

An interactive modeling program for DNA

M. Treger and E. Westhof

Laboratoire de Cristallographie Biologique, IBMC-CNRS, F-67084 Strasbourg-Cedex, France

An interactive program for modeling A-, B- and Z-DNA in a schematic and nonatomic representation has been developed for the Evans and Sutherland PS300. The program lets users display several molecules for which parameters determining the three-dimensional structure can be calculated either on the basis of theoretical models or inferred from experimental data. The calculation of the curvature and torsion of the helical axis, by a method based on Frenet's equations, makes it possible to quantify the effects of the parameters on the helical axis.

Keywords: DNA, modeling, curvature, torsion

Received 26 August 1987
Accepted 16 September 1987

Most programs for modeling and visualizing molecules represent the molecules in an atomic way (for examples, see References 1 and 2). Such detail is not always necessary or useful, especially in the case of large molecules like DNA, which is several hundred base pairs long. In this case, a nonatomic, simplified representation is more suitable because we are less interested in the detailed relative positions of the atoms than in the overall features of the molecule.

In the case of DNA, the display of the helical axis is interesting, since experimental data obtained on natural or synthetic DNA molecules have established that DNA can be bent.³⁻⁶ Several hypotheses concerning the origin of DNA bending have been proposed, and all of them assume that the curvature and flexibility of DNA are determined, at least partly, by the sequence via steric interactions between adjacent residues. This led to theoretical models for calculating the parameters of the DNA three-dimensional (3D) structure in terms of sequence by Calladine and Dickerson,^{7,8} Tung and Harvey,^{9,10} Trifonov,⁶ and Calladine and Drew.¹¹ The last model allows you to calculate, for each dimer of base pairs, an interval of permissible values for the roll angle, together with a preferential value, and to anticipate the extent of the flexibility induced by a sequence.

Levene and Crothers¹² have written a program for the Evans and Sutherland vector processor PS300 that displays the sugar-phosphate chains of DNA molecules whose parameters have been calculated on the basis of various theoretical models. Tung and Harvey¹³ have also developed a program for DNA modeling that makes it possible to draw DNA molecules in a schematic representation on a Tektronix. We have developed an inter-

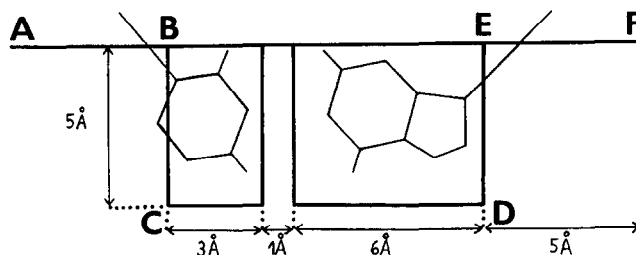


Figure 1. The schematic representation used for a base pair

active program that permits the schematic representation of DNA molecules in the right-handed A- and B-forms and in the left-handed Z-form. The helical parameters are obtained from three different theoretical approaches. The program also lets you display molecules whose parameters are inferred from experimental results (X-ray diffraction or NMR studies of oligomers), thereby making it possible to compare theoretical models and experimental data.

SCHEMATIC REPRESENTATION USED

Our representation is similar to that of Tung and Harvey:¹³ two rectangles for a base pair, as shown in Figure 1. The glycosyl base-sugar bonds are represented by segments. These bonds are not located on the great axis of a base pair, but are on the same side of it, in the minor groove. We have chosen to represent them by a segment extending on one side of each rectangle, so that it is situated on the side of the narrow groove. These bonds are blue, a color different from that of the bases. In this way, you can locate the narrow groove on the blue side and the large groove on the side of the bases. The sugar-phosphate chains are broken lines, and their vertices correspond to the points situated between the C(3') and C(4') atoms of the sugar rings. These points are at approximately the same level as the C(1') atoms bound to the bases. Color Plate 1 is the view obtained for DNA molecules in the standard A-, B- and Z-forms.

PARAMETERS DEFINING THE GEOMETRY OF THE DOUBLE HELIX

Defining the parameters

The parameters that determine the geometry of the double helix are as follows (see Figure 2):

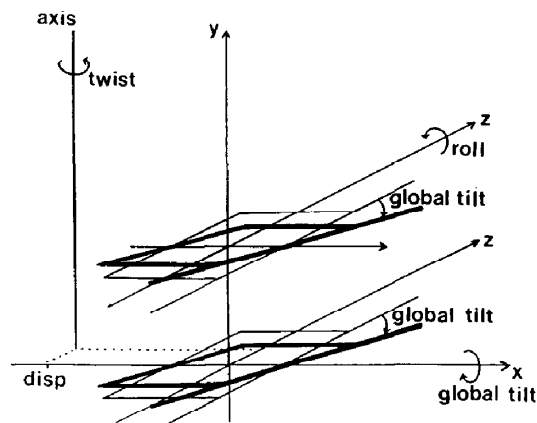


Figure 2. Definition of the helical parameters: global tilt = 0° in B-form, 13° in A-form; local tilt = 0° ; displacement = 0 \AA in B-form, -4.1 \AA in A-form. In this figure, the helix is in A-form

- The propeller-twist angle between bases of a hydrogen-bonded base pair. This is the angle between the plane of one base and the plane of the other base, the bases rotating in opposite directions about the great axis of the base pair. Conventionally, the propeller-twist angle is positive if, for an observer set on the great axis of a base pair, the nearest base rotates clockwise with regard to the farthest.
- The tilt angle of a base pair that defines the position of a base pair with regard to the helical axis. This angle is between the great axis of a base pair and a plane perpendicular to the helical axis.
- The twist angle between successive base pairs. This angle is between the great axis of a base pair without tilt and the orthogonal projection, on its mean plane, of the great axis of the following base pair.
- The roll angle between successive base pairs. This angle is between the mean plane containing one base pair without tilt and the mean plane containing the adjacent base pair without tilt.
- The distance between successive base pairs. This is the distance between the centers of gravity of the rectangles BCDE (Figure 1) corresponding to the two bases.

The translation parallel to the great axis of a base pair and the tilt angle between successive base pairs were neglected, since they do not vary in crystal structures of oligomers.¹⁴ The present set of parameters is sufficient to give, uniquely, the positions of all the base pairs in a DNA molecule.

Defining the axis and parameters that determine the axis path

In a helix represented in that way, the axis is a broken line where each vertex is associated with a base pair. In the case of the B-form of DNA, where the tilt angle is close to zero, the vertices of the broken line are the centers of gravity of the rectangles BCDE representing the base pairs, as shown in Figure 3. In an A-form helix, where the tilt angle is around 13° , a vertex of the axis associated with a base pair is located in the plane whose base pair would be situated in the absence of tilt, 4.1 \AA from the great axis of the base pair and

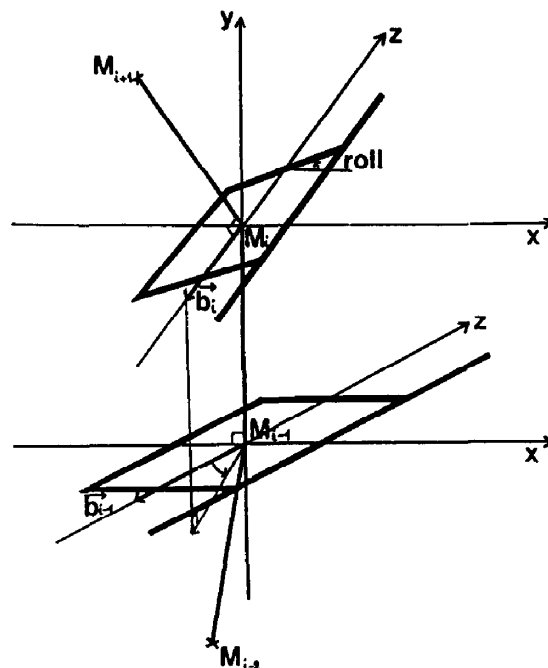


Figure 3. Definitions and notations used in the text. The helical axis is the broken line ($\dots, M_{i-2}, M_{i-1}, M_i, M_{i+1}, \dots$). In this figure, the helix is in B-form. \vec{t} : tangent unit vector; \vec{b} : third unit vector of local coordinate system; O : osculating plane; θ : torsion angle defined by points $M_{i-2}, M_{i-1}, M_i, M_{i+1}$

0.0088 \AA from the small axis on the side of the ascending strand $5'-3'$, as seen in Figure 2. The segment n of this broken line is perpendicular to the plane where the base pair n would be situated in the absence of tilt.

The axis path depends solely on the roll and twist angles and on the distance between base pairs; the tilt and propeller-twist angles do not modify the axis.

Calculating the parameters

The program incorporates three theoretical models for calculation of the parameters in terms of the sequence: those of Calladine and Dickerson,^{7,8} Tung and Harvey^{9,10} and Trifonov.⁶ These models let you calculate the roll and twist angles in terms of the base sequence.

The parameters for the double helix can also be inferred from experimental facts (X-ray diffraction or NMR studies of oligomers). And several molecules can be represented on the screen simultaneously and put to scale, translated or rotated with respect to each other interactively, allowing comparison between models and experimental results.

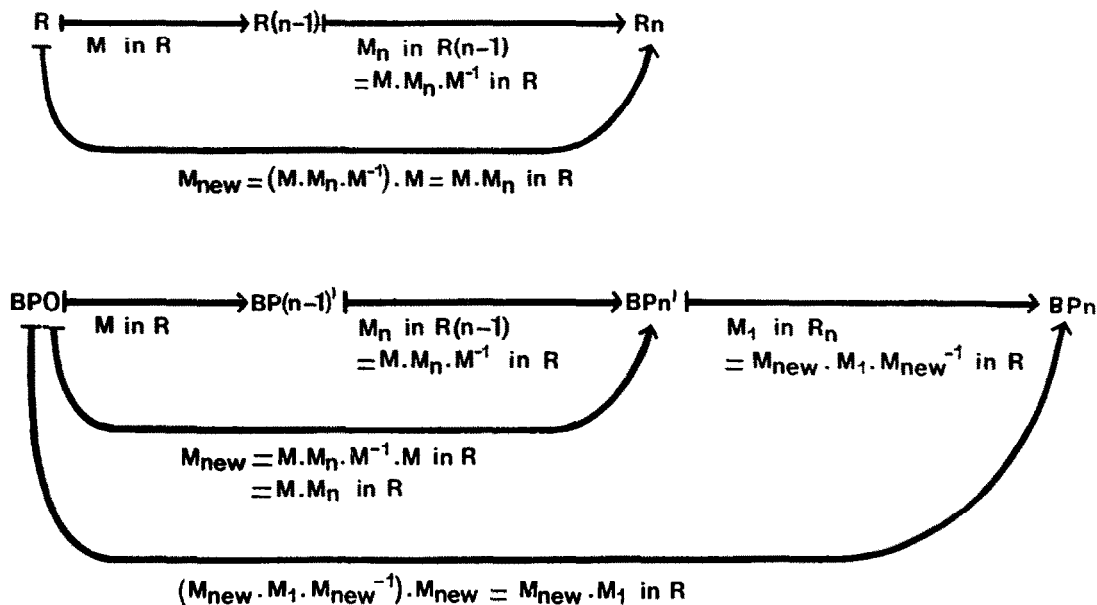


Figure 4. Definitions and notations used in the text. Method used for building the double helix: transformations for obtaining base pair n , BPn , from the first base pair, $BP1$, and $BP1$ from base pair $BP0$

BUILDING THE DOUBLE HELIX

The data required for constructing a double helix include the form of the helix (A or B or Z), the sequence of the bases, and the parameters for all base pairs.

In a double helix, the unit that is periodically repeated is a base pair. It is possible to compute, in the orthonormal left-handed coordinate system of the PS300, the coordinates of the points that define all the base pairs of the molecule, one after another from the first, in the following way (see Figure 4).

Calculating the coordinates of the first base pair

The first base pair, $BP1$, is obtained from another base pair, $BP0$, whose center of gravity coincides with the PS300 coordinate system R . $BP1$ is obtained from $BP0$ by a transformation taking into account the tilt and propeller-twist angles. It is the result of two transformations: a rotation of $BP0$ of the tilt angle about its small axis (Figure 2) and a rotation of the bases about an axis running through the points A and F (Figure 1), the two bases rotating in opposite directions of an angle equal to the half propeller-twist angle.

Calculating the coordinates of the base pair n

Let $BP(n-1)'$ be base pair $(n-1)$ of the molecule in the position that it would have in the absence of tilt and propeller-twist. Consider the coordinate system $R(n-1)$, local for $BP(n-1)'$, such that the plane xOz of $R(n-1)$ contains base pair $BP(n-1)'$.

At the beginning of stage n , the coordinates of all the base pairs from the first to $(n-1)$ have been computed in the PS300 coordinate system R and the matrix associated in R with the transformation $BP \rightarrow BP(n-1)'$ is known; it is noted M (Figure 4). It is also the matrix associated in R with the transformation $R \rightarrow R(n-1)$.

Let BPn' be base pair n in the molecule in the position

that it would have in the absence of tilt and propeller-twist. BPn' is obtained from $BP(n-1)'$ by a transformation composed of the following:

- A rotation of the roll angle about the great axis of $BP(n-1)'$, followed by
- A rotation of the twist angle about the segment of the helical axis local to $BP(n-1)'$, followed by
- A translation parallel to the segment of the helical axis local to $BP(n-1)'$ and of modulus equal to the value of the distance between $BP(n-1)$ and BPn (i.e., between the centers of gravity of the rectangles $B_{(n-1)}C_{(n-1)}D_{(n-1)}E_{(n-1)}$ and $B_nC_nD_nE_n$).

The matrix associated in the coordinate system $R(n-1)$ with this transformation is noted M_n . In the coordinate system R , the matrix of this transformation is $M.M_n.M^{-1}$. The same transformation produces the local frame of BPn' , Rn , from the local frame of $BP(n-1)'$, $R(n-1)$. The base pair n , BPn , is obtained from BPn' by a transformation taking into account the tilt and propeller-twist angles. It is the same transformation as $BP0 \rightarrow BP1$ but with different values for the angles; its matrix in Rn is M_1 . You can then compute the matrix associated in R with the transformation $BP0 \rightarrow BPn$ (Figure 4).

COLORATION OF ONE BASE TYPE IN A SELECTED MOLECULE

In most DNA molecules for which experimental data concerning curvature are available, the sequence of bases includes A series and T series, which are supposed to be responsible for the bending.^{3,4,15} To be able to locate the bases of one given type within a molecule displayed on the screen, especially with respect to the turns in the molecule, we have given the user the ability to color all the bases of one chosen type in a molecule by associating different hues with the four types of bases. Initially, the four types of bases have different hues, but their saturation is zero, so all bases appear white. The user chooses one type of bases with the help of a picking

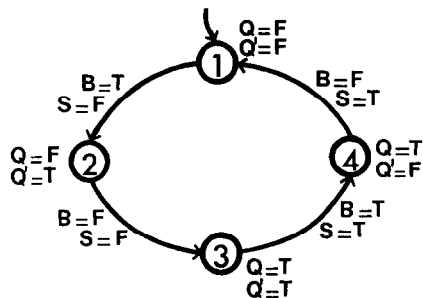


Figure 5. Definitions and notations used in the text. Transition graph of the automaton whose input is connected with the picking device for the base type and whose output is the saturation. The states encircled are the stable states. A state is stable if $Q_{t+1} = Q_t$ and $Q'_{t+1} = Q'_t$ when one does not modify the input B. Q_t is the value of Q at the time t; Q_{t+1} is the value of Q at the following time (t + 1). T and F stand for "true" and "false"

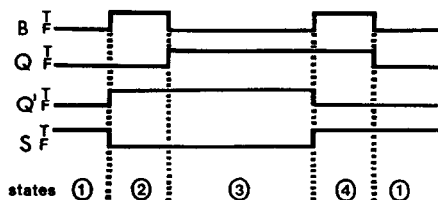


Figure 6. Definitions and notations used in the text. Chronogram of the variables B, Q, Q', S. T: true; F: false

device (a tablet with stylus) and a molecule with the help of a choice-indicating device (a programmed function keyboard). At each designation of one type of bases, the saturation of the bases of this type in the chosen molecule is modified. It takes alternately the values 0 and 1. The bases of this type in the molecule move alternately from the white state into the colored state.

Let's look at the idea behind the method. We have built a data-driven network of PS300 functions that lets us modify the hue of one or several types of bases in the selected molecule. This network is connected, on the one hand, with the choice-indicating device of the molecule and the picking device of the base type, and, on the other hand, with the entries corresponding to the saturation and the molecule number in the hierarchical description of the molecules. The part in the network connected with the picking device of the base type and whose output is the saturation is an automaton. The Huffman's method, which is used in sequential logic, has made it possible to build this automaton. It has one input (B of Boolean type) and one output (S: saturation of Boolean type). The variables B and S have the following values: B = true if a base is selected, false otherwise; S = true if the saturation is 0, false if it is 1.

This automaton has four states and two additional Boolean variables that depend on time (Q and Q'). Figure 5 shows the transition graph of the automaton. With the help of the transition tables of S, Q, Q', then the equations of S, Q, Q' by means of Karnaugh's method. From these equations you can deduce the chronogram (see Figure 6), the sequential network (see Figure 7) and then the data-driven network corresponding to this network.

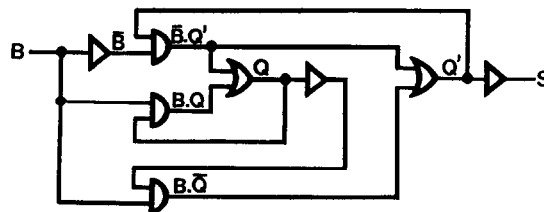


Figure 7. Definitions and notations used in the text. Sequential network whose input is B and output S. Logical operations: D : and D : or D : no

CURVATURE AND TORSION AT ANY VERTEX OF THE HELICAL AXIS

Frenet's equations make it possible to calculate curvature and torsion at any point of a regular and continuous curve in 3D space. Since the helical axis is a broken line, we have adapted these equations to the case of a broken line. If the vertex M_i is the point on the broken line of arc length s, we have chosen the next vertex, $M_{(i+1)}$, as the point of arc length $s + ds$ (i.e., ds is the length of the segment $(M_i, M_{(i+1)})$). Rackovsky and Scheraga¹⁶ have put forward a method in which all the segments of the broken line are assumed to have the same length, which they applied to the C_α chains of the polypeptides. We have used another method. At each vertex M_i on the helix axis, curvature ρ_i and torsion T_i are calculated in the following way:

- First vector of the Serret-Frenet's local frame:
The first vector of the local frame at the point M_i of a curve, τ_i , is the tangent unit vector at M_i . But at any vertex M_i within a broken line, a left tangent and a right tangent exist, τ_{ig} and τ_{id} (Figure 3) and they are different.

- Curvature at vertex M_i on the broken line:
By analogy with Frenet's formula:

$$\rho = |(\tau, \tau + d\tau)/ds|$$

where τ is the first vector of the Serret-Frenet's local frame and s is the arc length, we have defined ρ_i at M_i in the following way:

$$\rho_i = |(\tau_{ig}, \tau_{id})| / ((\|M_{i-1}M_i\|/2 + \|M_iM_{i+1}\|/2)) \quad (1)$$

$$= 2 \arccos(\langle \tau_{ig}, \tau_{id} \rangle) / (\|M_{i-1}M_i\| + \|M_iM_{i+1}\|)$$

$$\text{with } \tau_{ig} = M_{i-1}M_i / \|M_{i-1}M_i\| \text{ and } \tau_{id} = M_iM_{i+1} / \|M_iM_{i+1}\|$$

An equivalent formula has been proposed by Calladine and Drew.¹¹

- Torsion at vertex M_i on the broken line:
From Frenet's formula you can deduce the torsion at any point of a curve in 3D space:

$$T = (\mathbf{b}, \mathbf{b} + d\mathbf{b})/ds$$

where \mathbf{b} is the third vector of the local frame.

By analogy, torsion T_i at M_i can be written:

$$T_i = (\mathbf{b}_{i-1}, \mathbf{b}_i) / \|M_{i-1}M_i\|$$

Since the third vector of the local frame is perpendicular to the plane containing the first two vectors, \mathbf{b}_{i-1} and \mathbf{b}_i are defined by:

$$\mathbf{b}_{i-1} = (M_{i-1}M_{i-2} \wedge M_{i-1}M_i) / (\|M_{i-1}M_{i-2} \wedge M_{i-1}M_i\|)$$

if the points M_{i-2} , M_{i-1} , M_i are not aligned

and $\mathbf{b}_i = (\mathbf{M}_i\mathbf{M}_{i-1} \wedge \mathbf{M}_i\mathbf{M}_{i+1})/(\|\mathbf{M}_i\mathbf{M}_{i-1} \wedge \mathbf{M}_i\mathbf{M}_{i+1}\|)$
if the points M_{i-1} , M_i , M_{i+1} are not aligned.
Then torsion at M_i can be written:

$$|T_i| = \arccos(\langle \mathbf{b}_{i-1}, \mathbf{b}_i \rangle / \|\mathbf{M}_{i-1}\mathbf{M}_i\|) \quad (2)$$

But if the points M_{i-2} , M_{i-1} , M_i , M_{i+1} are situated in the same plane and are such that the vectors \mathbf{b}_{i-1} and \mathbf{b}_i are in opposite directions, torsion T_i must be zero, whereas according to equation 2 it has the value:

$$|T_i| = \pi / \|\mathbf{M}_{i-1}\mathbf{M}_i\|$$

Therefore, we have defined torsion T_i at point M_i in the following way:

$$|T_i| = \begin{cases} 0 & \text{if } (\mathbf{b}_{i-1}, \mathbf{b}_i) = \lambda\pi, \lambda \text{ positive or negative integer} \\ \arccos(\langle \mathbf{b}_{i-1}, \mathbf{b}_i \rangle / \|\mathbf{M}_{i-1}\mathbf{M}_i\|) & \text{otherwise} \end{cases}$$

Conventionally, for an infinitely small displacement along the curve, vector \mathbf{b} rotates about the right tangent in the negative direction if T is positive and in the positive direction if T is negative. One observes that $(\mathbf{b}_{i-1}, \mathbf{b}_i)$ is the torsion angle of the bonds $M_{i-2} - M_{i-1} - M_i - M_{i+1}$, except for the sign, and the sign of T_i is the same as that of the dot product $\langle \mathbf{b}_{i-1}, \mathbf{M}_i\mathbf{M}_{i+1} \rangle$.

In the method we used, there is no need to compute at each vertex of the broken line the second vector of the local frame. Besides, the equations giving ρ_i and T_i are still valid when the segment lengths are not equal, which is not the case in the method proposed by Rackovsky and Scheraga.¹⁶ Tung and Burks¹⁷ have put forward a much simpler method, not based on Frenet's equations. In their method, the bent path of the helical axis is approximated by the maximal distance from the helical axis to a segment running through the two ends of the axis. But this distance is not the radius of curvature, and when the axis is not plane, this distance is neither the radius of curvature nor the radius of torsion. Furthermore, two molecules for which the value of this distance is the same may have very different paths. In particular, they may be more or less compact and hence may display different electrophoretic behaviors, rendering the interpretation of their gel mobilities difficult.

The program computes the curvature and torsion at every vertex of the helical axis. If the curvature is zero at some vertices, only the vertices at which curvature is nonzero are considered together with the first and the last vertices, and the program computes the curvature and torsion at every vertex of this broken line by the same method.

Applying this method to particular cases of DNA molecules in which the twist angle has a constant value has made it possible to obtain the conditions in which the helical axis describes a plane curve and the conditions in which the helical axis forms a superhelix. This method can also be used to find relationships between, on the one hand, the axis parameters that determine the path of the axis (roll and twist angles, distance between base pairs) and, on the other hand, curvature and torsion of the axis.¹⁸

Color Plates 2 and 3 give examples of plane molecules, in B-form, in which the twist angle is constant and the roll angle is nonzero every n base pairs, displaying the effects of the roll angle on the path of the axis. Color Plate 4 shows a DNA molecule in B-form, in which

the parameters are calculated on the basis of Tung and Harvey's method, describing a left-handed superhelix.

CONCLUSIONS

An interactive program that allows the 3D schematic representation of DNA double helices in forms A, B and Z has been developed for the Evans and Sutherland vector processor PS300. The parameters determining the 3D structure of the helices can be either inferred from experimental data (X-ray diffraction or NMR studies of oligomers) or calculated on the basis of theoretical models by Calladine and Dickerson,^{7,8} Tung and Harvey^{9,10} and Trifonov.⁶ Several molecules can be represented simultaneously on the screen, put to scale and translated or rotated with respect to each other interactively. As a result, comparisons can be made between structures calculated on the basis of theoretical models and structures calculated on the basis of experimental data. The bases can be colored according to type so that you can localize them with respect to the bending regions. The molecules represented can be plotted on paper with a Versatec plotter directly from the screen, with stereographic view. To quantify the effects of the helical parameters on the axis path, the program incorporates the calculation of curvature and torsion at every vertex of the helical axis. The equations producing curvature and torsion are inferred from the Frenet's equations that have been adapted to the case of a broken line.

REFERENCES

- 1 Jones, T. A. Frodo: a graphics model building and refinement system for macromolecules. *J. Applied Crystallography*, 1978, **11**, 268-272
- 2 Chevrier, B., and Ripp, R. Program for the visualization and interactive study of molecules on a calligraphic display system. *J. Mol. Graphics*, 1986, **4**, 217-219
- 3 Koo, H.-S., Wu, H.-M., and Crothers, D. M. DNA bending at adenine-thymine tracts. *Nature*, 1986, **320**, 501-506
- 4 Hagerman, P. J. Sequence-directed curvature of DNA. *Nature*, 1986, **321**, 449-450
- 5 Lilley, D. DNA structure. Bent molecules, how and why? *Nature*, 1986, **320**, 487-488
- 6 Trifonov, E. N. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Research*, 1980, **8**, 4041-4053
- 7 Calladine, C. R. Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.*, 1982, **161**, 343-352
- 8 Dickerson, R. E. Base sequence and helix structure variation in B and A DNA. *J. Mol. Biol.*, 1983, **166**, 419-441
- 9 Tung, C.-S., and Harvey, S. C. A molecular mechanical model to predict the helix twist angles of B-DNA. *Nucleic Acids Research*, 1984, **12**, 3343-3356
- 10 Tung, C.-S., and Harvey, S. C. Base sequence, local helix structure, and macroscopic curvature of A-DNA and B-DNA. *J. Biol. Chem.*, 1986, **261**, 3700-3709

- 11 Calladine, C. R., and Drew, H. R. Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.*, 1986, **192**, 907-918
- 12 Levene, S. D., and Crothers, D. M. A computer graphics study of sequence-directed bending in DNA. *J. Biomolecular Structure and Dynamics*, 1983, **1**, 429-435
- 13 Tung, C.-S., and Harvey, S. C. Computer graphics program to reveal the dependence of the gross three-dimensional structure of the B-DNA double helix on primary structure. *Nucleic Acids Research*, 1986, **14**, 381-387
- 14 Dickerson, R.E., in *Biological Macromolecules and Assemblies* (Jurnak, F. A., and McPherson, A., eds.), Vol. 2, Wiley-Interscience, New York, 1985, pp. 471-494
- 15 Wu, H.-M., and Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature*, 1984, **308**, 509-513
- 16 Rackovsky, S., and Scheraga, H. A. Differential geometry and polymer conformation. 1. Comparison of protein conformations. *Macromolecules*, 1978, **11**, 1168-1174
- 17 Tung, C.-S., and Burks, C. A quantitative measure of DNA curvature enabling the comparison of predicted structures. *J. Biomolecular Structure and Dynamics*, 1987, **4**, 553-559
- 18 Treger, M., and Westhof, E. Evaluation of DNA curvature and torsion in *DNA Bending and Curvature* (Sarma, R. H., Sarma, M., Olson, W., and Sundaralingam, M., eds.), Adenine Press, New York, 1987