1472 Biophysical Journal Volume 68 April 1995 1472–1489

Comparison of Rotation Models for Describing DNA Conformations: Application to Static and Polymorphic Forms

J. Mazur* and R. L. Jernigan‡

*Frederick Biomedical Super Computing Laboratory, Program Resources, Inc./DYNCORP, National Cancer Institute/Frederick Cancer Research Foundation, and ‡Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health,
Bethesda, Maryland 20892 USA

ABSTRACT A new method, based on a space-fixed rotation axis, or local helix axis, is proposed for the calculation of the relative orientation variables for a sequence of base pairs. With this method, orientation variables are determined through the rotation of a base pair about this axis. These variables uniquely determine a set of helical variables, similar to the roll, tilt, and twist, commonly used for a description of spatial orientations of internally rigid base pairs. The proposed identification of roll and tilt with the direction cosines of the space-fixed rotation axis agrees well with their customary definitions as the openings of the angles between adjoining base pairs toward the minor groove and toward the ascending (5′ to 3′) backbone strand, respectively. These new variables permit a more direct physical comprehension of DNA conformations and also the behavior of self-complementary sequences. These direction cosines, together with the rotation angle about the space-fixed axis, form a set of three independent orientation variables of the bases that afford some advantages over the variously defined twist, roll, and tilt angles, either for static or average forms. An example for the static form of these variables is shown through their use to interpret crystal coordinates. An example for the average of orientation variables is based on statistical calculations. In this example, the orientation variables, together with the translational variables that describe the relative displacements of a pair of adjacent base pairs, form a canonically distributed ensemble in phase space spanned by these variables. Two sets of conformational variables are generated by using two different methods for performing rotation operations on the sequences of base pairs. The first method is based on the new single rotation about a space-fixed axis of rotation. This space-fixed axis of rotation is, in fact, the local helical axis as constructed previously by others. The second method is based on three consecutive rotations by Euler angles. Because of large flexibilities and anisotropies along various conformational variables of DNA base pairs, the two sets of generated conformational variables, based on these two different methods of performing rotation operations, lead to slightly different sets of structurally different, but energetically equivalent, spatial arrangements of the base pairs.

INTRODUCTION

A rotation matrix is used to determine the angular orientations of two neighboring base pairs. To extract rotation angles, this matrix must remain orthogonal when the direction cosines, that is, the nine elements of the rotation matrix, are replaced by their Boltzmann averages. In the canonical ensemble, there are only three angular rigid body variables. One problem is how to retain the orthogonality of the rotation matrix when the direction cosines are expressed in terms of the three Boltzmann-averaged variables. This problem will be detailed later in this paper. Here we will mention only the fact that the calculated conformational variables of the sequences of base pairs are not independent of the way the rotation matrix is constructed. Conformational flexibility, or fluctuations in translational and rotational parameters, depend on the rotation model employed. This is important in the interpretation of structures, whether from crystals, nuclear magnetic resonance, or conformational calculations.

In the conventional method (Sarai et al., 1988; Olson et al., 1988; Marky and Olson, 1994) of constructing the rotation matrix by performing three consecutive rotations about the Cartesian directions of the rotated base pair, the calculated Boltzmann averages of the direction cosines depend on the order in which these rotations are performed.

The 12 possible rotation sequences are designated by Goldstein (1981) as conventions. An example of using alternate rotation conventions is demonstrated by two such conventions used previously. One is the z convention (Goldstein, 1981). (The rotation convention is designated by the second rotation.) In this convention, there are three independent Euler angles, but the three plane rotations involve only two coordinates of the moving Cartesian axes. This convention was used by Sarai et al. (1988). The first and the last rotations were about the long (y) axis of the rotated base pair, and the second rotation was z, about the local helix axis. The other convention (designated by Goldstein as the zyx convention) is the one most frequently used (Olson et al., 1987; Marky and Olson, 1994; Soumpasis and Tung, 1988). In this convention, the last rotation is about the local helix axis. A few calculations based on the two conventions described above showed that the order of rotation contributes relatively little to the fluctuations and the variabilities of the conformational variables. However, the fact that the values of these variables, averaged in the canonical (Gibbsian) ensemble, can exhibit a dependence on the rotation model, is the main motivation for the current investigations.

The main impetus to the current investigation is the fact that other different rotation models, not restricted only to

---

Received for publication 31 March 1994 and in final form 4 January 1995.

Address reprint requests to Dr. Robert Jernigan, Laboratory of Mathematical Biology, DCBDC, NCI, NIH, MSC 5677, Rm. B-116, Bldg. 12B, 12 South Dr., Bethesda, MD 20892-5677. Tel.: 301-496-4783; Fax: 301-402-4724; E-mail: jernigan@lmmb.nci.nih.gov.

0006-3495/95/04/1472/18 $2.00

three sequential rotations and to one of the rotation conventions, could also be used for the conformational analysis. In the present paper, we are taking a different approach in which the specification of relative position of neighboring base pairs is reduced to a single rotation, but, instead of being about an axis of predefined orientation, it is about an axis for which the orientation must be determined to achieve the observed relative coordinate positions. This formulation has been well known in physics since the early 1800's. It was applied primarily to problems dealing with infinitesimal rotations as those are represented by single vectors. For finite rotations, two vectors are needed, one specifying the direction of the space-fixed axis and the other one for the position vector (Goldstein, 1981) of an atom in the rigid base pair with respect to some origin.

The main purpose of this paper is to develop this rotation model and compare it with the other rotation models. For these comparisons, we use first a rigid DNA crystal form with different models for performing rotation operations on the base pairs. Next, we perform statistical analysis with two different rotation models for calculations of the rotational variables to compare the two sets of different ensembles of the Boltzmann-averaged conformational variables. This comparison provides additional information about the fluctuations and conformational variabilities of base pair steps. The sequence dependence of fluctuations and conformational variabilities is manifested in the DNA flexibility and its polymorphism. These problems were the subject of previous publications (Sarai et al., 1988, 1989; Mazur et al., 1989). In these papers the environmental effects are also considered by assuming different forms for the dielectric solvent screening. It was suggested that conformational polymorphism of DNA is partially attributable to the intrinsic conformational preferences of some base pairs. A similar approach is adopted in the present work, except that the conformational analysis is performed by comparing the sets of Boltzmann averages of conformational variables obtained through different methods for performing rotation operations.

The following two different methods are selected for performing rotation operations, one based on three consecutive rotations, the other on a single rotation about a space-fixed axis. In the first method, three consecutive rotations, each characterized by a different Euler angle (the *zyx* convention), are employed. Notice that, in the *zyx* convention (the definition of the direction of the *x*, *y*, and *z* axes follows the Cambridge convention guidelines; see Dickerson et al., 1989), the first two rotations, by the *x* and *y* axes, can be interchanged, without significantly altering the average values of the rotation angles, as these angles (associated with tilt and roll, respectively) are small compared with the last rotation angle in this Cambridge convention (twist). In the new method presented here for determination of relative orientations of a set of two adjoining base pairs, the three consecutive rotations are replaced by a single rotation. This approach has been recently applied by Srinivasan et al. (1993) to characterize the three-dimensional structure of biopolymers. In the Srinivasan et al. (1993) paper, the spatial orientation of two neighboring rigid monomeric units is characterized by a single rotation axis. Srinivasan et al. apply their methodology to proteins and nucleic acids built of identical rigid monomeric units. An approach somewhat similar to the present one was first applied to nucleic acid structure analysis by Babcock et al. (1994). Raghunathan et al. (1993) also applied the single rotation model in their investigations on the molecular structure of a DNA triple helix.

In the single rotation model, the transformation matrix, which rotates a base pair to its new position in space, is unique in its structure, as its variables are the three direction cosines of the space-fixed axis and the rotation angle about this axis. As the space-fixed axis is represented by a unit vector, the three direction cosines that describe its orientation form an orthogonal set, thereby reducing the number of independent variables to three, the rotation angle and the two direction cosines. Alternatively, four orientation variables, that is, the three direction cosines of the rotation vector and the rotation angle, can be used to form linear combinations, resulting in four new variables, restricted by the orthogonality of the rotation matrix. A typical example of such variables is provided by the four symmetrical Euler parameters (Goldstein, 1981). This will be discussed in the following section.

The new rotation model, based on rotation about a single axis, offers an advantage over the model based on three consecutive rotations in providing better definitions of roll and tilt that are more consistent with their commonly used geometric models (Fratini et al., 1982). The concept that coordinate transformation between two rotating neighboring base pairs can be represented in terms of a single rotation axis and its spatial orientation is being applied to sequences of base pairs in the DNA. In this respect, the present work resembles that by Babcock et al. (1994). However, our approach and the one used by Babcock et al. (1994) are based on different physical modelings, that lead to closely related, but numerically slightly different, values for the computed orientational variables. The approach taken in this work will be presented in detail and compared with the physical model employed by Babcock et al. (1994).

In addition, complete description of spatial arrangements of adjoining base pairs requires specification of their relative translations. Bhattacharyya and Bansal (1988a), Bhattacharyya and Bansal (1988b), von Kitzing and Diekmann (1987), Lavery and Sklenar (1988), Lavery and Sklenar (1989), Soumpasis and Tung (1988), and Babcock et al. (1994) provide a self-consistent formulation that provides a complete description of the relative orientation of two adjoining base pairs in terms of three rotation and three translation parameters. The three translational variables, or degrees of freedom, are represented by two lateral displacements of a base pair along its long axis, connecting the C_6 and C_8 atoms (slide) and along its short axis, which points toward the minor groove (shift), and by rise, defined as the relative translation between the two sequential base pair centers (Bhattacharyya and Bansal, 1989).

A complete set of rotational and translational parameters for base pairs requires also the three rotational and the three translational degrees of freedom for the two bases linked together by hydrogen bonds. A transformation matrix is needed to orient two bases within a base pair. Mutual translations of two bases within a single base pair (denoted in the Cambridge convention as shear, stretch, and stagger) could also be considered. In the current work, these mutual translations of two bases within a single base pair are ignored.

The orientation variables defined by the Cambridge convention (Fratini et al., 1982; Olson et al., 1988) can be uniquely determined from the angles between the rotated axes of a second base pair and the axes of a first base pair in a pair of adjacent base pairs. The procedure was described, with small modifications, by Sarai et al. (1988). The coordinate system used there, but not here, to describe the relative orientation of two neighboring base pairs is as follows. The y axis connects atoms C_8 of purine and C_6 of pyrimidine in the base pairs and points toward the ascending 5′-3′ strand; the z axis is taken from the geometric center of the C_8-C_6 line and is normal to the base pair plane; and the x axis completes the right-handed orthogonal coordinate system, pointing toward the minor groove. Notice that the x axis does not have to lie in the plane of the base pair.

For the conventionally defined orientation variables, roll and tilt are defined as follows:

$$\text{Roll} = 90° - (x'z) \qquad \text{Tilt} = 90° - (y'z)$$

For clockwise rotation, twist is defined as:

$$\text{Twist} = 90° - (x'y)$$

This definition of twist differs from its common designation as $(y'y)$ (Olson et al., 1988; Sarai et al., 1988). In these equations, $'$ denotes the coordinates of the rotated second base pair with respect to the x, y, and z coordinates of the first, space-fixed base pair origin. In this procedure, $(x'z)$ denote the angle between the rotated x axis and unrotated z axis, $(y'z)$ denote the angle between the rotated y and unrotated z axis, and $(y'x)$ denote the angle between the rotated y and unrotated x axis.

The three rotational variables that uniquely determine the orientation of two adjoining base pairs, are defined by direction cosines between the rotated and unrotated sets of Cartesian axes. These definitions apply to any rotation convention. It is common knowledge that the rotational variables, defined in terms of Euler angles, are at variance with the Cambridge convention guidelines, as they are not invariant to the reversal of the order of rotations (Babcock et al., 1994; Sarai et al., 1988; Olson et al., 1988; Lavery and Sklenar, 1989).

Later in this paper, different definitions for tilt, roll, and twist will be presented, which satisfy the conditions imposed by the Cambridge convention that they be independent of the strand direction (5′ to 3′ versus 3′ to 5′) and the choice of reference strand. It will shown that, although three orientation variables cannot be expressed in terms of angles between pairs of rotated and unrotated axes, they can, nevertheless, be expressed in terms of linear combinations of two angles in which the rotated and unrotated axes are interchanged.

The emphasis in the current investigation is directed toward the evaluation of the direction cosines of the transformation matrix. The three orientation variables (tilt, roll, and twist) are calculated from the direction cosines for either one of the two rotation models described above. It should be remarked that the geometric coordinate transformation matrix method can be easily extended to sequences of a large number of consecutive base pairs. The rotational matrices, combined with the translational vectors, lead to generator matrices (see Flory, 1969; also, Maroun and Olson, 1988; Olson and Srinivasan, 1988), which are essential for structure analysis of chain molecules. It is of conceptual advantage to express parameters representing dimensional averages (such as end-to-end distances, radii of gyration, and helical axes) in terms of the averaged direction cosines in the various products of consecutive transformation matrices. Therefore, a good, even if approximate, way to evaluate direction cosines between adjoining base pairs from the available experimental data is essential for the evaluation of the orientation variables. Later in this paper we will describe an approach that first determines a set of three unit vectors associated with each base pair. This approach is applied to experimentally derived atomic coordinates of the bases. Direction cosines are then calculated by taking scalar products between sets of unit vectors for any two base pairs.

This paper is divided into several sections. First, the new method for the determination of space orientation of rigid base pairs is presented in detail. This method leads to a formulation of the rotation matrix in terms of the parameters of a rotation about a single axis of rotation. After this, the geometries of complementary and self-complementary sequences of base pairs are considered. In this section, a new definition of orientation variables is proposed that satisfies symmetry requirements that they be independent of the strand direction.

The remaining sections deal with computations. First, the two rotation models are compared for an experimental crystal structure. The method is described to calculate directly the direction cosines from the experimentally determined atomic coordinates. The last three sections deal with variabilities of base step conformations. These sections are oriented toward an examination of structural variability and the coexistence of polymorphic forms of DNA sequences. In particular, the structural variabilities of various sequences of base pairs are investigated. Typical calculations, performed on all 16 possible tetramers of the form MMNN, with both rotation models and two forms for dielectric solvent screening, are presented and the results discussed and analyzed.

Qualitative aspects of the current investigations should be emphasized. There might be other combinations of conformational variables and other methods of assembling them in phase space than the ones currently used. The parameters used to describe the force field and the partial atomic charges used currently are the same ones used in previous publications for a better comparison with our previously published data. The main purpose of the current investigation is to access the problem of structural variabilities by using an improved approach over those previously used. In the previous works, the emphasis was on fluctuations in conformational variables. In the present work, the emphasis is on the more technical aspect of using a different set of angular parameters in the canonical (or Gibbsian) ensembles of the conformational variables.

A NEW METHOD FOR DETERMINATION OF SPACE ORIENTATION OF SEQUENCES OF BASE PAIRS

We adopt a different coordinate frame that leads to the construction of a transformation matrix that does not require factorization of a total rotation of the base pair into the product of three axis rotations. It is based on the determination of a single rotation axis that remains invariant during a rotation. The problem is that, generally, the location of this axis cannot easily be found unless the eigenvector corresponding to the unit positive eigenvalue of a transformation matrix A (presented below) is evaluated. The search for this axis is facilitated by the fact that it should be located near the dyadic axis. The direction cosine of this axis with the z axis ought to be near 1. In other words it can be described as a local helix axis.

The description presented below for the rotation model about an arbitrary space-fixed axis is adapted from the entry on rotation and angular momentum in the Encyclopedia of Physics (van Dam et al., 1990)

Let a rotation $\{R\}$ be described by specifying its action on an arbitrary vector $\mathbf{r}$ (see Fig. 1). Each rotation $\{R(\phi, \mathbf{n})\}$ is characterized by a direction in space, $\mathbf{n}$, and a rotation angle ϕ. If rotation is about the unit vector $\mathbf{n}$ by an angle ϕ, and the rotation $\{R\}$ moves the original vector r to a new position r', then

$$\mathrm{r}' = \{R\}\mathrm{r}. \tag{1}$$

The relation between $\mathbf{r}'$ and $\mathbf{r}$ is

$$\mathbf{r}' = \{R\}\mathbf{r} = \mathbf{r}\cos\phi + \mathbf{n} \times \mathbf{r}\sin\phi + \mathbf{n}(\mathbf{n}\cdot\mathbf{r})(1-\cos\phi). \tag{2}$$

Equation 2 is sometimes referred to as the rotation formula (Goldstein, 1981). Derivation of Eq. 2 requires a considerable amount of vector analysis and is rarely presented in standard textbooks; however, the Textbook on Vectorial Mechanics (Milne, 1948) provides a detailed and complete derivation of this equation. Eq. 2 can be written in matrix form:

$$\mathbf{r}' = \mathbf{A}\mathbf{r}. \tag{3}$$

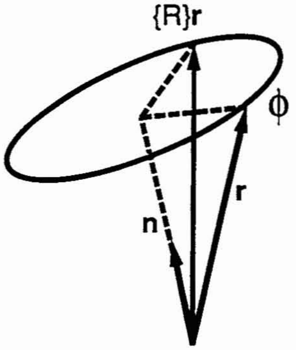

FIGURE 1 Geometrical interpretation of the action of a rotation $\{R(\phi, n)\}$ on an arbitrary vector $\mathbf{r}$ about a space-fixed axis n. (Adapted from van Dam et al., 1990).

A rotation vector that specifies both the direction of the axis and the amount of rotation is given by Eq. 4 below (see Milne, 1948; the vector $\mathbf{G}$ is also known as the Gibbs vector).

$$\mathbf{G} = (c_x\mathbf{u}_1 + c_y\mathbf{u}_2 + c_z\mathbf{u}_3)\tan(\phi/2) \tag{4}$$

The rotation vector $\mathbf{G}$ defines a finite rotation about the coordinate origin so that it brings any atom in a base pair to its new location after its rotation by the angle ϕ about the axis of rotation. In Eq. 4, $\mathbf{u}_i$, $i = 1, 2, 3$, are unit vectors along the three Cartesian directions x, y, and z, and c_x, c_y, and c_z are the direction cosines of the unit vector $\mathbf{n}$. These direction cosines form an orthonormal set, that is,

$$c_x^2 + c_y^2 + c_z^2 = 1. \tag{5}$$

Therefore, the rotation $\{R\}$ is characterized by the direction cosines of the vector $\mathbf{n}$ and by ϕ, the rotation angle. The sense, or the direction, of the rotation is determined by the sign of the rotation angle ϕ.

From Eq. 2 it follows that A, the transformation matrix, can be expressed as a sum of three matrices, one diagonal, one symmetric, and one skew-symmetric, each representing one of the terms in Eq. 2:

$$\begin{aligned} A = {} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\cos\phi \\ & + \begin{pmatrix} c_x^2 & c_xc_y & c_xc_z \\ c_xc_y & c_y^2 & c_yc_z \\ c_xc_z & c_yc_z & c_z^2 \end{pmatrix}(1-\cos\phi) \\ & + \begin{pmatrix} 0 & -c_z & c_y \\ c_z & 0 & -c_x \\ -c_y & c_x & 0 \end{pmatrix}\sin\phi \end{aligned} \tag{6}$$

Note that the matrix A describes positive rotations, defined by the right-hand rule, in accordance with the Cambridge convention rules.

Therefore, the transformation matrix is expressed in terms of three orientation variables. Two of these variables are the direction cosines of the space-fixed axis n, which form an orthogonal set, and the third independent variable is ϕ, the rotation angle about this axis. These three variables are being used instead of the three rotation angles in the rotation frame that was based on three consecutive planar rotations. These two sets of orientation variables, one based on three successive rotations, the other on the rotation about a space-fixed axis, are interconvertible. One set can be calculated from the other. In the following, reciprocal relations between these two sets will be derived. However, these reciprocal relations become only approximate if the respective orientation variables are replaced by their Boltzmann averages.

The equation for the transformation matrix is given below by Eq. 7. This equation is given in Jeffreys and Jeffreys (1946). The rotation angle ϕ is closely related to the usual twist angle. The small numerical difference between these two angles results from the change in the position of the space-fixed vector from that of the dyad axis, through its bending, which, in turn, is uniquely determined by roll and tilt angles. By the Cambridge convention, the sense of the rotation angle is determined so that the second base pair is rotated back toward the first, unrotated base pair. In Eq. 7, and in further developments, ϕ is taken to have the same sign as the twist angle.

$$A = \begin{pmatrix} \cos\phi + (1-\cos\phi)c_x^2 & (1-\cos\phi)c_xc_y - c_z\sin\phi & (1-\cos\phi)c_xc_z + c_y\sin\phi \\ (1-\cos\phi)c_xc_y + c_z\sin\phi & \cos\phi + (1-\cos\phi)c_y^2 & (1-\cos\phi)c_yc_z - c_x\sin\phi \\ (1-\cos\phi)c_xc_z - c_y\sin\phi & (1-\cos\phi)c_yc_z + c_x\sin\phi & \cos\phi + (1-\cos\phi)c_z^2 \end{pmatrix} \tag{7}$$

In Eq. 7, c_z can be replaced by

$$c_z = \pm(1.0 - c_x^2 - c_y^2)^{1/2}. \tag{8}$$

In current calculations, a plus sign is assigned in Eq. 8, as c_z measures the cosine of the angle between the positive rotation axis and the positive direction of the z axis. The matrix elements $A(i, j)$, $i, j = 1, 2, 3$ refer to the three Cartesian axes, x, y, and z.

The three components of the unit vector $\mathbf{n}$ and the two trigonometric functions of the rotation angle ϕ can serve as a basis for a set of four Euler parameters (Goldstein, 1981). These are $e_0 = \cos(\phi/2)$; $e_x = c_x\sin(\phi/2)$; $e_y = c_y\sin(\phi/2)$; and $e_z = c_z\sin(\phi/2)$. These four angular variables are related by the orthogonality condition expressed as

$$e_0^2 + e_x^2 + e_y^2 + e_z^2 = 1. \tag{9}$$

The transformation matrix A expressed in terms of these four Euler parameters is

$$A = \begin{pmatrix} e_0^2 + e_x^2 - e_y^2 - e_z^2 & 2(e_xe_y - e_0e_z) & 2(e_xe_z + e_0e_y) \\ 2(e_xe_y + e_0e_z) & e_0^2 - e_x^2 + e_y^2 - e_z^2 & 2(e_ye_z - e_0e_x) \\ 2(e_xe_z - e_0e_y) & 2(e_ye_z + e_0e_x) & e_0^2 - e_x^2 - e_y^2 + e_z^2 \end{pmatrix}. \tag{10}$$

There is a direct one-to-one correspondence between the angular variables of Eq. 7 and the angles that were used to rotate consecutively a base pair in each of the three planes of rotation and that are (only approximately) identified as twist, roll, and tilt. These variables are defined rigorously on the basis of the direction cosines of the transformation matrix A. This correspondence is shown below.

Let the nine direction cosines between the three Cartesian coordinates x', y', and z' of the rotated base pair or the base pair on the 3′ end of the base pair step and the Cartesian coordinates x, y, and z of the unrotated base pair be represented by their corresponding matrix elements. Thus, the direction cosine between rotated x' axis and unrotated y axis, which is given by the matrix element $A(1, 2)$ in the (x, y, z) frame of coordinates, is denoted as $a(x'y)$. Notice that the rotation angles in the transformation matrix orient the second base pair with respect to the first base pair.

Therefore, our old definitions for twist, roll, and tilt may be written as

$$\text{twist} \equiv \sin^{-1} a(x'y) \tag{11a}$$

$$\text{roll} \equiv \sin^{-1} a(x'z) \tag{11b}$$

$$\text{tilt} \equiv \sin^{-1} a(y'z). \tag{11c}$$

Bend angle is conveniently defined as $\cos^{-1} a(z'z)$.

From the orthogonality of the matrix A it follows that bend is given by

$$\text{bend} \approx (\text{roll}^2 + \text{tilt}^2)^{1/2}. \tag{12}$$

Equation 12 is readily obtained from the matrix A by identifying the matrix elements $a(x'z)$ and $a(y'z)$ with sin(roll) and sin(tilt), respectively, and expanding $\cos^{-1}(a(z', z))$ in a power series, retaining only quadratic terms in c_x and c_y. Bend can be expressed in terms of the new set of orientation variables. From Eqs. 7 and 8,

$$\begin{aligned} \text{bend} &= \cos^{-1}\{1 - (c_x^2 + c_y^2)(1 - \cos\phi)\} \\ &\approx 2\sin(\phi/2)(c_x^2 + c_y^2)^{1/2}. \end{aligned} \tag{12a}$$

Equation 12 follows from Eqs. 7, 11b, and 11c, as

$$\sin^2(\text{roll}) + \sin^2(\text{tilt}) = 2(c_x^2 + c_y^2)(1 - \cos\phi).$$

The direction cosines of matrix A and, therefore, roll, tilt, and bend are given in terms of c_x, c_y, and ϕ. Notice also that the rotation angle ϕ measures an amount of twist about the unit vector $\mathbf{n}$, which, essentially, is a bent dyad axis. In the absence of bend (both roll and tilt are 0), twist = ϕ, as expected. In the calculations described in this paper, values for twist, roll, and tilt are evaluated from the rotational variables, defined by Eq. 11.

A reverse procedure leading to calculations of the parameters that locate the space-fixed rotation axis and the rotation angle from the averaged rotation angles is less direct, as the transformation matrix is not uniquely determined when written in terms of these variables. Therefore, a general way to obtain the direction cosines c_x, c_y, and c_z and the rotation angle ϕ is to relate them to the direction cosines of the transformation matrix A rather than to the individual averaged rotation angles. First, the rotation angle is calculated from the trace of the matrix A as follows:

$$\cos\phi = 1/2[a(x'x) + a(y'y) + a(z'z) - 1]; \tag{13a}$$

then,

$$c_x = -[a(z'y) - a(y'z)]/2 \sin \phi \tag{13b}$$

$$c_y = [a(z'x) - a(x'z)]/2 \sin \phi \tag{13c}$$

$$c_z = [a(x'y) - a(y'x)]/2 \sin \phi . \tag{13d}$$

Equation 13, b-d, cannot be used when ϕ is 0 or 180°, as the rotation matrix is then symmetric.

Expression of new parameters in terms of old rotation parameters

We will apply these equations to the rotation model that approximates roll and tilt by the rotation angles $-\rho$ and τ. This model is based on the matrix of the following three right-handed rotations:

$$A = Z(\Omega)Y(\rho)X(\tau). \tag{14}$$

X, Y, and Z are single-rotation matrices that rotate a base pair along the x, y, or z axes, respectively.

The rotation angle ϕ, obtained from the trace of this matrix, is expressed as a function of Ω, ρ, and τ as

$$\cos \phi = 1/2[\cos \Omega(\cos \rho + \cos \tau) + \sin \Omega \sin \rho \sin \tau + \cos \rho \cos \tau - 1].$$

The approximation that replaces the direction cosines $a(x'z)$ and $a(y'z)$ by the rotation angles $-\rho$ and τ, respectively, requires that $\cos\rho$ and $\cos\tau$ be approximated as 1, and that $\sin\rho = \rho$ and $|\sin(\tau)| = \tau$ (ρ and τ are in radians). Then,

$$\begin{aligned} \phi &\approx \Omega \\ c_y &\approx 1/2[\tau + \rho \sin \Omega/(1 - \cos \Omega)] \\ c_x &\approx 1/2[-\rho + \tau \sin \Omega/(1 - \cos \Omega)]. \end{aligned} \tag{15}$$

Similar procedures apply to other rotation models. For example, when the rotation angles are the Euler angles α, β, and γ, the transformation matrix is the product of the following three rotations:

$$A = Y(\alpha)Z(\beta)Y(\gamma).$$

APPLICATIONS: ORIENTATION GEOMETRIES OF COMPLEMENTARY AND SELF-COMPLEMENTARY SEQUENCES OF BASE PAIRS

We now examine in more detail the new orientation variables and their conceptual relationships to the geometries of the base pairs. Consider first the conventional rotation model, based on three consecutive right-handed rotations. The transformation matrix A is defined as the product of three rotations as formulated in Eq. 14. If twist, roll, and tilt are defined by Eq. 11, a-c, then the same variables for the complementary sequences of base pairs are defined by Eq. 16, a-c, below:

$$\text{twist}' \equiv -\sin^{-1} a(y'x) \tag{16a}$$

$$\text{roll}' \equiv -\sin^{-1} a(z'x) \tag{16b}$$

$$\text{tilt}' \equiv \sin^{-1} a(z'y) \tag{16c}$$

(superscript ′ denotes the complementary base step).The symmetry considerations leading to the relations in Eq. 16, a-c, were detailed in a paper by Olson et al. (1987).

There are four base steps that are characterized by having identical base sequences along the ascending (5′-3′) and the descending (3′-5′) strands. Those are the CG, GC, AT, and TA steps. These steps form self-complementary sequences of base pairs. Because the two strands are identical, the direction cosines of these base steps must remain invariant to the reversal of their directions. For these self-complementary base steps the following relations hold:

$$\begin{aligned} a(y'z) &= a(z'y) \\ a(z'x) &= -a(x'z) \\ a(x'y) &= -a(y'x) \end{aligned} \tag{17}$$

Equation 17 follows from Eqs. 11, a-c, and 16, a-c. The following equations present the three rotation variables for sequences of neighboring base pairs and for their complements:

$$\begin{aligned} \sin(\text{roll}) &= -\sin \rho \\ \sin(\text{tilt}) &= \sin \tau \cos \rho \approx \tau \\ \sin(\text{twist}) &= \sin \Omega \cos \rho \approx \sin \Omega \\ \sin(\text{roll}') &= -\sin \tau \sin \Omega - \cos \Omega \sin \rho \cos \tau \\ &\approx -\rho \cos \Omega - \tau \sin \Omega \\ \sin(\text{tilt}') &= \sin \rho \sin \Omega \cos \tau - \sin \tau \cos \Omega \\ &\approx \rho \sin \Omega - \tau \cos \Omega \\ \sin(\text{twist}') &= \sin \Omega \cos \tau - \cos \Omega \sin \rho \sin \tau \approx \sin \Omega, \end{aligned} \tag{18}$$

where τ, ρ, and Ω are in radians. These equations, in slightly different forms, were derived by Olson et al. (1987).

Olson et al. (1987) correctly argue that the conventionally defined roll, tilt, and twist that describe complementary base steps differ in their values and, therefore, the geometries of, say, AA base steps differ from those of TT, as is evidenced by a casual inspection of Eq. 18. As pointed out by Olson et al. (1987), either rolls or tilts in complementary sets might have their signs in opposite senses to the ones imposed by the Cambridge convention. For example, rolls in complementary sets of base steps will have opposite signs if the ratio of roll to tilt in one of the base steps is less than tan(twist); usually ~0.73. Tilts in complementary sets of base steps have the same signs if the ratio of roll to tilt is less than −cot(twist); usually ~−1.45.

The differences between the geometries of the complementary base steps do not apply to bends. As a rule, any rotational variable represented by a diagonal term of the transition matrix is invariant in a set of two complementary base steps. Bend is defined as the direction cosine between the rotated and unrotated local dyad axes. Therefore, bend is invariant for all complementary sets of base steps. The invariancy of bends holds also for base pairs at longer separations with intervening identical complementary sequences, such as between base pairs one and four in the two tetramers AAAA and TTTT.

Because, for the self-complementary base steps, roll=roll′ and tilt=−tilt′, the following relation follows:

$$\sin\rho(1-\cos\Omega\cos\tau) = -\sin\Omega\sin\tau,$$

and, because $\sin\rho \approx \rho$, $\sin\tau \approx \tau$, $\cos\tau \approx 1$,

$$\sin\rho/\sin\tau \approx -\mathrm{roll/tilt} = \sin\Omega/(1-\cos\Omega), \qquad (19)$$

For the standard B-form twist Ω of ~36°, the roll/tilt ratio is roughly -3, a result confirmed by detailed calculations performed on the canonical ensembles of the structural variables.

Therefore, the self-complementary base steps (i.e., CG, GC, AT, and TA) have only two independent rotational degrees of freedom. However, the nonplanarity of the base pairs and chain-end effects in short polyoligomers render Eq. 19 to be only approximately valid.

Equation 19 can be derived from any rotation scheme used in the derivation of the transformation matrix A. Consider, for example, the standard Euler angles (α, β, and γ), defined in a paper by Sarai et al. (1988) (z convention; see Goldstein, 1981) and used therein for the determination of the spatial orientation in a base pair step. Symmetry considerations require that α and γ, the first and third rotation angles, be equal for all self-complementary base pair steps, as they involve rotations about the original and final y axes in the sequence of three consecutive rotations. Therefore, in this z convention, $a(y'z) = a(z'y)$, $a(x'z) = -a(z'x)$, leading directly to the roll/tilt ratio being a function of only the twist.

The above considerations resolve the problem of providing a consistent definition for roll and tilt. If tilt were defined as an angle that measures a compression, or an expansion, of a backbone strand, one would expect that tilt would be 0 in these four self-complementary pairs of base steps, regardless of the frame of coordinates. The proposed new definitions of the orientation variables of base pairs (see below), expressed in terms of the x, y, and z components of the rotation vector, obviates this difficulty, because, as is shown below, the symmetry-dictated requirement that $c_x=0$ for the self-complementary base pair steps requires a vanishing tilt. With the conventional definition of tilt, the self-complementary pairs of base pair steps have a nonvanishing tilt. However, the amount of this tilt is uniquely determined by the other two rotational variables, that is, by the roll and by the twist (Eq. 19). In this model, tilt can be 0 only in the absence of twist. Part of the difficulty is related to the often encountered inability to conceptually perceive problems that require three dimensions. For example, mechanical models (Ulanovsky and Trifonov, 1987) that were used as a basis for the definition of roll and tilt (Dickerson et al., 1989), are basically two-dimensional models. However, a stretching, or a contraction, of strands, which is used to define a tilt, can be depicted as a combined effect of all three rotational variables, not just as tilt. After a base pair is subjected to two rotations that define roll and twist, a fixed amount of a third rotation, identified as tilt, is needed to complete the three-dimensionality of the rotational space group and to preserve the internal symmetries of all four self-complementary base steps. In conclusion, tilt, uniquely determined from the direction cosines between the z direction of the rotated base pair and the x direction of the unrotated base pair, is non-zero, but its magnitude depends on the other two rotational variables.

We present here an alternate method that evaluates twist, roll, and tilt in the sets of complementary base pair steps. This method is based on the existence of the space-fixed rotation axis. The rotation matrix, written in terms of the rotation angle about this axis and its vector components was presented in Eq. 7. From this equation and from the unit vectors along the Cartesian directions in each one of the two complementary dimer base steps, the following relations between direction cosines of the space-fixed direction axis are obtained:

$$\begin{aligned}\sin(\mathrm{roll}') &= -(1-\cos\phi)c_xc_z - c_y\sin\phi\\ \sin(\mathrm{tilt}') &= (1-\cos\phi)c_yc_z - c_x\sin\phi\\ \sin(\mathrm{twist}') &= -(1-\cos\phi)c_xc_y + c_z\sin\phi\end{aligned} \qquad (20a)$$

For comparison, equations for the first base step in the complementary sets of the base steps are

$$\begin{aligned}\sin(\mathrm{roll}) &= (1-\cos\phi)c_xc_z - c_y\sin\phi\\ \sin(\mathrm{tilt}) &= (1-\cos\phi)c_yc_z + c_x\sin\phi\\ \sin(\mathrm{twist}) &= (1-\cos\phi)c_xc_y + c_z\sin\phi\end{aligned} \qquad (20b)$$

The orientation variables in the complementary sets of two base steps are presented below:

$$c'_x = -c_x \qquad c'_y = c_y \qquad c'_z = c_z \qquad \phi' = \phi. \qquad (21)$$

From Eq. 20, the following relations between the two sets of angular variables are obtained as

$$\begin{aligned}c_x &= -0.5[\sin(\mathrm{tilt}') - \sin(\mathrm{tilt})]/\sin\phi\\ c_y &= -0.5[\sin(\mathrm{roll}') + \sin(\mathrm{roll})]/\sin\phi\\ c_z &= 0.5[\sin(\mathrm{twist}') + \sin(\mathrm{twist})]/\sin\phi\\ \phi &\approx \Omega.\end{aligned} \qquad (22)$$

For the self-complementary base pair steps, $c_x = 0$, as required by Eq. 17.

Equation 22 is of great interest, as the angles between the rotation axis and the positive y and x axes are proportional, respectively, to the sums and differences of the two rolls and tilts of the two base steps that form a complementary set of sequences. Notice that c_y has the opposite sign of roll. Equation 22 provides a scenario for the relations between the orientation averages based on these two different coordinate inertia frames; c_x and c_y, which are the direction cosines of the space-fixed axis, are linear combinations of the two sines of the rotation angles in the interchanged sequence of rotations of the two base pair steps that form a mutually complementary set of base pairs. In other words, the coordinates of the space-fixed axis do not depend on the order of rotation operations in the complementary two-base-pair step.

It is suggested that the three direction cosines of the space-fixed rotation axis and the rotation angle be used to provide alternate definitions for the three rotational neighboring base pair parameters defined, in the Cambridge convention, as the tilt, roll, and twist. The orientation parameters that define the magnitude and orientation of the space-fixed rotation axis were shown (see Eqs. 21 and 22) to be independent of the direction of the double helix (5′ to 3′ or 3′ to 5′) and of the choice of primary versus secondary backbone strand (see also Babcock et al., 1994). Moreover, c_x and c_y describe the characteristics associated with the Cambridge convention requirements for tilt and roll, respectively, in terms of their action on strand expansion and contraction (for tilt) or on opening of base pair sequences towards the minor groove (for roll).

Based on the geometric arguments presented above, it is suggested to define three orientation variables for the adjacent base pairs in terms of the orientation variables of the space-fixed rotation axis, which replace the commonly defined tilt, roll, and twist. These proposed variables are designated in the present manuscript as tilt*, roll*, and twist*. If the orientation variables of the space-fixed rotation axis are its direction cosines c_x, c_y, and c_z and the rotation angle ϕ, then

$$\begin{aligned} \text{tilt}^* &= \sin^{-1}(c_x \sin\phi) \\ \text{roll}^* &= -\sin^{-1}(c_y \sin\phi) \qquad (23) \\ \text{twist}^* &= \sin^{-1}(c_z \sin\phi). \end{aligned}$$

It should be noticed that the above equations are invariant when ϕ is replaced by $(180 - \phi)$. This ambiguity does not present a problem for typical double-stranded structures, such as considered in the present paper, but could present a problem for structures such as hairpin loops.

The rotation angle ϕ is expressed in terms of the above defined roll, tilt, and twist as

$$\phi = \sin^{-1}[\sqrt{\sin^2(\text{tilt}^*) + \sin^2(\text{roll}^*) + \sin^2(\text{twist}^*)}] \qquad (24)$$

In the alternative rotation model, in which the orientation variables of the rotation axis are the four Euler parameters (see Eqs. 9 and 10), Eq. 23 is replaced by the following:

$$\begin{aligned} \text{tilt}^* &= \sin^{-1}(2e_x e_0) \\ \text{roll}^* &= -\sin^{-1}(2e_y e_0) \qquad (25) \\ \text{twist}^* &= \sin^{-1}(2e_z e_0), \end{aligned}$$

and tilt*, roll*, and twist* can be expressed in terms of the direction cosines of the transformation matrix as follows:

$$\text{tilt}^* = -\sin^{-1}([a(z'y) - a(y'z)]/2) \qquad (26a)$$

$$\text{roll}^* = -\sin^{-1}([a(z'x) - a(x'z)]/2) \qquad (26b)$$

$$\text{twist}^* = -\sin^{-1}([a(y'x) - a(x'y)]/2). \qquad (26c)$$

Simply stated, the sines of tilt, roll, and twist (Eq. 25) are the averages of these quantities that hold in the sets of complementary and self-complementary base pair steps and defined therein as rotations about their x, y, and z axes, respectively (eqs. 11).

In a series of recent publications (Babcock et al., 1993; Babcock et al., 1994; Babcock and Olson, 1994), the concept of coordinate transformation between two adjoining base pairs, based on a single rotation axis and its spatial orientation, was used, for the first time, to define rotational parameters that are in accordance with the guidelines set at a 1988 workshop to establish conventions. The proposed definitions for tilt, roll, and twist (Babcock et al., 1994) strongly resemble our definitions for these parameters, designated as tilt*, roll*, and twist*, and given by Eq. 23. However, in Babcock et al. (1994), c_x, c_y, and c_z, that is, the direction cosines of the vector of the axis of rotation, are represented as ratios of the angles tilt, roll, and twist, respectively, to the overall rotation angle ϕ. In our formulation (see Eq. 23), these angular rotation parameters are replaced by their sines. In the Babcock et al. (1994) formulation,

$$|\text{tilt}^{**}| = c_x\phi \quad |\text{roll}^{**}| = c_y\phi \quad \text{twist}^{**} = c_z\phi. \qquad (27)$$

Therefore, Eq. 24 is replaced by

$$\phi = \sqrt{(\text{tilt}^{**})^2 + (\text{roll}^{**})^2 + (\text{twist}^{**})^2}. \qquad (28)$$

Thus, their model is equivalent to a linearized version of ours in which sin x has been approximated by x. Numerical differences in the calculated absolute values of rotational variables according to our method and the one used by Babcock et al. (1994) are, therefore, rather small. Nonetheless, there are conceptual differences in the physical approaches presented above and those that underlie the methodology of Babcock et al. (1993, 1994) and Babcock and Olson (1994).

In the Babcock et al. (1994) methodology, the formulation of the transformation matrix is associated with their concept of simultaneous rotations about the x, y, and z axes being performed at constant angular velocities. The angular velocity of rotation is defined as an instantaneous rate of change in the rotation angle (Goldstein, 1981). The direction of this velocity lies along the axis of the infinitesimal rotation. This concept of simultaneous (infinitesimal) rotations is extended to the finite case by assuming a constant velocity. The cited linearizations of c_x, c_y, and c_z, the direction cosines of the axis of rotation, can be associated with these authors' extension of the model of infinitesimal rotations to finite rotations.

The method of infinitesimal (that is, commuting and, therefore, simultaneous) rotations toward the evaluation of molecular structure of polynucleotides was used by Sklenar et al. (1986). These authors developed a computer program, named SIR (for successive infinitesimal rotations), leading to a set of differential equations, which are solved by stepwise numerical integrations.

In the present work, the rotation parameters, regardless of the rotation model, are always expressed in terms of the direction cosines of the transformation matrix. In the conventional model, based on three consecutive rotations, there is a one-to-one correspondence between rotation parameters and the individual direction cosines of the transformation matrix. With the rotation model based on a single rotation by a space-fixed direction, the rotation parameters are given in terms of the differences in direction cosines of a transformation matrix and its transpose. It is therefore possible that the angular parameters measure amounts of rotation about a set of local reference axes that are intermediate between those of the two adjoining base pairs.

The following section shows how the direction cosines are determined from the sets of three perpendicular unit vectors associated with each base pair, which, in turn, can be calculated from data, whether experimental or model derived.

COMPARISON OF THE TWO ROTATION MODELS WITH FIXED CONFORMATION

It is interesting to compare directly the three orientation variables, based on the new and old rotation models, with these quantities determined from an experimental data set. With the old model, only three direction cosines are required to evaluate these variables (Eq. 11). With the new model, the entire transformation matrix is involved (Eq. 13, a-d). As all direction cosines cannot be readily assessed from the experimental data, we designed a simple approach to evaluate the two sets of three unit vectors assigned to each base pair. The nine direction cosines of the rotation matrix are then calculated by taking scalar products between unit vectors along the Cartesian coordinate system assigned to each base pair in a set of two, usually adjacent, base pairs. The first step in these calculations is the determination of the sets of three unit vectors associated with each base pair.

The procedure adopted is as follows. We first select a vector along the long axis of a base pair, with direction from the − strand to the + strand. Here this vector is taken in the direction of the axis connecting the C_6 atom of the pyrimidine base with the C_8 atom of the matched purine base. This vector is designated as the y vector. Other choices are also possible. One of them is described in the Cambridge convention as the line that passes through the C_6 atom and is parallel to the $C_{1'}$-$C_{1'}$ vector. Next, another vector, taken in the general direction that is transverse to the y vector (that is, it bisects this vector, and maintains the general direction toward the minor groove), is determined as follows.

First, another longitudinal axis is determined, replacing the C_6-C_8 axis, as this axis deviates considerably from a hypothetical longitudinal line that coincides with the long principal axis of the inertia tensor of a base pair. (Soumpasis and Tung (1988) defined the coordinate system of a base pair in terms of the three principal axes of inertia). For our purposes, a reasonable substitute for the long inertia axis is the axis that connects the geometric centers of the two bases that constitute a base pair. This hypothetical line separates the atoms into two sets, located on either side of this axis, in the base pair. The Cartesian coordinates of these centers are determined by taking averages over the coordinates in each of these two groups of atoms. The vector of the transverse direction is located in the direction of the line connecting these two geometric centers.

We list below the two groups of atoms located on either side of the hypothetical axis that connects the geometric centers of the two bases in a base pair. Hydrogen atoms are not included.

1. For d(C-G) base pair, group one is C: N_1, C_2, O_2, N_3; G: N_1, C_2, N_2, N_3, C_4, N_9; group two is C: C_6, C_5, C_4, N_4; G: O_6, C_6, C_5, N_7, C_8.
2. For d(A-T) base pair, group one is A: N_9, C_5, N_1, C_2, N_3, C_4; T: N_1, C_2, O_2, N_3, C_6; group two is A: C_8, N_7, C_6, O_6; T: O_4, C_4, C_5, C_{5M}.

The atoms in group one are located on the minor groove side and the atoms in group two are on the major groove side of base pairs. A nearly equivalent alternative to this transverse vector is the vector that connects the center of the transverse vector with the center of the longitudinal vector. This second alternative is the basis of the calculations presented here.

At a first glance, it seems to be rather unusual that two different longitudinal directions are used. One is for the determination of the longitudinal, or y vector. The other one is to determine the location of the longitudinal direction that separates the two sets of atoms with geometric centers connected by the transverse axis. The motivation for these procedures is the need to conform closely to the Cambridge convention for the general direction of the long axis. However, the C_6-C_8 axis does not provide a similar number of atoms located on either side of this axis. As a consequence, the base pair plane would differ in the amount of tip and inclination angles about its local helix axis (defined in the Cambridge convention) from these angles. This would affect the calculated values of tilt and roll between adjoining base pairs.

Alternatively, the vector that lies in the direction of the axis that connects the geometric centers of the two bases could have been selected as the y vector. However, the calculated twists based on this alternate vector differ considerably from the twists based on the C_6-C_8 vector. The large differences in the twist values can be qualitatively seen in results reported by Soumpasis and Tung (1988). These authors compare the Fratini et al. (1982) twist, roll, and tilt with these parameters calculated on the basis of the molecular model in which the x, y, and z axes are taken to be the three principal axes of inertia.

The longitudinal vector, selected here as the vector that connects the C_6 and C_8 atoms, defines the y vector of the base pair plane. The two vectors, the longitudinal and the transverse, define the base pair plane.

A vector perpendicular to the base plane is obtained by taking the cross product of the longitudinal and transverse vectors. This vector is designated as the z vector. Next, the orthogonal coordinate system is completed by taking a cross product between the longitudinal vector (the y vector) and the vector perpendicular to the base pair plane (the z vector). This vector lies along the short axis in the base pair plane of the long and transverse axes, pointing toward the minor groove. It is denoted as the x vector. The three unit vectors directed along the x, y, and z axes are obtained by dividing the x vector, y vector, and z vector by their lengths, respectively.

The direction cosines, which specify the relative orientation in space of any two base pairs, are obtained by taking dot products between the three vectors that determine the spatial orientations of the base pair planes. If these two base pairs are also adjacent along the chain, the calculated directions can be used to determine any three rotational variables, as defined for the rotation model used. Therefore, tilt, roll, and twist can be determined from the direction cosines according to Eq. 11, a-c. Alternatively, the parameters of the rotation about a single axis can be calculated from the direction cosines according to Eq. 13, a-d.

Actually, the two bases that constitute a base pair are not usually exactly coplanar because of the presence of propeller twist and buckle. The base pair plane, determined by the above described method simply ignores the intra-base-pair degrees of freedom such as propeller twist and buckle. The same approach outlined above can also be applied for the calculations of direction cosines between adjacent individual bases along each of the two backbone strands of the polynucleotide. As individual bases are planar, any vector connecting two atoms that straddle the longitudinal axis can be designated as the transverse vector.

With the old rotation model, twist, roll, and tilt are calculated directly from the direction cosines, by using Eq. 11. With the new model, twist is replaced by the rotation angle ϕ and is calculated from the trace of the transformation matrix, Eq. 13a; $c_x \sin\phi$ and $c_y \sin\phi$, Eq. 13, b-c, which define tilt and roll with the new rotation model, are calculated from the direction cosines of the transition matrix in terms of the transformation matrix asymmetries. We have chosen merely as an example to calculate the orientation variables based on the x-ray coordinates of the d(CGCGAATTCGCG) dodecamer (Drew et al, 1981) for these two rotation models; results are shown in Table 1. Data for local helix twist, based on the analysis of Dickerson and Drew (1981), are also presented in Table 1. In this paper, Dickerson and Drew presented two sets of twist data, one designated as global twist angles, the other as local twist angles. The local twist angles, which are shown in Table 1, are more closely related to the present model than are the global twist angles. This is because a local twist angle is given for each base pair step separately, as if the remainder of the molecule did not exist.

As expected, the two models show some differences in their tilt values. With the new model, absolute values for the tilt tend to be smaller than for the old model, as the direction cosines, $a(z'y)$ and $a(y'z)$, the difference of which is taken in Eq. 13b, can be of the same sign. On the other hand, the two terms in the roll expression, $a(z'x)$ and $a(x'z)$, are always of opposite signs, and, therefore, $c_y \sin\phi$, Eq. 13c, tends to be reasonably well approximated by $-\sin^{-1}a(x'z)$, Eq. 11. The comparison of roll* = $\sin^{-1}(c_y \sin\phi)$ with roll, Eq. 11b, is of interest. Symmetry considerations require that these two quantities be equal for all self-complementary base pair steps. This is because, for these base pair steps, $a(z'x) = a(x'z)$. This is indeed very nearly so for the CG, GC, and AT sequences (Table 1).

The twist, when defined as the rotation angle ϕ about the space-fixed direction, calculated as $\cos^{-1}[1/2(TrA\text{-}1)]$, tends to be in slightly better agreement with its definition, based on the local twist angle by using vectors between the two $C_{1'}$ atoms and their attached N atoms of one base pair and the corresponding atoms of the adjoining base pair (see Dickerson and Drew, 1981), than does the twist calculated from Eq. 11a.

Table 1 lists the three new orientation variables for each direction axis of the neighboring two base pairs, obtained from the experimentally determined atomic coordinates of the dodecamer. These are the rotation angle ϕ (Eq. 24), roll* (Eq. 26a), and tilt* (Eq. 26b). For comparison, the conventionally defined twist, roll, and tilt (Eq. 11, a-c) are also presented.

The magnitude and the direction of the rotation vector **G**, Eq. 4, for each base pair step, can be readily determined from

TABLE 1 Orientation variables calculated from direction cosines for rotation models I and II, based on the rigid x-ray form 1BNA

	ϕ	Twist	Twist†	Roll*	Roll	Tilt*	Tilt	Bend
Step								
CG	37.9	36.8	38.3	7.1	6.5	1.6	4.0	7.7
GC	38.9	38.6	39.6	−3.3	−3.0	−0.3	−2.0	3.6
CG	29.5	26.2	33.5	12.7	12.9	−0.6	2.4	13.1
GA	37.0	36.8	37.4	0.1	1.0	3.6	3.6	3.6
AA	37.4	37.3	37.5	1.0	0.5	1.6	1.9	2.0
AT	33.7	33.0	32.2	−5.8	−5.3	−1.5	−3.3	6.3
TT	35.6	35.5	36.0	0.1	1.0	−3.2	−3.2	3.3
TC	40.6	40.2	41.4	−2.8	−1.6	−3.1	−4.1	4.4
CG	30.0	28.6	32.3	8.4	8.3	0.2	2.4	8.7
GC	43.0	41.6	44.7	−9.7	−10.5	2.0	−1.6	10.6
CG	35.7	35.7	37.0	−1.0	−0.7	−0.8	−1.1	1.3
Average	36.3	35.5	37.3					

$\phi = \cos^{-1} 1/2[a(x'x)+a(y'y)+a(z'z)-1]$; twist = $\sin^{-1} a(x'y)$; bend = $\cos^{-1} a(z'z)$; roll* = $-\sin^{-1}(c_y \sin\phi)$; roll = $\sin^{-1} a(x'z)$; tilt* = $\sin^{-1}(c_x \sin\phi)$; tilt = $\sin^{-1} a(y'z)$.

Twist† is defined as the angle between the vectors between $C_{1'}$ and attached N atom of one base and the equivalent atoms of the neighboring nucleotide, taken from Dickerson and Drew (1981).

these three variables, as the three direction cosines, c_x, c_y, and c_z, form an orthogonal set. Fig. 2 displays these direction vectors. For a better visual display, the magnitudes of these vectors are shown with an arbitrary length. The origin point for each vector is the center of the C_6-C_8 vector.

COMPARISON OF THE TWO ROTATION MODELS WITH CONFORMATIONAL VARIABILITY

Introduction of the new model for the spatial orientation of sequences of base pairs provides an opportunity to compare the two sets of different ensembles of the canonically distributed and Boltzmann-averaged ensembles of conformational variables. This comparison can provide some information about the relative structural stabilities of the base steps and their sequence dependence of conformation.

The conformational states of neighboring base pairs are described by three translational and three orientation variables. Here we are performing calculations only on interacting bases; i.e., backbone atoms are ignored. The three orientation variables are either identified as the three rotational angles about the Cartesian axes (old rotation model) or, alternatively, as the rotation angle about a space-fixed axis and the direction cosines of the unit vector along this axis (new rotation model). The three translational and the three orientation variables are energy minimized by assuming that they are canonically distributed in phase space. For each base pair, a canonical ensemble consists of sets of these six degrees of freedom. All combinations of these degrees of freedom, or variables, are generated. Therefore, each variable is in adiabatic contact with all other variables except with itself. Boltzmann averages of each variable over this total set of conformations are then calculated.

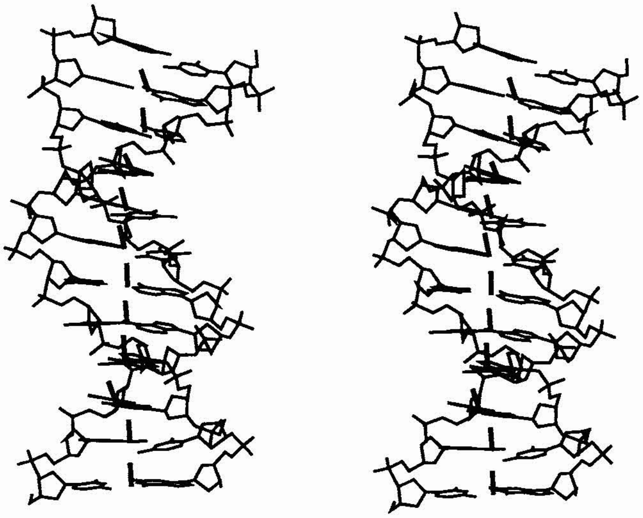

FIGURE 2 Stereo view of the dodecamer d(CGCGAATTCGCG), obtained from 1BNA, Brookhaven File (Bernstein et al., 1977). The individual rotation vectors for successive base pairs are superimposed onto the projected dodecamer. These vectors, which are the direction vectors **G** (Eq. 4), show the position and the orientation of the rotation axis.

In a previous publication (Mazur et al., 1989), we listed the conformational variables that spanned the phase space, and we specified their ranges and the number of values assigned to each variable. Only one variable for the intra-base-pair rotation, the propeller twist, was incorporated in the previous statistical calculations (see Sarai et al., 1988; Sarai et al., 1989). This variable was incorporated in the phase space, spanned by the three translation and the three rotation variables that orient two neighboring base pairs. In the current calculations, all three intra-base-pair rotational variables (i.e., propeller twist, buckle, and opening) are included. The large numbers of combinations of the various incremental values in practice restrict the number of these variables that can be considered in mutual adiabatic contact, hence the rotational variables that specify orientation of the two bases in a base pair, such as propeller twist, buckle, and opening, are taken to form a separate ensemble of canonically distributed variables. That is, the average values of the conformational variables for the single base pair and for the pair of adjacent base pairs (or steps) are calculated in succession.

For chains of three or more base pairs, the calculations are performed iteratively; that is, one base pair at a time is averaged. After the averages of the values of conformational variables are calculated in succession, the whole process is repeated a number of times, until one of the following situations is reached: a stationary steady-state solution is obtained or the calculated Boltzmann-averaged variables fluctuate in their values, without further decreases in the total energy. The inability to obtain a steady-state solution is more typical of the calculated buckle and propeller twist than of the other base pair parameters.

Previously, in a set of n base pairs, each iteration cycle consisted of n computations of the conformational averages for each base pair, taken in succession, except that the translational and rotational variables of the first base pair were assumed to be fixed in space, and only its propeller twist was allowed to be energy minimized.

In the present series of statistical calculations, each iteration cycle consists of $2n-1$ computations of the conformational averages. The computations of the intra-base-pair variables and of the inter-base variables that mutually orient, in space, two adjoining base pairs are performed successively. This new algorithm considerably reduces the number of conformers generated for each step of the iteration. In addition, the description of the rotational orientations of the two hydrogen bond-linked bases in a base pair is now more complete, because of the inclusion of buckle and opening angles, in addition to the previously included propeller twist.

After each set of the Boltzmann averages was computed, the calculated variables replace these variables that were computed during the previous iteration steps. Therefore, each base pair is energy minimized in the force field formed by the already energy-minimized base pairs. At the end of calculations, the entire chain is built of base pairs with the Boltzmann-averaged conformational properties.

This paper is concerned primarily with the rotation variables that orient, in space, two adjoining base pairs. As only one base pair is averaged at a time, the averages for a sequence of base pairs are also performed iteratively.

The overall rotation matrix must remain orthogonal even when the rotation variables are replaced by their Boltzmann averages. Therefore, the construction of the canonical ensemble is restricted by the need to preserve the orthogonality of the transformation matrix. If, instead, all nine matrix elements, or direction cosines, were replaced by their Boltzmann averages, the loss of orthogonality would lead to a translation and contraction, in addition to a rotation, of the initial vector. Moreover, such new coordinates would be energetically strained, as the initial distances between atoms in adjacent base pairs would be altered by a rotation with a nonorthogonal matrix.

In the transformation matrix that relates the spatial orientations of adjoining base pairs, all rotational degrees of freedom are replaced by their Boltzmann-averaged values. Therefore, the calculated values of the orientation variables, such as twist, roll, or tilt, will depend on the particular set of distributed values of the degrees of freedom. With the old model, these variables are the angles of rotation in each one of the three consecutive planar rotations of the base pairs. Different permutations of the order in which these rotations are performed can lead to somewhat different sets of the orientation variables, which, in turn, lead to different spatial orientations of the base pairs. Past calculations did show that the total energy of a set of base pairs is usually relatively unaffected by the sequence in which the three rotations are performed.

With the new model, there are four orientation variables, the rotation angle about the space-fixed axis and its three direction cosines with the Cartesian axes. These four variables can form linear combinations, resulting in four new variables, as they are restricted by the required orthogonality of the transformation matrix A. One such combination is presented by the four Euler parameters, Eq. 9. Therefore, different spatial orientations of the base pairs could also occur with the new rotation model. There is an advantage to using the four Euler parameters for averages of rotational variables; the required ranges of these variables are usually substantially narrower than for the direction cosines of the usual rotation angles. We have demonstrated this to be important in some recent unpublished calculations for intrabase pair degrees of freedom.

The possibility that different sets of orientation variables can have the same total potential energy is regarded as an example of the structural variability (see below) in a sequence of base pairs. The variabilities are associated with the presence of large fluctuations in most of the translational and rotational variables of individual base pairs. As the amounts of these fluctuations were shown to be sequence dependent (Sarai et al., 1989; Olson et al., 1993), the variabilities are also sequence dependent. The fact that different rotational schemes can lead to systems with different geometries but having the same total energy is of great importance. This shows that different numerical procedures could be used for constructing series of conformers that could physically coexist.

The term variability can be associated with structural degeneracy. Variability, like degeneracy, indicates that there is more than one conformation associated with the same energy. For the system to be considered to be structurally degenerate, there must be more than one linearly independent eigenvector associated with the same eigenvalue. In the present modeling, the eigenvalue is the total potential energy of the sequence of base pairs. This eigenvalue, which is a rather complicated function of the translational and rotational variables, or coordinates, can be considered to be structurally degenerate if there is more than one set of conformational variables (eigenvectors) corresponding to the same total energy (eigenvalue). The structures that have identical energies do not differ in their atomic compositions, but they do differ in their atomic coordinates. Degeneracy, or variability, and fluctuations are not synonymous. However, large fluctuations in some variables often lead to states having the same total energy but different sets of structural variables.

It is desirable to look more closely into this structural variability by adopting, in conjunction with the old rotation model, different formulations for successive rotations, leading to different sets of canonically distributed rotation angles. Euler angles, used previously (Sarai et al., 1988) serve as an example. With these angles, roll and tilt are no longer approximated by single rotation angles. Therefore, the values for roll and tilt, calculated from the Boltzmann-averaged Euler angles, are expected to differ from their values obtained on the basis of a rotation scheme that closely identifies them with the rotation angles about the y and x axes, respectively. However, actual calculations have shown that these two sets of values for the roll and the tilt exhibit rather small numerical differences. This is because most rotation schemes used previously, except for the recent publications by Babcock and Olson (1993) and by Babcock et al. (1994), in which a single rotation model was used, were based on three consecutive rotations. Therefore, we propose to explore further the problem of the multiplicity of solutions by comparing the sets of the averaged conformational variables obtained with the new rotation model with the ones obtained from the old model, which was based on one particular sequence of rotations that are closely associated with the twist, roll, and tilt. This particular rotation model was used by Marky and Olson (1994), Soumpasis and Tung (1988), and others. Specifically,

$$A = Z(\Omega)Y(\rho)X(\tau).$$

X, Y, and Z are single-rotation matrices that rotate a base pair around the x, y, or z axes, respectively. This rotation model is designated as the zyx rotation convention (see Goldstein, 1981).

In this paper, the structural variability of the sequences composed of two adjoining base pairs is investigated by comparing their Boltzmann-averaged rotational and translational variables obtained from the two rotation models that describe DNA conformations.

Care must be taken to preserve the orthogonality of the transformation matrix when the rotational variables are replaced by their Boltzmann averages. This orthogonality is always maintained when these variables are the rotational angles. This was the case with the old model, the one based on successive rotations of base pairs about Cartesian directions. With the new model, the three independent variables are the direction cosines c_x and c_y and the rotation angle ϕ, and c_z is recalculated from Eq. 8 at each step taken in the energy calculation. However, when taking the averages of these variables in a canonical ensemble, the orthogonality relation, Eq. 8, might not hold. To preserve the orthogonality of the transformation matrix, the three Boltzmann-averaged direction cosines are renormalized so that

$$[\langle c_x\rangle^2 + \langle c_y\rangle^2 + \langle c_z\rangle^2]C = 1, \tag{29}$$

where C is the normalization constant.

When the four Euler parameters, Eq. 11, are replaced by their Boltzmann averages, the following orthogonality condition must be satisfied:

$$[\langle e_0\rangle^2 + \langle e_x\rangle^2 + \langle e_y\rangle^2 + \langle e_z\rangle^2]C = 1, \tag{30}$$

where C is another normalization constant. Notice that the normalization condition, Eq. 30, includes all four angular variables. Yet another normalization relationship is required, which will normalize the trigonometric functions of rotation angle ϕ, that is,

$$(\langle \cos\phi\rangle^2 + \langle \sin\phi\rangle^2)C = 1. \tag{31}$$

There are two approaches for construction of orthogonal matrices in which the rotational variables are replaced by their Boltzmann averages. In one method, the angular variables are replaced by their Boltzmann averages, assuming that those variables are canonically distributed. In the second approach, the trigonometric functions of the angular variables, rather than the angles, form a canonically distributed set. In this second routine, the Boltzmann-averaged trigonometric functions of the angular variables are normalized (Eq. 31). If normalization is imposed, it was shown that those two approaches lead to numerically identical results.

The equivalency of these two methods of averaging is based on the following relations:

$$\cos\langle\phi\rangle = \langle\cos\phi\rangle\, C^{1/2} \quad \text{and} \quad \sin\langle\phi\rangle = \langle\sin\phi\rangle\, C^{1/2}. \tag{32}$$

The normalization constant C is given by

$$C = (\langle\cos\phi\rangle^2 + \langle\sin\phi\rangle^2)^{-1}. \tag{33}$$

In Eq. 32, the angle ϕ could be any rotational angle subject to the normalization of its sine and cosine functions.

Equation 32 might be proved analytically. At the present time, we are unable to make such a proof. Statistical calculations performed on several different sequences of base pair steps indicate that Eq. 32 is accurate. It should be emphasized that the present method of averaging treats the rotation angle and its trigonometric functions as stochastic variables, as their probability distribution, or ensemble, is continuous and nonnegative. Numerical considerations restrict the validity of Eq. 32 to rotation angles in the ($0 \leq \phi \leq \pi/2$) range, which is certainly true for the usual base pair orientation angles. The restricted range of the rotation angles for which Eq. 32 holds is easily demonstrated by considering arithmetic averages of randomly distributed sets of angles. It should be emphasized that, although Eq. 32 is presumably exact within the indicated range of the rotation angles, the replacement of trigonometric functions of these angles by their Boltzmann averages is an approximation. This can be attested by replacing all nine direction cosines of the transformation matrix by their averages. The ensuing matrix is no longer orthogonal.

The deviations of the normalization constants from unity provide an additional estimate of the extent of fluctuations in these angular variables. The values for the normalization constants were shown, on the basis of sample calculations, to be sequence dependent.

For both rotation models, the calculations were performed for tetramers NNMM. Data that refer to the central base pairs of the tetramers were extracted and used in the preparation

TABLE 2 Average energy, twist, bend, and displacement for rotation models I and II

	Energy I (kcal/mol)	Energy II (kcal/mol)	Twist I	Twist II	Bend I	Bend II	Displacement I (Å)	Displacement II (Å)
Base Step								
AA	34.0	33.9	39.0	39.2	1.82	1.69	0.33	0.0
TT	34.0	34.0	38.6	39.1	1.27	1.02	0.50	0.1
GG	2.5	3.0	32.5	35.5	1.51	2.35	0.79	0.74
CC	3.5	3.4	31.2	37.2	1.59	1.60	1.12	0.58
AT	34.8	34.6	27.0	32.2	0.97	0.26	0.87	0.42
TA	35.7	35.9	32.4	32.6	6.08	6.33	0.85	0.70
GC	0.0	0.7	37.9	38.9	0.78	0.62	0.51	0.43
CG	0.6	0.4	28.0	27.9	5.86	7.27	0.42	0.45
TG	17.6	17.8	34.8	35.3	4.60	4.86	0.25	0.11
CA	18.5	18.6	35.4	35.0	4.36	3.96	0.35	0.05
GA	18.7	18.6	35.3	36.3	2.07	2.24	0.30	0.35
TC	18.2	18.4	36.7	37.3	2.78	1.84	0.52	0.35
AC	17.8	18.0	34.6	37.2	1.91	1.73	0.47	0.42
GT	19.2	19.2	33.3	37.4	1.26	1.64	0.30	0.42
AG	17.0	17.1	30.8	32.5	1.44	1.62	0.52	0.38
CT	18.0	17.9	28.5	34.3	0.94	1.71	0.70	0.23
Average	18.1	18.2	33.5	35.5	2.44	2.69	0.53	0.34
SD	0.3		3.0		0.67		0.26	

$\epsilon_0 = 3$; $\alpha_0 = 1.2$ Å^{-1}; $r_0 = 3.0$ Å; twist = $\sin^{-1}\langle a(x'y)\rangle$; bend = $\cos^{-1}\langle a(z'z)\rangle$ (twist and bend are given in degrees).

of Tables 2 and 3. This procedure was used to attenuate and reduce the chain end effects. However, chain end effects still maintain some influence on the calculated orientation variables at the interior of the tetramer. This is particularly evident for the intra-base-pair buckle, which tends to be directed toward the interior of the sets of base pairs. This undesirable chain end effect can affect the symmetry relations in the complementary sets of base steps. Because of the approximate nature of computations, and because the tetramers may not be sufficiently long to eliminate completely the chain end effects, it was deemed necessary to perform independent calculations on all 16 permutations of the NM base steps in NNMM tetramers.

In our previous publications (Sarai et al., 1988, 1989; Mazur et al., 1989), buckle and opening were not included, but propeller twist was included in the phase space spanned by the rotational and translational variables of a base pair. Buckle was shown to interfere particularly with the twist and slide variables in base steps that show tendencies to assume the A conformation, particularly the GG/CC and AT steps. In these base steps, the presence of buckle increases twist and, consequently, decreases their mutual displacement, as twist and slide are anticorrelated. The total energy tends to decrease by 2–3 kcal/mol when buckle is included. It is conceivable that exclusion of buckle could provide a better description for the spatial orientations of the base pairs in the interior of a double helix sufficiently long so that end effects will be eliminated. However, exclusion of buckle will not provide a complete description of the mutual orientations for the two bases weakly linked by hydrogen bonds. In addition, crystal data indicate that buckle is considerable even in the interior of the polyoligomer.

The existence of different sets of Boltzmann-averaged orientation variables, strongly associated with their fluctuations and anisotropies (Sarai et al., 1989), leads to a concept termed in this paper structural variability. Detailed calculations did show (see Tables 2 and 3) that the total energies for the NNMM tetramers are almost independent of the rotation frame used for their spatial orientations. The two sets of canonically distributed variables obtained with either of the two models for setting the angular orientations of the sequences of base pairs are not restricted to the angular variables of a base pair. The phase space is also spanned by the three relative translations between adjacent base pairs; hence, the *x*, *y*, and *z* Cartesian components of the orientation vector of a base pair are displaced by the Boltzmann-averaged shift, slide, and rise, respectively. There is a problem of whether or not the rotation and translation operations performed on a sequence of base pairs commute. As a rule, these operations do not commute because of the presence of the molecular force field. (In most applications of rotation and translation operations in molecular physics, translations are assumed to be force-free; hence, the problem of their commutation does not arise). To explore the effect of commutation of the translation and rotation operations on the calculated energies and on the values of the Boltzmann-averaged orientational variables, we performed some exploratory calculations in which these two operations were performed in reverse order. It was found that, as expected on the basis of mathematical considerations, the Boltzmann averages and the total energies should be independent of the order in which the translation and rotation operations are performed. The calculations performed on tetramers show insignificant numerical differences.

A comparison of the two sets of rotation and translation variables obtained with either method of modeling base pair rotations provides us with an opportunity to inspect the possible ranges of these variables and how they are affected by a specific base pair. This is discussed in the following section.

CALCULATIONS

In the following calculations, the interaction energy is calculated as a sum over all atom pair contributions in the form

TABLE 3 Average energy, twist, bend, and displacement for rotation models I and II

	Energy I (kcal/mol)	Energy II (kcal/mol)	Twist I	Twist II	Bend I	Bend II	Displacement I (Å)	Displacement II (Å)
Base Step								
AA	34.0	33.9	35.1	37.5	1.24	0.41	0.025	0.012
TT	34.0	33.9	35.4	38.0	1.29	0.21	0.029	0.018
GG	2.0	1.8	33.6	37.9	2.66	2.10	0.059	0.041
CC	2.1	1.9	33.3	38.3	2.44	2.12	0.067	0.034
AT	34.1	33.9	31.0	35.7	1.48	1.09	0.125	0.100
TA	35.4	35.2	29.0	31.1	4.65	3.95	0.022	0.00
GC	0.9	0.5	30.5	34.6	1.99	0.63	0.068	0.057
CG	0.2	0.0	27.3	29.9	4.69	4.73	0.028	0.029
TG	17.8	17.6	30.3	33.0	4.59	4.36	0.018	0.014
CA	17.6	17.5	29.0	32.7	4.69	4.37	0.011	0.010
GA	18.7	18.5	30.8	34.4	2.50	1.52	0.037	0.037
TC	18.5	18.6	31.7	35.9	2.52	1.40	0.057	0.024
AC	17.8	17.7	32.3	36.4	1.37	0.44	0.068	0.044
GT	18.1	18.0	31.9	36.6	1.59	0.23	0.053	0.067
AG	17.1	17.0	31.0	34.5	2.22	1.76	0.05	0.011
CT	17.3	17.1	30.3	34.6	2.35	1.24	0.035	0.015
Average	17.8	17.7	31.4	35.2	2.64	1.91	0.045	0.032
SD	0.2		3.9		0.87		0.19	

$\epsilon_0 = 6$; $\alpha_0 = 0.5$ Å^{-1}; $r_0 = 3.0$ Å; twist $= \sin^{-1} \langle a(x'y) \rangle$; bend $= \cos^{-1} \langle a(z'z) \rangle$ (twist and bend are given in degrees).

of attractive dispersion and short range repulsion, with 6–12 potential and 10–12 potential for the hydrogen bonds, and electrostatic interactions. For electrostatic interactions, the electrostatic potential V_{es} (in kcal/mol), which takes account of solvent screening action, was taken as (Mazur and Jernigan, 1991)

$$V_{el} = 332 \sum_{ij} (1/\epsilon_0 r_{ij}) C_i C_j \exp(-\alpha_0 (r_{ij} - r_0)) \qquad r_{ij} > r_0$$

$$V_{el} = 332 \sum_{ij} (1/\epsilon_0 r_{ij}) C_i C_j, \qquad r_{ij} < r_0$$

where α_0 is the ionic screening parameter in units of Å^{-1} and ϵ_0 is the limiting value of the dielectric constant at distances less than r_0. The parameters for the 6–12 potential were taken from Zhurkin et al. (1981). The parameters for the 10–12 potential for the hydrogen bonds were taken from Poltev and Shulyupina (1986). C_i and C_j are partial atomic charges taken from Zhurkin et al. (1981). r_0 is the contact distance between interacting charges and is qualitatively related to the sums of van der Waals radii. A value of 3.0 Å is adopted for this distance, which is usually the closest distance between atoms located in the adjoining base pair planes. Two different parameter sets for the electrostatic interactions were used, with different values of α_0, the ionic screening parameter, and ϵ_0, the limiting dielectric constant at distances less than r_0: one dielectric model with $\alpha_0 = 1.2$ Å^{-1} and $\epsilon_0 = 3$ and another dielectric model with $\alpha_0 = 0.5$ Å^{-1} and $\epsilon_0 = 6$. The first dielectric model represents a highly solvated system at a high relative humidity. The other dielectric model represents a low solvated (low humidity) system. A simple interpretation of these electrostatic parameters in terms of solvent conditions was previously presented by Mazur et al. (1989) and by Mazur and Jernigan (1991). The purpose of introducing two different electrostatic models is to examine whether the differences between the orientation coordinates of the base pairs obtained with the two rotation models discussed above, that is, the variabilities of the base pair sequences, might be affected by the details of the solvent model.

Results presented below are for the middle base step of a tetramer. Data that deal with the variables that describe intra-base-pair rotations are not presented here, although they were included in the calculations, as the statistical methods used in the calculations (specifically, the propeller twist, buckle, and opening) are identical for both models of base pair rotations and, therefore, not much affected by the presence of variabilities in the rotational and translational variables of the base steps. In Tables 2 and 3, data for the total energy, twist, bend, and displacement are presented. The displacement is defined here as a square root of ($\text{slide}^2 + \text{shift}^2$). Slide and shift are determined from the centers of the lines joining the C_8 atom of purine and the C_6 atom of pyrimidine in the two adjacent and stacked base pairs. Bend is calculated from the averaged $a(z'z)$. Calculations were carried out for all 16 NM base steps of base pairs at the center of NNMM tetramers. The same ranges of variables, the same increments, and the same total number of points were used in both coordinate systems. With the previously reported calculations (Sarai et al., 1989; Mazur et al., 1989), appropriate ranges for twist and slide were found to vary from one base step to another, and they were centered, as closely as possible, at their minimal free energy values. In the present calculations, the same ranges were used everywhere, except for the pyrimidine-purine base steps, for which the ranges for angular variables about x and y directions were shifted to reflect their significantly larger bends. These procedures were followed to assess better the effect of structural variabilities.

Here we list the ranges for all translation and rotation variables, together with their increments and the number of points in their summands.Angular variables (in degrees):

1. Twist, Ω, and ϕ, 35.9 ± 16; step, 4; 9 points.
2. Roll, ρ, and $\sin^{-1} c_y$, (−15,15); step, 3; 11 points.
3. Tilt, τ, and $\sin^{-1} c_x$, (−15,15); step, 3; 11 points.
4. Propeller twist, 11.5 ± 7.5; step, 2.5; 7 points.
5. Buckle, (−7.5,7.5); step, 2.5; 7 points.
6. Opening, (−2,2); step, 2.5; 5 points.

For pyrimidine-purine base steps, $\sin^{-1} c_y$ and $\sin^{-1} c_x$, (−24,12); interval, 3; 13 points.

Translations (in Å):

1. Slide, (−0.8,2.0); step, 0.4; 8 points.
2. Shift, (−0.8,0.8); step, 0.4; 5 points.
3. Rise, (−0.1,0.1); step, 0.5; 5 points.

In the Tables and in the Discussion below, model I refers to the new rotation model used for the construction of the coordinate frame for base pairs in the double helix and model II refers to the previously used rotation model, in which the transformation matrix is expressed as a product of three matrices for rotations about three axes.

Energies are presented as excess energies over the energy of a tetramer with the lowest total energy. The tetramers with the lowest energies that serve as reference points are the GGCC in Table 2 and the CCGG in Table 3. The last lines in Tables 2 and 3 present the standard deviations of the differences in the displayed averages for the two models.

As mentioned above, the orientation variables that are presented in Tables 2 and 3 (twist and bend) are uniquely represented by single direction cosines of the transformation matrix A. These direction cosines are expressed in terms of the Boltzmann-averaged orientation variables. The Boltzmann-averaged orientation variables are calculated separately for rotation model I and for rotation model II. For the rotation model I, the Boltzmann-averaged variables are the rotation angle about the space-fixed axis and the three direction cosines of the rotation axis with respect to the three Cartesian directions. These three averaged direction cosines are normalized. For the rotation model II, the Boltzmann-averaged orientation variables are the three rotation angles Ω, ρ, and τ (Dickerson, et al., 1989), which approximate the averaged twist, roll, and tilt, respectively. For each rotation model, the orientation variables are evaluated in the phase space spanned by the translational and internal rotational variables of a base pair. As explained above, calculations of the conformational variables (translations and rotations) are performed iteratively; that is, one base pair is averaged at a time, and their average values are calculated in succession.

In addition, the intra-base-pair rotational variables (propeller twist, buckle, and opening) are also calculated iteratively, assuming that they form separate ensembles in phase space. Tables 2 and 3 tabulate average twist and bend, defined uniquely in terms of the direction cosines of the transformation matrix. Average twist is defined as $\sin^{-1}\langle a(x'y)\rangle$ and average bend as $\cos^{-1}\langle a(z'z)\rangle$, for both models.

It is not necessary to restrict the calculations of the orientation variables to those expressed in terms of single direction cosines of the matrix A. Certain orientation variables, such as the direction cosines of the rotation axis, are expressed in terms of linear combinations of several direction cosines. For example, $\langle\Omega\rangle$, the average twist, could be replaced by $\langle\phi\rangle$, which is defined as $\cos^{-1}(Tr[A] - 1)/2$. $Tr[A]$ is a trace of the Boltzmann-averaged transformation matrix, properly normalized to maintain its orthogonality. Therefore, three direction cosines of the transformation matrix A are included in this definition of $\langle\phi\rangle$.

In addition to the averaged twist and bend, Tables 2 and 3 also present displacement, defined as the square root of the sum of squares of the averaged slide and shift, which are the averaged translations of the base pairs in the two directions normal to their pseudo-dyad axis. Two values of displacement for the two different rotation models are presented in these tables. The displacement is often identified as slide, as shift is usually considerably smaller than slide. AC/GT and GA/TC base steps are exceptions to this. Because slide can be regarded as a reaction coordinate for the B-A transition (Mazur et al., 1989), attention is given to its direction, or sign. For most base steps, slide is negative, as it is directed toward the ascending (5′-3′) strand (Dickerson et al., 1989). A notable exception is the CG base step, with a slide of ~0.3Å.

DISCUSSION

Results presented in Tables 2 and 3 are discussed below.

First energies for the two models of the NM base steps are practically identical. With the highly screened dielectric (Table 2), energy differences exceed 0.2 kcal/mol only for the GG and GC sequences. Average differences in the energies between models I and II are only 0.09 kcal/mol, for the high-screened dielectric situation (Table 2), and −0.16 kcal/mol for the low-screened dielectric (Table 3). The mean square deviations of the energy differences are 0.28 and 0.16 kcal/mol, respectively. Considering the fact that the energies cover a range of ~30 kcal/mol, the calculated energy differences and their root mean square deviations are quite negligible, justifying the conjecture that the two sets of averaged orientation and translational variables, having nearly identical total energies, might be considered to represent the structural variabilities of base pair steps.

Second, notice that the model I is more undertwisted than model II. As twist and slide are anticorrelated, decreased amounts of twist are accompanied by small increases in the base pair displacements. The twist coordinate is characterized by large fluctuations and remains virtually constant in energy over the large range of its allowed values, which is, in the present work, 20–52°. Therefore, large variabilities ensue for this coordinate. Twist differences are larger with lower dielectric screening (Table 3). This fact is explained by relatively larger structural stabilities of systems in which electrostatic interactions are short in range, through a larger screening action of the solvent, as the minima in the potential energies tend to be steeper. This could also mean that the B form of DNA displays larger variabilities for this coordinate; that is, it is conformationally less stable than the A form.

Of particular interest are the AA/TT and GG/CC base steps. Tables 2 and 3 show little variability in the AA/TT base steps. Their conformational variability is more evident in the lower electrostatically screened case. To the contrary, Table 2 shows larger differences between the GG and CC base steps, both in their twist and displacement coordinates. Highly solvated GG/CC base steps are actually known to tend to undergo a B-A transition. The data in Table 2 imply a possible coexistence of these base steps in the A and B polymorphs of DNA (Mazur et al., 1991). However, a complete B-A transition would require a sugar pucker rotation from its $C_{2'}$-*endo* mode associated with B form to the $C_{3'}$-*endo* mode of the A form of the DNA. Because of the absence of the backbone atoms in the calculations, there is no certain way that the aforementioned variability of the GG/CC base steps can be related to the A and B polymorphs. Different polymorphic forms can also exist within the B-DNA family.

Overall, the AT sequences tend to be more overtwisted than the GC sequences. The AA/TT are also the most overtwisted base steps. The CG base step remains the most underwound with either model. The base steps that are considerably overtwisted, in particular the AA/TT, show smaller structural differences, or variabilities, than is the case with the undertwisted base steps. This conforms with the overall assessment, that the AA(TT) base steps are more typically B form than A form, and, therefore, are less likely to undergo a conformational transition to A-form DNA, which is known to be underwound in comparison with the B form. In addition to GG/CC, large twist differences between the two models I and II are found in the following base steps: AC/GT, AG/CT, and AT, for both dielectric models. AT was found previously (Sarai et al., 1989) to display a particularly large fluctuation in its twist coordinate. Of these base steps, GG/CC, AT, and AG/CT were previously shown to be the ones most likely to be found in the A form of DNA (Mazur et al., 1989). In addition, TA also shows some preference for the A form.

A third point is that, overall, the bend coordinate shows smaller structural variabilities than does the twist coordinate. The differences between the two models in the average values for the bend coordinate, taken over all base steps, are less than 1 degree. The mean square deviation for these differences, is 0.67°, with high electrostatic screening, and 0.87°, with low screening (Tables 2 and 3). Bends are found to be considerably larger for all base steps in which a pyrimidine base precedes a purine, that is, for TA, CG, and CA/TG, than for other steps. This unique property of these base steps is well maintained for both coordinate models and for both forms of the dielectric screening. These unusually large bends (about twice as large) suggest a mini-kinked sequence of base pairs (Ul'yanov and Zhurkin, 1982) rather than a gradually bent one, with most of the bend located at the junction points between pyrimidine and purine nucleotides. All other base steps show smaller values for bends and larger relative variances in their values. This is particularly true for the GG/CC base steps. Sarai et al. (1988) did show that these base steps are characterized by large anisotropies in their roll and tilt fluctuations. It was observed previously (Mazur et al., 1989) that the large roll angles (or bends, as roll contributes the major part of the bend) in the pyrimidine-purine base steps are accompanied by smaller twists. The data presented above confirm this finding. These four base steps have an average twist of 32.7° (high dielectric screening, Table 2) and 30.3° (low dielectric screening, Table 3). All other base steps manifest an average twist of ~35.1 degrees (high dielectric screening) or 34.3° (low dielectric screening).

Fourth, twist and displacement (or slide) are anticorrelated. This is seen in the averaged twists and displacements, taken over all 16 base steps, for both models (last line, Tables 2 and 3).

It was proposed earlier in this paper that the normalization constant, which preserves the unit vector in its normalized form when its eigenvectors are replaced by their Boltzmann averages, could reflect the fluctuations of the orientation variables. Our calculations do show that there is a general correlation between the normalization constant and the twist differences. The normalization constants for all 16 base steps range between 0.991 and 0.976. Higher values of the normalization constant are typical of all pyrimidine-purine base steps and of the AA/TT base step. All of these base steps show greater conformational stability than the others. It is possible that large bends, typical of the pyrimidine-purine base steps, tend to lock the base pairs into relatively steeper potential energy wells, which could contribute to their relative conformational stabilities.

The least stable base steps are the above mentioned GG/CC, AT, and CT/AG. Those are the base steps that are also more likely to assume the A-form DNA. AT also shows an unusually low value for its normalization constant (0.976). This also confirms the finding reported previously (Sarai et al., 1989) about the exceptional flatness of the free energy dependence on twist for this base step.

Similar conclusions obtain for the normalization constant for the trigonometric functions of the rotation angle and its dependence on the specific base pair step. The calculated values of this constant tend to be closer to unity, their values for a straight double helix, than is the normalization constant for the average unit vector. Nonetheless, the same conclusions hold for its correlation with the variability of the twist.

We previously mentioned that the presence of buckle in the force field can interfere with the twist and displacement variables, pushing the conformation slightly toward the B form. This is particularly evident in GG/CC and AT sequences. We repeated some of the calculations with the buckle removed from the force field. The results show that the conformational variability of these base steps is maintained. We can therefore conjecture that this variability is not related to chain end effects.

In conclusion, the structural variabilities are mostly manifested in the twist coordinate. The twist is known to be strongly anticorrelated with the slide, or the mutual displacement of adjacent base pairs. Large slides and undertwist are typical of A-form DNA. This variability points to the possible coexistence, under certain conditions, of B-DNA and A-DNA forms. This is only a possibility, as other factors (solvent action or coupling with backbone) could modify some of these variabilities. Thus, the AA/TT and the pyrimidine-purine pairs of base steps are stiffer than the other base steps. Large electrostatic screening enhances this structural stiffness. Bends are shown to be structurally invariant; that is, they are not seriously affected by solvent action or by small changes in the helix character of the molecule, such as B or A forms of DNA.

The new model I is more biased toward underwound helices with accompanying larger displacements than is model II. Overall, the qualitative differences between these two models, in particular for sequences of base pairs known to be structurally stiff, are rather small. However, as pointed out earlier, there are conceptual advantages to the new approach.

Both models I and II are centered around the straight double helix. Therefore, the method of averaging the conformational variables in their canonically distributed ensembles is restricted to conformational space in the vicinity of the straight B-DNA molecule. A more complete picture of the conformational preferences of base pairs would be obtained if different or broader ranges of conformational variables were used. This could lead to an exploration of other conformational energy minima that, because of a presence of potential barriers, could not be reached from the initial conformational models selected in the present study.

In this paper, a novel rotation model, based on a single rotation about a space-fixed axis, was applied to calculations of the relative orientation variables of a sequence of base pairs. This type of model, first used by Babcock et al. (1994) for nucleic acid structure analysis, has conceptual advantages over previously used rotation schemes to orient a base pair unit in space. The new model establishes a transformation matrix that orients two neighboring base pairs, for which the elements, or direction cosines between the unit vectors of the two base pairs, depend only on the direction cosines of the space-invariant rotation axis and on the angle of rotation about this axis. In this model, roll and tilt are no longer identified as rotations about particular long and short directions in one of the two adjoining base pairs. Rather, roll and tilt are directly associated with the opening of the angles toward the minor groove and toward the backbone strand, respectively. A method to calculate these angles from any coordinate set has been presented. Moreover, the new model preserves the symmetry requirements for complementary sets of base pair steps as, in these sets, the two rolls are equal and the two tilts are equal but have opposite signs. For the self-complementary base pair steps, (CG, GC, TA, and AT), tilt is zero, as required by the symmetry considerations. These symmetry-based conclusions ignore the nonplanarity of base pairs. Table 1 shows that, in the self-complementary base pairs (only CG, GC, and AT are sampled in the dodecamer), tilt, based on the new rotation model, is considerably smaller than the one based on the conventional rotation model, but its absolute average value of ~1° could be associated with the nonplanarity (buckle, propeller twist) of the base pairs.

The authors thank Dr. Akinori Sarai of Riken Life Science Center, Tsukuba, Ibaraki, Japan, for his interest and helpful discussions in various aspects of this work. The authors acknowledge Dr. G. Raghunathan of the Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, for bringing our attention to the existence of this space-fixed rotation axis approach that has been used here for the specification of the orientation variables of sets of base pairs.

We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Frederick Biomedical Supercomputer Center at the Frederick Biomedical Supercomputer Center of the Frederick Cancer Research and Development Center under contract N01-CO-74102 with Program Resources, Inc./DYNCORP. The content of this article does not necessarily reflect the views of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organisations imply endorsement by the U.S. government.

REFERENCES

Babcock, M. S., and W. K. Olson. 1994. The effect of mathematics and coordinate system on comparability and "dependencies" of nucleic acid structure parameters. *J. Mol. Biol.* 237:98–124.

Babcock, M. S., E. P. D. Pednault, and W. K. Olson. 1993. A users guide to a collection of new analysis programs. *J. Biomol. Struct. Dyn.* 11:597–628.

Babcock, M. S., E. P. D. Pednault, and W. K. Olson. 1994. Nucleic acid structure analysis: mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* 237:125–156.

Bhattacharyya, D., and M. Bansal. 1988a. A general procedure for generation of curved DNA molecules. *J. Biomol. Struct. Dyn.* 6:93–104.

Bhattacharyya, D., and M. Bansal. 1988b. A self-consistent formulation for analysis and generation of non-uniform DNA structures. *J. Biomol. Struct. Dyn.* 6:635–653.

Bernstein, F. C., G. J. Williams, E. E. Meyer, Jr., J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structure. *J. Mol. Biol.* 25:535–542.

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

Dickerson, R. E., and H. R. Drew. 1981. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* 149:761–786.

Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA.* 78:2179–2183.

Flory, P. J. 1969. *Statistical Mechanics of Chain Molecules.* Wiley-Interscience, NY. 19–22.

Fratini, A. V., M. L. Kopka, H. R. Drew, and R. E. Dickerson. 1982. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT^{Br}CGCG. *J. Biol. Chem.* 237:14686–14707.

Goldstein, H. 1981. Classical Mechanics, 2nd. ed. Addison Wesley Publishing Co., New York. 672 pp.

Jeffreys, H., and B. S. Jeffreys. 1946. Methods of Mathematical Physics. Cambridge University Press, Cambridge, UK. 718 pp.

Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biol. Struct. Dyn.* 6:63–91.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

Marky, N. L., and W. K. Olson. 1994. Configurational statistics of the DNA duplex: extended generator matrices to treat the rotations and translations of adjacent residues. *Biopolymers.* 34:109–120.

Maroun, R. C., and W. K. Olson. 1988. Base sequence effects in double helical DNA. III. Average properties of curved DNA. *Biopolymers.* 27: 585–603.

Mazur, J., and R. L. Jernigan. 1991. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers.* 31:1615–1629.

Mazur, J., A. Sarai, and R. L. Jernigan. 1989. Sequence dependence of the B-A conformational transition of DNA. *Biopolymers.* 28:1223–1233.

Milne, E. A. 1948. Vectorial Mechanics. Interscience Publishers, New York. 143–148.

Olson, W. K., N. L. Marky, R. L. Jernigan, and V. B. Zhurkin. 1993. Influence of fluctuations on DNA curvature: a comparison of flexible and static wedge models of intrinsically bent DNA. *J. Mol. Biol.* 232:530–554.

Olson, W. K., and A. R. Srinivasan. 1988. The translation of DNA primary base sequence into three dimensional structure. *Comp. Appl. Biosci.* 4:133–142.

Olson, W. K., A. R. Srinivasan, M. Hao, and J. L. Nauss. 1988. Base sequence dependent models of DNA curvature: the geometric interdependence of complementary residues. *In* Structure and Expression, Vol. 3. W. K. Olson, M. H. Sarma, R. Sarma, and M. Sundaralingam, editors. Alan Liss, New York. 225–242.

Poltev, V. I., and N. V. Shulyupina. 1986. Simulation of interactions between nucleic acid bases by refined atom-atom potential functions. *J. Biomol. Struct. Dyn.* 3:739–765.

Raghunathan, G., H. T. Miles, and V. Sasisekharan. 1993. Symmetry and molecular structure of a DNA triple helix: $d(T)_n.d(A)_n.d(T)_n$. *Biochemistry.* 32:455–462.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1988. Sequence dependence of DNA conformations: means and fluctuations. *In Structure and Expression*, Vol. 3. W. K. Olson, M. H. Sarma, R. Sarma, and M. Sundaralingam, editors. Adenine Press, NY. 213–223.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1989. Sequence dependence of DNA conformational flexibility. *Biochemistry.* 28:7842–7849.

Sklenar, H., R. Lavery, and B. Pullman. 1986. The flexibility of the nucleic acids. I. "Sir", a novel approach to the variation of polymer geometry in constrained systems. *J. Biol. Struct. Dyn.* 3:967–987.

Soumpasis, D. M., and C.-S. Tung. 1988. A rigorous base pair oriented description of DNA structures. *J. Biomol. Struct. Dyn.* 6:397–420.

Srinivasan, R., V. Geetha, J. Seetharaman, and S. Mohan. 1993. A unique or essentially single parametric characterisation of biopolymeric structures. *J. Biomol. Struct. Dyn.* 11:583–596.

Ulanovsky, L. E., and E. N. Trifonov. 1987. Estimation of wedge components in curved DNA. *Nature.* 326:720–722.

Ul'yanov, N. B., and V. B. Zhurkin. 1982. Study of the flexibility of complementary dinucleoside phosphates by the Monte-Carlo method. *Molek. Biol.* 16:1075–1085.

van Dam, H., J. D. Louck, and L. C. Biederharn. 1990. Rotation and angular momentum. *In* Encyclopedia of Physics. R. G. Lerner and G. L. Trigg, editors. VCH Publishers, New York. 1076–1081.

von Kitzing, E., and S. Diekmann. 1987. Molecular mechanics calculations of $dA_{12}.dT_{12}$ and of the curved molecule $d(GCTCGAAAAA)_4$. $d(TTTTTCGAGC)_4$. *Eur. Biophys. J.* 15:13–26.

Zhurkin, V. B., V. I. Poltev, and V. I. Florent'ev. 1981. Atom-atom potential functions for conformational calculations of nucleic acids. *Molek. Biol.* 14:887–895.