin, V. B. (1989) *EMBO J.*, **8**, 1–4.

40 Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C. -S., Wang, A. H. -J. and Zhurkin, V. B. (1989) *J. Biomol. Struct. Dynam.* **6**, 627–634.

41 Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C. -S., Wang, A. H. -J. and Zhurkin, V. B. (1989) *Nucleic Acids Res.* **17**, 1797–1803.

42 Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C. -S., Wang, A. H. -J. and Zhurkin, V. B. (1989) *J. Mol. Biol.* **206**, 787–791.

43 Guzikevich-Guerstein, G. and Shakked, Z. (1996) *Nature Struct. Biol.*, **3**, 32–37.

44 Young, M. A., Ravishanker, G., Beveridge, D. L. and Berman, H. M. (1995) *Biophys. J.*, **68**, 2454–2468.

45 Zhurkin, V. B., Gorin, A. A., Charakhchyan, A. A. and Ulyanov, N. B. (1990) In Beveridge, D. L. and Lavery, R. (eds), *Theoretical Biochemistry and Molecular Biophysics*. Adenine Press, Schenectady NY, pp. 411–432.

46 Maroun, R. C. and Olson, W. I. (1988) *Biopolymers*, **27**, 585–603.

47 El Hassan, M. A. and Calladine, C. R. (1997) *Phil. Trans. Roy. Soc. Lond.*, **A355**, 43–100.

48 Dickerson, R. E., Goodsell, D. S. and Neidle, S. A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3579–3583.

49 Goodsell, D. S. and Dickerson, R. E. (1994) *Nucleic Acids Res.*, **22**, 5497–5503.

50 Goodsell, D. S., Kaczor-Grzeskowiak, M. and Dickerson, R. E. (1994) *J. Mol. Biol.*, **239**, 79–96.

51 Otwinowski, Z., Schevitz, R. W., Zhang, R. -G., Lawson, C. L., Joachimiak, A. J., Marmorstein, R., Luisi, B. F. and Sigler, P. B. (1988) *Nature*, **335**, 321–329.

52 Mondragon, A. and Harrison, S. C. (1991) *J. Mol. Biol.*, **219**, 321–334.

APPENDIX. ANGLES BETWEEN ALL NORMAL VECTOR PAIRS

1. Catabolite Activator Protein

CAPS G-A-A-A-G-T-G-T-G-A-C-A-T-A-T-G-T-C-A-C-A-C-T-T-T-C-G A08__PDR006

J=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
I= 1	0	4	2	6	6	9	9	6	4	47	45	46	50	49	54	46	50	52	48	71	69	75	72	76	77	77	83	78	79
I= 2	4	0	5	9	9	12	12	10	8	49	47	48	51	50	56	48	52	54	50	75	73	79	76	79	81	80	86	82	83
I= 3	2	5	0	4	4	7	7	5	2	45	43	44	48	47	52	44	48	49	46	69	68	73	70	74	76	75	81	77	78
I= 4	6	9	4	0	0	2	3	3	1	41	39	40	44	43	48	40	44	45	42	65	64	69	66	70	72	71	77	73	74
I= 5	6	9	4	0	0	2	3	2	1	41	39	40	44	43	48	40	44	45	42	65	64	69	66	70	72	71	77	73	73
I= 6	9	12	7	2	2	0	1	4	4	38	36	37	41	40	45	37	41	43	39	62	61	66	63	67	69	68	74	70	71
I= 7	9	12	7	3	3	1	0	3	4	38	37	37	42	41	46	38	41	43	39	62	60	66	63	67	68	68	74	70	70
I= 8	6	10	5	3	2	4	3	0	3	41	40	41	45	44	49	41	45	46	42	65	63	69	66	69	71	71	76	72	73
I= 9	4	8	2	1	1	4	4	3	0	42	41	41	45	44	49	41	45	47	43	67	65	71	68	71	73	72	78	74	75
I= 10	47	49	45	41	41	38	38	41	42	0	7	0	7	10	10	4	3	5	4	31	30	36	31	37	39	37	40	42	41
I= 11	45	47	43	39	39	36	37	40	41	7	0	6	4	4	8	2	9	8	12	38	37	43	38	43	45	43	46	49	48
I= 12	46	48	44	40	40	37	37	41	41	0	6	0	7	10	10	4	4	5	5	32	31	37	32	37	39	37	41	43	42
I= 13	50	51	48	44	44	41	42	45	45	7	4	7	0	3	4	4	8	5	12	35	34	41	35	41	42	40	42	47	45
I= 14	49	50	47	43	43	40	41	44	44	10	4	10	3	0	6	6	11	9	14	38	38	44	38	44	46	43	45	50	48
I= 15	54	56	52	48	48	45	46	49	49	10	8	10	4	6	0	8	10	6	14	34	33	39	34	39	41	38	40	45	43
I= 16	46	48	44	40	40	37	38	41	41	4	2	4	4	6	8	0	7	6	9	35	34	40	35	41	42	40	43	46	45
I= 17	50	52	48	44	44	41	41	45	45	3	9	4	8	11	10	7	0	3	4	28	27	33	28	33	35	33	36	39	38
I= 18	52	54	49	45	45	43	43	46	47	5	8	5	5	9	6	6	3	0	7	29	29	35	30	35	37	35	37	41	39
I= 19	48	50	46	42	42	39	39	42	43	4	12	5	12	14	14	9	4	7	0	27	26	32	28	33	35	33	37	38	37
I= 20	71	75	69	65	65	62	62	65	67	31	38	32	35	38	34	35	28	29	27	0	1	5	1	5	7	6	12	11	9
I= 21	69	73	68	64	64	61	60	63	65	30	37	31	34	38	33	34	27	29	26	1	0	6	2	6	8	7	13	12	10
I= 22	75	79	73	69	69	66	66	69	71	36	43	37	41	44	39	40	33	35	32	5	6	0	5	0	2	3	10	6	4
I= 23	72	76	70	66	66	63	63	66	68	31	38	32	35	38	34	35	28	30	28	1	2	5	0	5	7	5	10	11	9
I= 24	76	79	74	70	70	67	67	69	71	37	43	37	41	44	39	41	33	35	33	5	6	0	5	0	1	2	9	6	4
I= 25	77	81	76	72	72	69	68	71	73	39	45	39	42	46	41	42	35	37	35	7	8	2	7	1	0	2	8	4	2
I= 26	77	80	75	71	71	68	68	71	72	37	43	37	40	43	38	40	33	35	33	6	7	3	5	2	2	0	7	7	4
I= 27	83	86	81	77	77	74	74	76	78	40	46	41	42	45	40	43	36	37	37	12	13	10	10	9	8	7	0	12	9
I= 28	78	82	77	73	73	70	70	72	74	42	49	43	47	50	45	46	39	41	38	11	12	6	11	6	4	7	12	0	3
I= 29	79	83	78	74	73	71	70	73	75	41	48	42	45	48	43	45	38	39	37	9	10	4	9	4	2	4	9	3	0

2. Integration Host Factor

IHF G-T-G-C-A-A-C-A-A-A-T-T-G-A-T-A-A-G-C-A-A-T-G-C-T-T-T-T-T-G-G-C F11__PDT040

J=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
I= 1	0	1	4	2	7	8	13	12	18	24	35	80	77	73	75	79	81	80	79	71	143	151	150	151	155	155	155	160	163	159	160	164	165
I= 2	1	0	4	4	5	6	11	12	18	24	34	79	76	71	74	78	80	78	78	69	141	149	149	150	154	154	154	159	161	158	159	164	165
I= 3	4	4	0	6	7	9	13	8	14	20	32	80	78	73	76	79	82	80	79	71	140	148	148	150	153	153	153	158	161	158	157	160	160
I= 4	2	4	6	0	9	11	15	14	20	26	38	83	80	75	78	82	84	82	82	73	145	153	153	154	158	158	158	163	165	162	163	166	166
I= 5	7	5	7	9	0	1	5	12	17	22	31	73	70	66	68	72	74	73	72	64	136	144	143	144	148	148	148	153	156	152	154	160	163
I= 6	8	6	9	11	1	0	4	12	17	22	30	72	69	64	67	71	73	71	71	62	135	143	142	143	147	147	147	152	154	151	153	160	163
I= 7	13	11	13	15	5	4	0	14	18	22	28	67	65	60	63	66	68	67	66	58	130	138	138	139	142	142	143	147	150	146	149	156	160
I= 8	12	12	8	14	12	12	14	0	5	11	23	76	76	71	73	77	79	78	77	69	133	141	141	143	147	146	146	150	154	151	149	151	152
I= 9	18	18	14	20	17	17	18	5	0	5	19	75	75	71	73	77	79	77	76	69	129	136	137	138	142	141	142	146	149	146	144	145	147
I= 10	24	24	20	26	22	22	22	11	5	0	14	73	74	70	72	75	77	75	75	68	123	130	131	133	137	136	136	140	144	141	138	140	141
I= 11	35	34	32	38	31	30	28	23	19	14	0	60	63	59	60	64	66	63	63	57	109	117	117	119	123	122	122	126	130	127	125	129	132
I= 12	80	79	80	83	73	72	67	76	75	73	60	0	11	12	9	8	9	7	8	13	65	73	71	71	75	75	75	80	82	79	83	93	98
I= 13	77	76	78	80	70	69	65	76	75	74	63	11	0	4	3	2	4	4	3	6	71	78	76	76	79	80	80	85	86	83	88	99	104
I= 14	73	71	73	75	66	64	60	71	71	70	59	12	4	0	3	6	8	7	6	2	75	83	80	80	83	84	85	89	90	87	92	103	108
I= 15	75	74	76	78	68	67	63	73	73	72	60	9	3	3	0	3	6	4	3	4	72	80	77	77	80	82	82	86	88	85	89	100	105
I= 16	79	78	79	82	72	71	66	77	77	75	64	8	2	6	3	0	2	1	0	8	69	76	74	74	77	78	78	83	84	81	86	97	101
I= 17	81	80	82	84	74	73	68	79	79	77	66	9	4	8	6	2	0	2	2	10	67	74	72	72	75	76	76	81	82	79	84	95	100
I= 18	80	78	80	82	73	71	67	78	77	75	63	7	4	7	4	1	2	0	1	9	68	75	73	73	76	77	77	82	83	80	85	96	100
I= 19	79	78	79	82	72	71	66	77	76	75	63	8	3	6	3	0	2	1	0	8	69	76	74	74	77	78	78	83	84	81	86	96	101
I= 20	71	69	71	73	64	62	58	69	69	68	57	13	6	2	4	8	10	9	8	0	77	84	82	82	85	86	86	91	92	89	94	105	110
I= 21	143	141	140	145	136	135	130	133	129	123	109	65	71	75	72	69	67	68	69	77	0	8	7	10	13	12	13	17	21	17	18	28	33
I= 22	151	149	148	153	144	143	138	141	136	130	117	73	78	83	80	76	74	75	76	84	8	0	3	5	8	5	6	9	13	10	20	25	
I= 23	150	149	148	153	143	142	138	141	137	131	117	71	76	80	77	74	72	73	74	82	7	3	0	2	6	4	5	9	13	9	11	22	27
I= 24	151	150	150	154	144	143	139	143	138	133	119	71	76	80	77	74	72	73	74	82	10	5	2	0	3	4	4	9	11	8	12	23	27
I= 25	155	154	153	158	148	147	142	147	142	137	123	75	79	83	80	77	75	76	77	85	13	8	6	3	0	2	2	5	7	4	10	21	25
I= 26	155	154	153	158	148	147	142	146	141	136	122	75	80	84	82	78	76	77	78	86	12	5	4	4	2	0	0	5	8	4	8	19	24
I= 27	155	154	153	158	148	147	143	146	142	136	122	75	80	85	82	78	76	77	78	86	13	6	5	4	2	0	0	4	7	4	8	19	24
I= 28	160	159	158	163	153	152	147	150	146	140	126	80	85	89	86	83	81	82	83	91	17	9	9	9	5	5	4	0	4	2	5	15	19
I= 29	163	161	161	165	156	154	150	154	149	144	130	82	86	90	88	84	82	83	84	92	21	13	13	11	7	8	7	4	0	3	8	16	20
I= 30	159	158	158	162	152	151	146	151	146	141	127	79	83	87	85	81	79	80	81	89	17	10	9	8	4	4	4	2	3	0	7	17	22
I= 31	160	159	157	163	154	153	149	149	144	138	125	83	88	92	89	86	84	85	86	94	18	10	11	12	10	8	8	5	8	7	0	11	15
I= 32	164	164	160	166	160	160	156	151	145	140	129	93	99	103	100	97	95	96	96	105	28	20	22	23	21	19	19	15	16	17	11	0	4
I= 33	165	165	160	166	163	163	160	152	147	141	132	98	104	108	105	101	100	100	101	110	33	25	27	27	25	24	24	19	20	22	15	4	0

3. Human TATA-Binding Protein

TBPJ C-G-T-A-T-A-T-A-T-A-C-G E04__PDT024

J=	1	2	3	4	5	6	7	8	9	10	11	12
I= 1	0	5	6	46	68	72	83	91	90	74	70	68
I= 2	5	0	1	43	66	70	83	94	94	80	76	73
I= 3	6	1	0	44	67	72	84	95	95	81	76	74
I= 4	46	43	44	0	23	28	45	67	73	96	101	98
I= 5	68	66	67	23	0	4	24	50	59	97	107	104
I= 6	72	70	72	28	4	0	20	47	56	96	107	104
I= 7	83	83	84	45	24	20	0	26	36	82	96	93
I= 8	91	94	95	67	50	47	26	0	10	60	74	73
I= 9	90	94	95	73	59	56	36	10	0	50	64	63
I= 10	74	80	81	96	97	96	82	60	50	0	14	13
I= 11	70	76	76	101	107	107	96	74	64	14	0	3
I= 12	68	73	74	98	104	104	93	73	63	13	3	0

4. MAT a1/a2 Homeodomain

MAT1 C-A-T-G-T-A-A-T-T-T-A-T-T-A-C-A-T-C-A B03__PDT028

J=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
I= 1	0	6	7	14	18	13	10	14	12	24	24	28	27	38	44	57	59	58	55
I= 2	6	0	1	9	13	11	11	12	12	24	25	29	28	36	42	56	58	59	56
I= 3	7	1	0	10	13	12	12	13	13	25	26	30	29	37	43	57	59	60	57
I= 4	14	9	10	0	3	5	11	6	9	17	18	24	22	28	34	48	49	51	49
I= 5	18	13	13	3	0	8	13	8	12	17	18	25	22	26	32	46	47	50	48
I= 6	13	11	12	5	8	0	6	1	4	13	14	19	17	25	31	45	46	47	45
I= 7	10	11	12	11	13	6	0	6	2	14	14	18	17	27	33	47	48	48	45
I= 8	14	12	13	6	8	1	6	0	4	12	13	18	16	24	30	44	45	46	44
I= 9	12	12	13	9	12	4	2	4	0	12	13	17	15	25	32	45	46	47	43
I= 10	24	24	25	17	17	13	14	12	12	0	1	8	4	13	19	32	34	34	32
I= 11	24	25	26	18	18	14	14	13	13	1	0	6	3	13	19	32	34	34	31
I= 12	28	29	30	24	25	19	18	18	17	8	6	0	3	15	20	30	32	31	27
I= 13	27	28	29	22	22	17	17	16	15	4	3	3	0	12	18	30	31	31	27
I= 14	38	36	37	28	26	25	27	24	25	13	13	15	12	0	6	20	21	23	22
I= 15	44	42	43	34	32	31	33	30	32	19	19	20	18	6	0	14	15	19	19
I= 16	57	56	57	48	46	45	47	44	45	32	32	30	30	20	14	0	2	6	11
I= 17	59	58	59	49	47	46	48	45	46	34	34	32	31	21	15	2	0	8	13
I= 18	58	59	60	51	50	47	48	46	47	34	34	31	31	23	19	6	8	0	6
I= 19	55	56	57	49	48	45	45	44	43	32	31	27	27	22	19	11	13	6	0

5. Even-skipped Homeodomain

EVEN T-A-A-T-T-G-A-A-T-T B04__PDT031

	J= 1	2	3	4	5	6	7	8	9	10
I= 1	0	9	9	7	7	12	11	2	7	4
I= 2	9	0	3	7	5	8	5	8	2	5
I= 3	9	3	0	4	2	4	2	7	2	5
I= 4	7	7	4	0	1	5	4	4	4	3
I= 5	7	5	2	1	0	5	4	5	2	3
I= 6	12	8	4	5	5	0	3	10	7	8
I= 7	11	5	2	4	4	3	0	9	4	6
I= 8	2	8	7	4	5	10	9	0	6	3
I= 9	7	2	2	4	2	7	4	6	0	3
I= 10	4	5	5	3	3	8	6	3	3	0

6. Lambda Repressor

LAMR A-T-A-C-C-A-C-T-G-G-C-G-G-T-G-A-T-A-T A01__PDR010

	J= 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
I= 1	0	2	1	4	7	16	16	21	22	18	21	23	23	20	27	27	25	24	24
I= 2	2	0	1	3	6	14	14	19	21	16	19	21	21	19	26	26	24	23	23
I= 3	1	1	0	2	5	14	14	19	21	16	19	22	21	18	25	25	23	23	22
I= 4	4	3	2	0	3	11	11	16	18	14	17	19	18	16	23	23	21	20	20
I= 5	7	6	5	3	0	9	9	14	15	11	14	17	16	12	19	20	18	17	16
I= 6	16	14	14	11	9	0	1	5	6	2	5	9	7	6	14	15	14	11	12
I= 7	16	14	14	11	9	1	0	5	6	2	5	8	6	7	15	16	15	12	14
I= 8	21	19	19	16	14	5	5	0	3	2	2	4	1	8	15	16	17	13	15
I= 9	22	21	21	18	15	6	6	3	0	5	1	7	3	5	12	12	13	9	11
I= 10	18	16	16	14	11	2	2	2	5	0	3	6	4	7	15	15	16	12	14
I= 11	21	19	19	17	14	5	5	2	1	3	0	6	2	5	13	13	14	10	12
I= 12	23	21	22	19	17	9	8	4	7	6	6	0	4	12	19	19	21	16	19
I= 13	23	21	21	18	16	7	6	1	3	4	2	4	0	8	15	15	16	12	15
I= 14	20	19	18	16	12	6	7	8	5	7	5	12	8	0	8	9	9	5	7
I= 15	27	26	25	23	19	14	15	15	12	15	13	19	15	8	0	0	3	2	3
I= 16	27	26	25	23	20	15	16	16	12	15	13	19	15	9	0	0	3	3	3
I= 17	25	24	23	21	18	14	15	17	13	16	14	21	16	9	3	3	0	4	1
I= 18	24	23	23	20	17	11	12	13	9	12	10	16	12	5	2	3	4	0	3
I= 19	24	23	22	20	16	12	14	15	11	14	12	19	15	7	3	3	1	3	0

7. Mat alpha-2 Homoeodomain

MAT2 C-A-T-G-T-A-A-T-T-C-A-T-T-T-A-C-A-C-G-C B02__PDT005

	J= 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I= 1	0	9	9	15	15	11	9	11	12	6	3	5	5	12	17	18	20	20	13	8
I= 2	9	0	1	9	10	14	15	14	19	13	10	14	7	10	14	17	24	22	20	17
I= 3	9	1	0	11	11	15	16	16	20	14	10	14	8	12	16	18	25	23	21	18
I= 4	15	9	11	0	0	10	13	11	17	15	13	16	9	4	5	8	18	15	19	20
I= 5	15	10	11	0	0	10	14	11	17	15	13	16	9	4	5	8	18	15	19	20
I= 6	11	14	15	10	10	0	3	0	6	7	8	8	7	5	9	8	9	8	8	11
I= 7	9	15	16	13	14	3	0	3	3	3	6	5	7	9	12	12	10	10	5	7
I= 8	11	14	16	11	11	0	3	0	5	6	8	8	7	6	9	8	9	8	8	11
I= 9	12	19	20	17	17	6	3	5	0	6	10	7	11	12	15	14	8	10	2	8
I= 10	6	13	14	15	15	7	3	6	6	0	4	1	6	10	15	15	14	14	7	5
I= 11	3	10	10	13	13	8	6	8	10	4	0	4	3	9	15	15	17	16	11	7
I= 12	5	14	14	16	16	8	5	8	7	1	4	0	7	11	16	16	15	16	8	3
I= 13	5	7	8	9	9	7	7	7	11	6	3	7	0	6	12	13	17	16	13	11
I= 14	12	10	12	4	4	5	9	6	12	10	9	11	6	0	5	6	14	11	14	15
I= 15	17	14	16	5	5	9	12	9	15	15	15	16	12	5	0	2	14	10	17	20
I= 16	18	17	18	8	8	8	12	8	14	15	15	16	13	6	2	0	11	7	16	20
I= 17	20	24	25	18	18	9	10	9	8	14	17	15	17	14	14	11	0	3	9	16
I= 18	20	22	23	15	15	8	10	8	10	14	16	16	16	11	10	7	3	0	11	18
I= 19	13	20	21	19	19	8	5	8	2	7	11	8	13	14	17	16	9	11	0	7
I= 20	8	17	18	20	20	11	7	11	8	5	7	3	11	15	20	20	16	18	7	0

8. Eco RV Restriction Enzyme, non-cognate sequence

EcoV C-G-A-G-C-T-C-G F06A_PDE002

	J= 1	2	3	4	5	6	7	8
I= 1	0	6	11	8	8	8	4	7
I= 2	6	0	5	1	2	1	2	1
I= 3	11	5	0	3	3	3	7	4
I= 4	8	1	3	0	0	0	3	0
I= 5	8	2	3	0	0	0	3	0
I= 6	8	1	3	0	0	0	3	0
I= 7	4	2	7	3	3	3	0	2
I= 8	7	1	4	0	0	0	2	0

9. Eco RV Restriction Enzyme, cognate sequence

EcoV A-A-G-A-T-A-T-C-T-T F09__PDE015

	J= 1	2	3	4	5	6	7	8	9	10
I= 1	0	1	2	9	8	52	50	49	49	48
I= 2	1	0	1	11	9	50	48	48	47	47
I= 3	2	1	0	12	11	50	48	47	47	46
I= 4	9	11	12	0	1	55	53	54	53	51
I= 5	8	9	11	1	0	55	53	54	53	51
I= 6	52	50	50	55	55	0	2	6	5	4
I= 7	50	48	48	53	53	2	0	6	5	1
I= 8	49	48	47	54	54	6	6	0	1	6
I= 9	49	47	47	53	53	5	5	1	0	5
I= 10	48	47	46	51	51	4	1	6	5	0