J. Mol. Biol. (1997) 273, 668–680

JMB

Structure and Conformation of Helical Nucleic Acids: Analysis Program (SCHNAaP)

Xiang-Jun Lu[1], M. A. El Hassan[2] and C. A. Hunter[1]*

[1]*Krebs Institute for Biomolecular Science Department of Chemistry University of Sheffield Sheffield, S3 7HF, UK*

[2]*Department of Engineering University of Cambridge Cambridge, CB2 1PZ, UK*

**Corresponding author*

We present a new versatile program, SCHNAaP, for the analysis of double-helical nucleic acid structures. The program uses mathematically rigorous and fully reversible procedures for calculating the structural parameters: the Cambridge University Engineering Department Helix computation Scheme (CEHS) is used to determine the local helical parameters and an analogous procedure is used to determine the global helical parameters. These parameters form a complete set that conforms to the ''Cambridge Accord'' on definitions and nomenclature of nucleic acid structure parameters. In addition to the two standard Watson-Crick base-pairs, the program handles mismatched base-pairs and chemically modified bases. An analysis of the sugar-phosphate backbone conformation is included. Standardized base-stacking diagrams of each dinucleotide step with reference to the mid-step triad are generated. Structures are classified as one of the four polymorphic families, *A*/*B*, *Z*, *W* or *R*, although *W*- and *R*-DNA (two types of hypothetical structure) have yet to be observed experimentally.

Keywords: DNA conformation; step parameters; base-pair parameters; global helical parameters; base-stacking diagrams

Introduction

Since the ''Cambridge Accord'' on the definitions and nomenclature of nucleic acid structure parameters was formulated (Diekmann, 1989), several groups have developed different approaches to determining these parameters (El Hassan & Calladine, 1995; Bansal *et al.*, 1995; Mazur & Jernigan, 1995; Babcock & Olson, 1994; Babcock *et al.*, 1994; Jursa, 1994; Tung *et al.*, 1994; Lavery & Sklenar, 1989). These methods , although they all fall within the guidelines of the Cambridge Accord, are subtly different in a way that causes problems in comparing parameters and structures analyzed or generated by the various alternative procedures (Olson, 1996). NewHelix (Dickerson, 1985; Fratini *et al.*, 1982) has been widely used, especially among experimental biologists. The parameters it defines, while intuitively easy to understand, are incomplete (Diekmann, 1989; El Hassan & Calladine, 1995). Moreover, the parameters are defined with reference to a best-fit straight global helical axis, which makes it unsuitable for dealing with strongly curved DNAs as commonly seen in DNA-protein complexes. Curves (Lavery & Sklenar, 1988, 1989), on the other hand, calculates parameters with respect to an optimal ''curved'' global helical axis. The method of Babcock *et al.* (1994) uses a local axis for each dinucleotide step, but uses a single rotation to define the relative orientation between successive base-pairs. The CEHS scheme (El Hassan & Calladine, 1995) uses the concept of a mid-step triad (MST) as recommended by the Cambridge Accord to work out the six rotational and translational parameters. It is rigorous, reversible and conceptually simpler than the above methods

Here, we present the computer implementation of the CEHS scheme and its extension to calculate a set of global helical parameters. The program also generates base-stacking diagrams of each dinucleotide step with reference to the mid-step triad, and automatically classifies DNA structures as one of the four polymorphic families, *A*/*B*, *Z*, *W* or *R*, although *W*- and *R*-DNA have yet to be

Abbreviations used: CEHS, Cambridge University Engineering Department helix Computation scheme; MST, mid-step triad; MBT, mid-base triad; PDB, Protein Data Bank; NDB, nucleic acid data base.

0022–2836/97/430668–13 $25.00/0/mb971346

Table 1. The nomenclature and symbols of the rotational and translational parameters recommended by the ''Cambridge Accord''

		Local					
		Step parameters		Base-pair parameters		Global	
Motion	Axis	Name	Symbol	Name	Symbol	Name	Symbol
Rotation	*z*	Twist	Ω	Opening	σ	Twist_g	Ω_g
	y	Roll	ρ	Propeller	ω	Tip	θ
	x	Tilt	τ	Buckle	κ	Inclination	η
Translation	*z*	Rise	D_z	Stagger	S_z	Rise_g	*dz*
	y	Slide	D_y	Stretch	S_y	*y*-displacement	*dy*
	x	Shift	D_x	Shear	S_x	*x*-displacement	*dx*

observed experimentally. The program will handle most structures including right or left-handed helices with Watson-Crick or mismatched base-pairs and intact or chemically modified bases. The program provides the following structural parameters: (1) local CEHS base-pair parameters; (2) local CEHS base-step parameters; (3) global helical parameters as defined below; (4) sugar-phosphate backbone parameters; (5) standardized base-stacking diagrams, in which each step is oriented with respect to its mid-step-triad (MST); and (6) classification of the structure as an *A*, *B*, *Z*, *W* or *R* polymorph.

The procedure

Definitions of the structural parameters

The nomenclature and symbols of various rotational and translational parameters, both local and global, as recommended by the Cambridge Accord are given in Table 1 (Diekmann, 1989).

Reference triads

The position of the *i*th base-pair (or base) is defined by a (1 × 3) vector, $\mathbf{r}_i$, representing the origin of the triad; and the orientation is defined by a (3 × 3) matrix, $\mathbf{T}_i$, where each column gives the direction cosines of the *x*, *y*, and *z*-axes of its reference triad, denoted by the unit vectors $\hat{x}_i$, $\hat{y}_i$ and $\hat{z}_i$, respectively.

In order to define the reference triads of the bases and base-pairs, we need to arbitrarily assign one strand as strand I and the other as strand II. The positive sense of each strand is along the 5′ → 3′ direction. The reference triads are now defined as follows (Figure 1).

(1) Base-pair triad

(i) The origin is the midpoint of the line connecting C_8 for purines (R) and C_6 for pyrimidines (Y).

(ii) The *y*-axis lies along the RC_8-YC_6 line. Its positive direction points from strand II to strand I.

(iii) The *z*-axis is defined as follows. We calculate the normal to the least-squares plane (Blow, 1960; Schomaker *et al.*, 1959) through all atoms of the base-pair (excluding hydrogen and C_1' atoms) and let its positive direction be along the 5′ → 3′ direction of strand I. In general, the *y*-axis and the normal vector are not exactly orthogonal to each other, due to the non-planarity of the base-pair. We therefore decompose the base-pair normal into two components (Stephenson, 1973); one is parallel with the *y*-axis, and the other perpendicular to it. It is the perpendicular component that is used as the *z*-axis.

(iv) The *x*-axis completes a right-handed triad with the *y* and *z*-axes. In *A*/*B* and *W*-DNA, the positive *x*-axis direction points along the short axis of the base-pair from the minor groove side to the major groove side. However, in *Z* and *R*-DNA, the positive *x*-axis direction points from the major groove side to the minor groove side (see discussion of polymorph structures later).

(2) Base triads

(i) The origin is given by the midpoint of the N_1-C_4 line for R, and the N_3-C_6 line for Y.

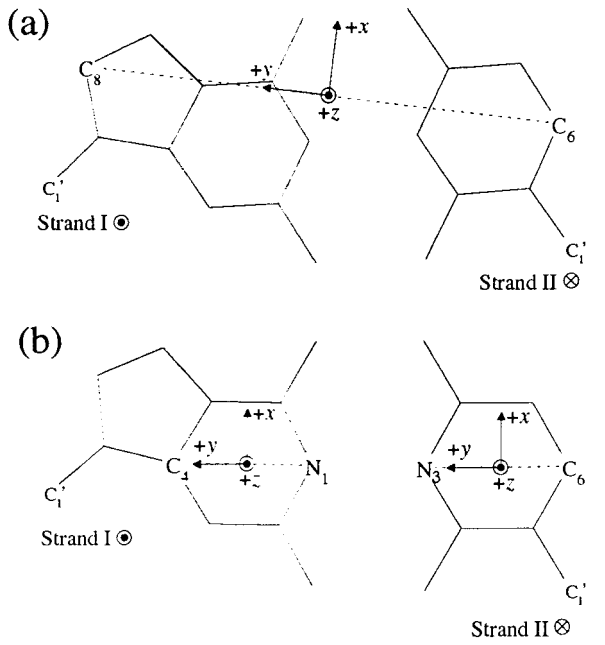

Figure 1. Base-pair (a) and base (b) reference for *A*/*B* and *W*-DNA used in SCHNAaP; ⊙ points out of the plane of the paper, and ⊗ points into the plane of the paper.

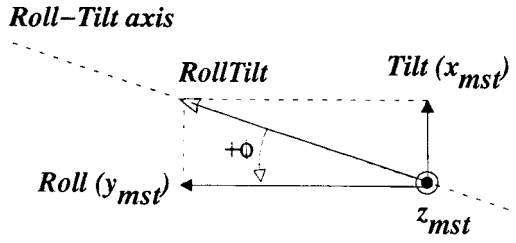

Figure 2. Roll/Tilt vectorial addition and decomposition.

(ii) The y-axis lies along the N_1-C_4 line for R and the N_3-C_6 line for Y. Its positive direction points from strand II to strand I.

(iii) The z-axis is defined as the normal to the least-squares plane through all atoms in the base (excluding hydrogen and the C_1' atom). Its positive direction is along the $5' \rightarrow 3'$ direction of strand I.

(iv) The x-axis completes a right-handed triad.

Local base-step and base-pair parameters

We will begin with a summary of the key features of the CEHS scheme (El Hassan & Calladine, 1995) on which this program is based. CEHS is a local scheme using the concept of the MST as recommended by the Cambridge Accord to ensure the base-step parameters calculated are invariant upon inversion of the direction in which the step is read.

Determination of the base-step parameters

A key feature of CEHS is the use of Roll/Tilt vectorial addition to avoid the non-commutativity problem associated with the angular addition. The two rotations, Roll and Tilt, about the y and x-axes of the MST, respectively, are combined to give a single rotation of angle RollTilt (Γ) about the Roll-Tilt axis, which is oriented at an angle of ϕ to the MST y-axis (Figure 2). The analysis procedure described below first determines the values of Γ and ϕ, and then decomposes these into values of Roll and Tilt.

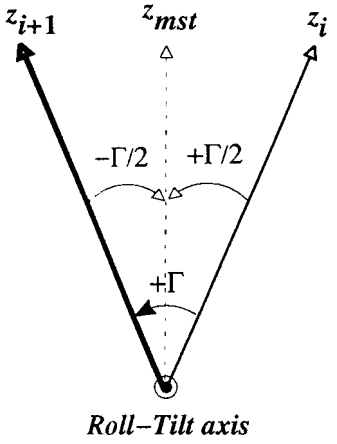

Figure 3. Rotations about the Roll-Tilt axis with a view down the end of the Roll-Tilt axis vector. Only the z-axes of the three triads are shown for clarity.

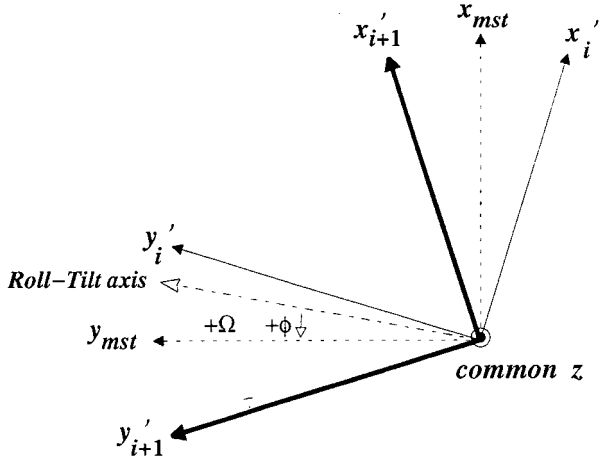

Figure 4. The transformed base-pair reference triads and the MST. Those for base-pairs $i+1$ and i are represented by thick and thin lines, respectively, and the MST is represented by broken lines.

The procedure used in SCHNAaP to define the MST and the local CEHS base-step parameters is as follows (Figures 3 and 4).

(1) We first of all calculate the angle RollTilt (Γ), which is defined as the magnitude of the angle between $\hat{z}_i$ and $\hat{z}_{i+1}$:

$$\Gamma = \cos^{-1}(\hat{z}_i \cdot \hat{z}_{i+1}) \qquad (1)$$

(2) Next we determine the Roll-Tilt axis (rt), which is given by:

$$rt = \hat{z}_i \times \hat{z}_{i+1} \qquad (2)$$

rt is then normalized to give $\widehat{rt}$.

(3) We then rotate base-pair i by $+\Gamma/2$ about the Roll-Tilt axis, and base-pair $i+1$ by $-\Gamma/2$ about the Roll-Tilt axis to the transformed orientation matrices $\mathbf{T}_i'$ and $\mathbf{T}_{i+1}'$ (Figure 3):

$$T_i' = R_{\mathrm{rt}}(+\Gamma/2)T_i, \quad T_{i+1}' = R_{\mathrm{rt}}(-\Gamma/2)T_{i+1} \qquad (3)$$

Here $\mathbf{R}_{\mathrm{rt}}$ (θ) refers to an orthogonal matrix that rotates the reference frame about the Roll-Tilt axis through an angle θ.

The x-y planes of the transformed base-pairs are now precisely parallel with each other, and their (coincident) z-axes coincide with the MST z-axis. The x (and y) axis of the MST lies along the bisector of the angle between the x (and y) axes of the transformed ith and $(i+1)$th base-pairs. Thus the directions of the MST axes, $\mathbf{T}_{\mathrm{mst}}$, are obtained by averaging and normalizing the two base-pair triads, $\mathbf{T}_i'$ and $\mathbf{T}_{i+1}'$ (Figure 4). The position of the MST origin is defined by:

$$r_{\mathrm{mst}} = (r_i + r_{i+1})/2 \qquad (4)$$

(4) Twist, Ω, is the angle between the two transformed y (or x) axes of the base-pairs (Figure 4). Its magnitude is given by:

$$\Omega = \cos^{-1}(\hat{y}'_i \cdot \hat{y}'_{i+1})$$

$$\text{If } (\hat{y}'_i \times \hat{y}'_{i+1}) \cdot \hat{z}_{\text{mst}} > 0, \quad \Omega > 0$$

$$\text{If } (\hat{y}'_i \times \hat{y}'_{i+1}) \cdot \hat{z}_{\text{mst}} < 0, \quad \Omega < 0 \qquad (5)$$

This ''sign-control'' mechanism is generally applicable to other situations where the sign of an angle is important (for example, the ϕ angle below), and is used in SCHNAaP for defining the signs of all angular parameters (Stephenson, 1973; Lavery & Sklenar, 1989).

(5) The angle between the Roll-tilt axis and the MST y-axis is ϕ (Figures 2 and 4):

$$\phi = \cos^{-1}(\widehat{rt} \cdot \hat{y}_{\text{mst}})$$

$$\text{If } (\widehat{rt} \times \hat{y}_{\text{mst}}) \cdot \hat{z}_{\text{mst}} > 0, \quad \phi > 0$$

$$\text{If } (\widehat{rt} \times \hat{y}_{\text{mst}}) \cdot \hat{z}_{\text{mst}} < 0, \quad \phi < 0 \qquad (6)$$

ϕ lies in the range of $[-180^\circ \rightarrow +180^\circ]$.

(6) Roll and Tilt, which are defined as the components of RollTilt along the y and x-axes of the MST, respectively, are given by the following formulae (Figure 2):

$$\rho = \Gamma\cos(\phi), \quad \tau = \Gamma\sin(\phi) \qquad (7)$$

(7) Shift, Slide and Rise, which are defined as the components of the relative displacement of the two base-pair triads along the x-, y- and z-axes of the MST, respectively, are given by:

$$[D_x D_y D_z] = (r_{i+1} - r_i)T_{\text{mst}} \qquad (8)$$

By using the concept of the MST, the CEHS scheme guarantees that the same numerical value of the parameters can be obtained, regardless of whether the dinucleotide step is reckoned along the $5' \rightarrow 3'$ direction of strand I or strand II. However, the signs of the x-axis parameters, Tilt and Shift, are sensitive to the direction in which the step is analyzed. As pointed out by El Hassan & Calladine (1995), this causes ambiguities for these parameters in the sequence-symmetric steps, GC/GC, CG/CG, AT/AT and TA/TA. In principle, there are four different sign combinations for Tilt and Shift (Table 2). If we consider two physically identical base-pairs, there are only two different types of step geometry: cases 1 and 2 are identical, as are cases 3 and 4. However, for a real structure where the two base-pairs usually have different conformations and sequence environments, all four cases in Table 2 are different. The signs of these parameters are therefore preserved in the SCHNAaP output. If the labeling of strand I and strand II is reversed (in practice, if the order in which they appear in the input file is reversed), then the signs of Tilt and Shift will be reversed for all steps.

Table 2. Four possible sign combinations for Tilt and Shift

	1	2	3	4
Tilt	+	−	+	−
Shift	+	−	−	+

Determination of the base-pair parameters

The procedures described for the above base-step parameters are directly applicable to the analysis of the base-pair geometry. The two rotations, Buckle and Opening, about the x and z-axes of the mid-base-triad (MBT) respectively, are combined in the same way as Roll and Tilt. The rotation angle is BuckleOpening (γ), and the rotation axis is the Buckle-Opening axis, which lies in the MBT x-z plane at an angle ϕ′ to the MBT x-axis (Figure 5). The analysis procedure described below first determines the values of γ and ϕ′, and then decomposes these into values of Buckle and Opening.

The base-pair i, the base on strand I is denoted iI, and the base on strand II, iII.

(1) We first calculate BuckleOpening (γ), which is defined as the magnitude of the angle between $\hat{y}_{i\text{II}}$ and $\hat{y}_{i\text{I}}$:

$$\gamma = \cos^{-1}(\hat{y}_{i\text{II}} \cdot \hat{y}_{i\text{I}}) \qquad (9)$$

(2) Next we determine the Buckle-Opening axis (bo) which is given by:

$$bo = \hat{y}_{i\text{II}} \times \hat{y}_{i\text{I}} \qquad (10)$$

bo is then normalized to give $\widehat{bo}$

(3) We then rotate base iII by $+\gamma/2$ about the Buckle-Opening axis, and base iI by $-\gamma/2$ about the Buckle-Opening axis to give the transformed orientation matrices $\mathbf{T}'_{i\text{II}}$ and $\mathbf{T}'_{i\text{I}}$:

$$T'_{i\text{II}} = R_{bo}(+\gamma/2)T_{i\text{II}}, \quad T'_{i\text{I}} = R_{bo}(-\gamma/2)T_{i\text{I}} \qquad (11)$$

The x-z planes of the transformed bases are now precisely parallel with each other, and their (coincident) y-axes coincide with the MBT y-axis. The

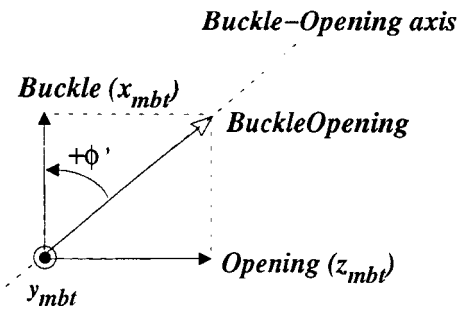

Figure 5. Buckle/Opening vectorial addition and decomposition.

x (and z) axis of the MBT lies along the bisector of the angle between the x (and z) axes of the transformed iI and iII bases. Thus the directions of the MBT axes, $\mathbf{T}_{\mathrm{mbt}i}$, are obtained by averaging and normalizing the two base triads, $\mathbf{T}'_{i\mathrm{II}}$ and $\mathbf{T}'_{i\mathrm{I}}$.

$$r_{\mathrm{mbt}} = (r_{iI} + r_{iI})/2 \tag{12}$$

(4) Propeller, ω, is the angle between the two transformed x (or z) axes of the bases. Its magnitude is given by:

$$\omega = \cos^{-1}(\hat{x}'_{iII} \cdot \hat{x}'_{iI})$$

$$\text{If } (\hat{x}'_{iII} \times \hat{x}'_{iI}) \cdot \hat{y}_{\mathrm{mbt}} > 0, \quad \omega > 0$$

$$\text{If } (\hat{x}'_{iII} \times \hat{x}'_{iI}) \cdot \hat{y}_{\mathrm{mbt}} < 0 \quad \omega < 0 \tag{13}$$

(5) The angle between the Buckle-Opening axis and the MBT x-axis is ϕ':

$$\phi' - \cos^{-1}(\widehat{bo} \cdot \hat{x}_{\mathrm{mbt}})$$

$$\text{If } (\widehat{bo} \times \hat{x}_{\mathrm{mbt}}) \cdot \hat{y}_{\mathrm{mbt}} > 0, \quad \phi' > 0$$

$$\text{If } (\widehat{bo} \times \hat{x}_{\mathrm{mbt}}) \cdot \hat{y}_{\mathrm{mbt}} < 0, \quad \phi' < 0 \tag{14}$$

(6) Buckle and Opening are calculated by the following formula:

$$\kappa = \gamma \cos(\phi'), \quad \sigma = \gamma \sin(\phi') \tag{15}$$

(7) The displacement parameters are obtained by:

$$[S_x S_y S_z] = (r_{iI} - r_{iII}) T_{\mathrm{mbt}} \tag{16}$$

Just as for Tilt and Shift, Buckle and Shear are sensitive to the way in which the base-pair is reckoned, and their signs will be reversed if the labeling of strands I and II is reversed. The values of Buckle given by SCHNAaP and NewHelix (Fratini *et al.*, 1982; Dickerson, 1985; El Hassan & Calladine, 1995, and NewHel93 User's Manual) have opposite signs, but SCHNAaP conforms with the sign convention recommended by the Cambridge Accord.

Global helical parameters

In addition to the local parameters that provide information on the fine detail of the base-stacking geometries, global helical parameters, which describe the geometry of the base-pairs relative to a global reference frame, may be useful (Lavery & Sklenar, 1989, 1988; Bhattacharyya & Bansal, 1989). We have therefore developed a mathematically rigorous set of global parameters using the definitions recommended by the Cambridge Accord (Table 1). We use a similar approach to that discussed above for the local parameters. Thus Tip and Inclination are combined in the same way as Roll and Tilt (Figure 6). The new set of global parameters is

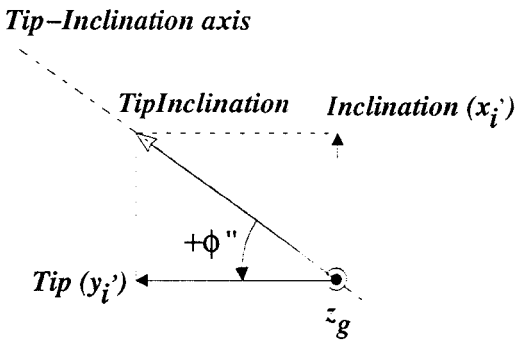

Figure 6. Tip/Inclination vectorial addition and decomposition.

defined relative to a global reference frame which is determined as follows.

(1) The global z-axis is the ''best-fit'' helical axis calculated using the algorithm developed by Rosenberg *et al.* (1976). The vectors used to define the helical axis are a combination of C'_1 and RN_9/YN_1 equivalent atom pairs along the same strand, as recommended by Dickerson (NewHel93 User's Manual).
(2) The global x and y-axes are given by the x and y-axes of the first base-pair's reference triad, respectively, after its z-axis has been aligned with the global z-axis.
(3) The x and y coordinates of the origin are determined as described (Rosenberg *et al.*, 1976). The z coordinate is defined as the midpoint of the RC_8-YC_6 line of the first base-pair.

Denoting the unit vector along the global helical z-axis by $\hat{z}_g$, we can write down the analysis procedure as follows.
(1) We first calculate TipInclination (Λ), which is defined as the magnitude of the angle between $\hat{z}_g$ and $\hat{z}_i$:

$$\Lambda = \cos^{-1}(\hat{z}_g \cdot \hat{z}_i) \tag{17}$$

(2) Next we determine the Tip-Inclination axis (ti), which is given by:

$$ti = \hat{z}_g \times \hat{z}_i \tag{18}$$

ti is then normalized to give $\widehat{ti}$.
(3) We then rotate base-pair i by $-\Lambda$ about the Tip-Inclination axis to give the transformed orientation matrix $\mathbf{T}'_i$. This aligns its z-axis with the global z-axis, and hence its x-y plane is now perfectly parallel with the global x-y dyad.

$$T'_i = R_{ti}(-\Lambda) T_i \tag{19}$$

Note that since we have defined the global x and y-axes as the x and y-axes of the first base-pair after its z-axis has been aligned with the global z-axis, it follows that the direction cosines of the global triad are simply given by the transformed orientation matrix of the first base-pair, i.e. $\mathbf{T}'_1$.

(4) The angle between the Tip-Inclination axis and y'_i is ϕ'' (Figure 6).

$$\phi'' - \cos^{-1}(\widehat{ti} \cdot \hat{y}'_i)$$

$$\text{If } (\widehat{ti} \times \hat{y}'_i) \cdot \hat{z}_g > 0, \quad \phi'' > 0$$

$$\text{If } (\widehat{ti} \times \hat{y}'_i) \cdot \hat{z}_g < 0, \quad \phi'' < 0 \qquad (20)$$

(5) Tip and Inclination are given by the following formulae:

$$\theta = \Lambda \cos(\phi''), \quad \eta = \Lambda \sin(\phi'') \qquad (21)$$

(6) The global Twist, Ω_g, is the angle between the successive y (or x) axes of the transformed base-pair triads:

$$\Omega_g = \cos^{-1}(\hat{y}'_i \cdot \hat{y}'_{i+1})$$

$$\text{If } (\hat{y}'_i \times \hat{y}'_{i+1}) \cdot \hat{z}_g > 0, \quad \Omega_g > 0$$

$$\text{If } (\hat{y}'_i \times \hat{y}'_{i+1}) \cdot \hat{z}_g < 0, \quad \Omega_g < 0 \qquad (22)$$

(7) The displacement parameters are obtained by:

$$[d_x d_y d_z] = r_i T'_i \qquad (23)$$

(8) the global Rise, Rise_g, is the difference between successive z-displacements:

$$\text{Rise}_g = d_{z_{i+1}} - d_{z_i} \qquad (24)$$

In this scheme, the base-pair reference triad is related to the global reference triad in an analogous way to that in which the $(i+1)$th base-pair triad is related to the MST in the CEHS scheme. It is important to note that in calculating the global helical parameters, we first rotate each base-pair until the z-axis of its reference triad is perfectly aligned with the global helix axis. The angle through which we rotate each base-pair to bring about this alignment is the TipInclination, which we define to be a ''vector-sum'' of Tilt and Inclination. We then proceed to calculate the global Twist, ϕ'', etc. This is in contrast to a directly orthogonal projection of the RC_8-YC_6 vector as in NewHelix and the scheme of von Kitzing & Diekmann (1987). The advantage of doing so is that it ensures exact and straightforward rebuilding of the molecular structure in question starting from the global parameters deduced from this analysis procedure (see the companion paper, Lu *et al.*, 1997). Table 3 gives a comparison between the global parameters calculated using this new method and those from NewHelix for three oligomers (El Hassan & Calladine, 1995). These are the *B*-form dodecamer d(CGCAAAAAAGCG) (Nelson *et al.*, 1987); the *B*-form dodecamer d(CGATCGATCG) (Grzeskowiak *et al.*, 1991); and the *A*-form octamer d(GGGGCCCC) (McCall *et al.*, 1985). In fact, the results are very similar, because the angle between the directly projected RC_8-YC_6 vector and the one after rotation is very small.

Table 3. Root-mean-square (RMS) differences between the global parameters as given by NewHelix and the present work for the three oligomers

	Dodecamer	Decamer	Octamer
Inclination	0.88	0.38	1.12
Tip	0.02	0.00	0.06
Twist_g	0.27	0.05	0.40
x-displacement	0.00	0.00	0.01
y-displacement	0.00	0.00	0.03
Rise_g	0.00	0.00	0.00

Translational parameters are in Å and rotational parameters in degrees

Classification of structure polymorphs

DNA structures are traditionally classified into *A*, *B* and *Z*-forms, as in the Brookhaven Protein Data Bank (PDB; Bernstein *et al.*, 1997) and the Nucleic Acid Data Base (NDB; Berman *et al.*, 1992). Given that oligonucleotide structures often do not strictly satisfy all of the criteria for a single polymorph, such classification can be ambiguous (Calladine & Drew, 1992). Within one polymorphic family great structural variability is possible (Kennard & Hunter, 1991; Dickerson, 1992), but there are some structural parameters that clearly distinguish between families. *Z*-DNA is usually distinguished from *A* and *B*-DNA by the sense of the helical Twist, negative for *Z*-DNA and positive for *A* and *B*-DNA. However, a recently proposed left-handed structure, *W*-DNA (Ansevin & Wang, 1990; Dickerson, 1992; Figure 7), has negative helical Twist, yet is structurally very different from *Z*-DNA. In order to unambiguously distinguish between these different classes of double-helix, some new criteria are required.

Consideration of the structural differences between *A*/*B*, *Z* and *W*-DNA reveals that there are in principle four possible arrangements for antiparallel nucleic acid duplexes (Figure 7; Watson & Crick, 1953; Zhurkin *et al.*, 1978; Olson *et al.*, 1982; Hopkins, 1983). If a double-helix is viewed with the $5' \rightarrow 3'$ direction of strand I pointing upwards and to the left of strand II, the bases can be arranged so that either the major groove or minor groove faces the viewer. This is the essential difference between the two left-handed *Z* and *W*-DNA structures. It is possible to have the right-handed versions of these two structures: one is the *A*/*B*-DNA structure and the other is as yet uncharacterized experimentally (Hopkins, 1981). We propose to call this fourth polymorph *R*-DNA : Right-handed but with Reversed base orientation as compared to *A*/*B*-DNA. These four types of structure can now readily be distinguished by the sense of the helical twist and the orientation of the bases relative to the sugar-phosphate backbone.

We use the procedure described above to calculate the base and base-pair triads for all cases. The value of Twist distinguishes left-handed *Z*- or *W*-DNA (negative Twist) from right-handed *A*/*B* or *R*-DNA (positive Twist) structures. The direction

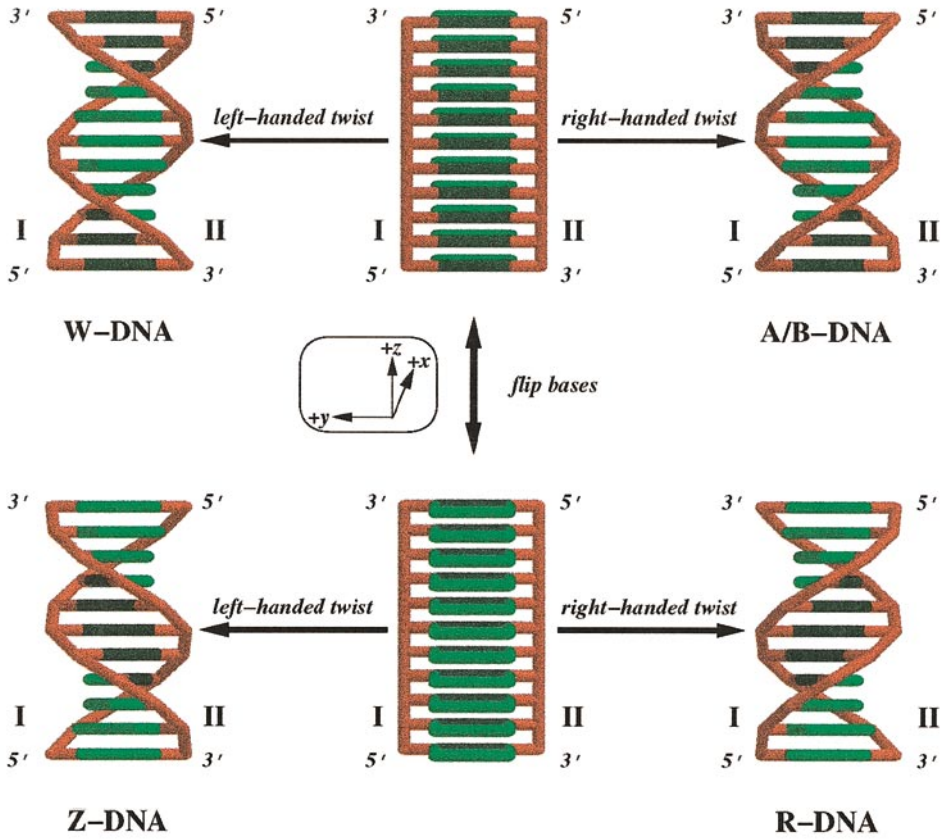

Figure 7. A representation of four possible arrangements for antiparallel nucleic acid duplexes. Left-handed *W* and *Z*-DNA are shown on the left (the characteristic zig-zag backbone pattern is not represented for simplicity). Right-handed *A*/*B* and hypothetical *R*-DNA are shown on the right. The Twist free ladder forms are shown in the middle column. In the top row, the minor groove faces the viewer, while in the bottom row, the major groove faces the viewer. The SCHNAaP coordinate system is also shown. These structures were generated using SCHNArP (see accompanying paper) with Twist = ± 36° (0° for the ladder forms), Rise = 3.34 Å, and all other step parameters are set to zero. Color scheme: the minor groove side, dark green; the major groove side, light green; and the backbone, red.

of the *x*-axis of the base-pair triad gives the orientation of the bases relative to the backbone: it can point towards the major groove side as in *A*/*B* and *W*-DNA or towards the minor groove side as in *R* and *Z*-DNA. This direction is determined using the dot product of the $C'_1 \rightarrow RN_9$ (or $C'_1 \rightarrow YN_1$) vector with the base-pair triad *x*-axis: positive for *A*/*B* and *W*-DNA; negative for *R* and *Z*-DNA. In this scheme, the sense of many of the parameters calculated for the two kinds of base-backbone orientation differs. For example, positive Roll in *A*/*B*-DNA and *W*-DNA corresponds to the opening-up of the minor groove, whereas for *R*-DNA and *Z*-DNA, positive Roll means opening-up of the major groove. Similarly, negative *x*-displacement in *A*/*B* and *W*-DNA means that the helical axis passes closer to the major groove side, whereas negative *x*-displacement in *R* and *Z*-DNA means that the helical axis passes closer to the minor groove side.

In order to distinguish *A* and *B*-DNA, a further set of criteria is required. The sugar conformations, C'_3-*endo* for *A*-DNA and C'_2-*endo* for *B*-DNA as traditionally claimed, are generally poor discriminators because of their flexibility, especially in *B*-DNA (Dickerson, 1988). Among the six local base-step parameters, Slide and Roll have the most discrimination power (Calladine & Drew, 1984; Gorin *et al.*, 1995), but there is no clear distinction between *A* and *B*-DNA with regard to either parameter. The position of the phosphorous atom relative to the base-pair (Calladine & Drew, 1984) and the global parameter, *x*-displacement (Dickerson, 1988), however, are excellent criteria for distinguishing between these two structural families. In SCHNAaP, oligonucleotide structures are thus classified into *A* or *B*-DNA according to the average *x*-displacement and Z_p, the *z* coordinate of the phosphorous atom with respect to the MST (El Hassan, 1993). Among the 60 oligomers analyzed (El Hassan & Calladine, 1996), the two distinct clusters in the average Z_p *versus* *x*-displacement scatter plot in Figure 8 illustrate this clearly.

Base-step stacking diagrams

Base-step stacking diagrams allow a qualitative visual assessment of the stacking interactions

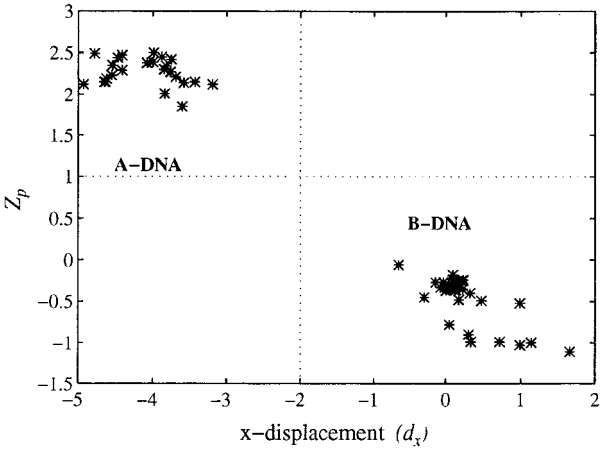

Figure 8. Scatter plot of average Z_p *versus* *x*-displacement for 60 oligomers. Broken lines mark the criteria used in SCHNAaP to discriminate between *A* and *B* helices

between successive bases, and they appear frequently in the literature on DNA structures. However, different reference frames are used by different authors, and this can be somewhat misleading. The base-stacking may be viewed in orientations that are: (i) perpendicular to the mean plane through the lower base-pair (Hunter *et al.*, 1989; Nunn & Neidle, 1996); (ii) perpendicular to the mean plane through the upper base-pair (Portmann *et al.*, 1995); (iii) perpendicular to the mean plane through both base-pairs (Malinina *et al.*, 1994); or (iv) along the global helical axis (Prive *et al.*, 1991). Each one of these methods gives a different view of the same step. The MST concept seems well suited for setting up ''standardized'' base-step stacking diagrams. If the base-step is viewed relative to the MST reference frame, the resulting stacking diagram is uniquely defined, local to the dinucleotide step and independent of whether the view is from the upper or lower base-pair. The central GC step in the octamer d(GGGGCCCC) (McCall *et al.*, 1985) is used as an example in Figure 9, and its local base-step and base-pair parameters are given in Table 4. This step has a large positive Roll angle, and so the appearance of the stacking geometries in Figure 9 (a) and (b), and (c) and (d) which are oriented with respect to the planes of the lower and upper base-pairs, respectively, are quite different. In contrast, the stacking geometries in Figure 9 (e) and (f) are

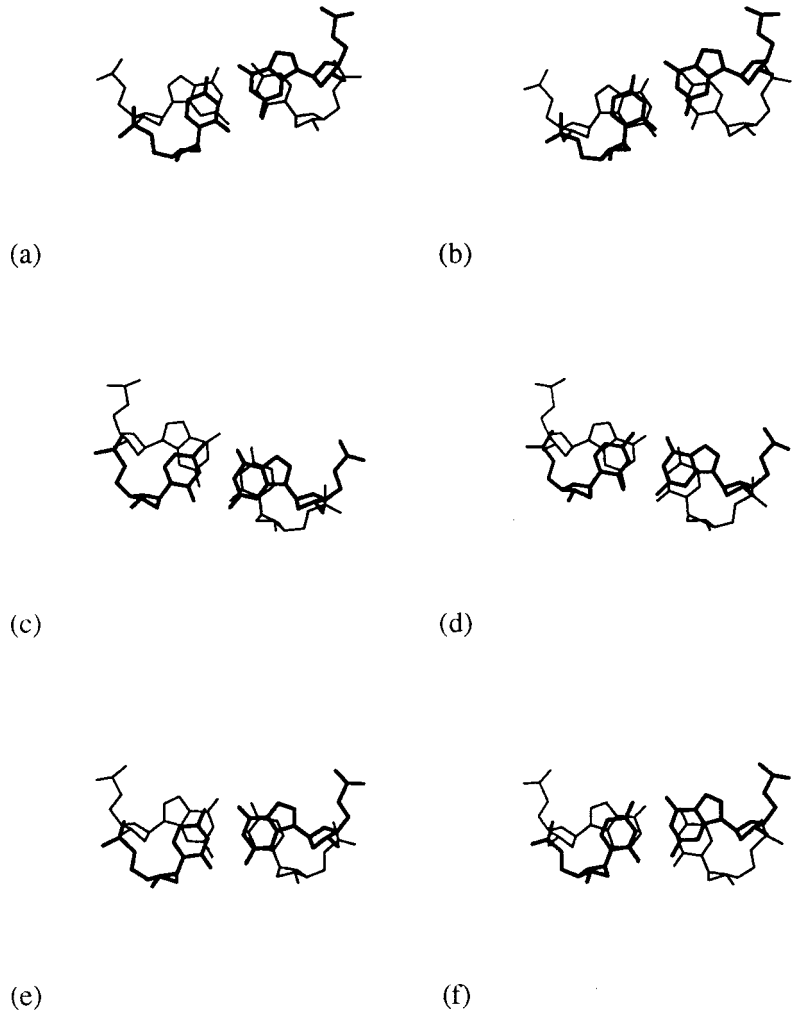

Figure 9. Base-step stacking diagrams for the central GC step (base-pairs 4 and 5) in the octamer d(GGGGCCCC) (McCall *et al.*, 1985; Table 4). Base-pairs drawn with thick lines are stacked above base-pairs drawn with thin lines. The left column corresponds to the strand designation given in the PDB and NDB entry, and in the right column these two strands are reversed. In (a) and (b) the step is viewed with respect to the lower base-pair; in (c) and (d) viewed with respect to the upper base-pair; and in (e) and (f) viewed with respect to the MST.

Table 4. CEHS base-pair and base-step parameters for the central GC step (base-pairs 4 and 5) in the octamer d(GGGGCCCC) (McCall *et al*., 1985)

	bp	Propeller	Opening	Buckle	Stretch	Stagger	Shear
4	G-C	−4.07	2.34	−4.58	5.43	−0.05	0.13
5	C-G	−11.90	0.87	8.62	5.52	−0.16	0.00
	Step	Twist	Roll	Tilt	Rise	Slide	Shift
	GC/GC	31.03	11.14	−0.69	3.50	−0.77	−0.47

Translational parameters are in Å and rotational parameters in degrees

identical because the base-step is oriented with reference to the MST.

Other features of the program

In order to make SCHNAaP as widely applicable as possible, we have devised schemes to handle modified and mispaired bases. Chemical modifications do not pose a problem as long as they do not change the basic chemical composition of the core pyrimidine and purine ring structures. Thus, the program handles bases with additional substituents such as methylcytosine, and bases with missing substituents such as inosine. The point here is that such modifications do not affect the atoms used for defining the various reference frames used by SCHNAaP.

CEHS was originally designed to deal with DNA involving only standard Watson-Crick purine-pyrimidine base-pairs, but SCHNAaP can also handle base-pairs. For purine-pyrimidine mispairs, the *y*-axis is defined by the RC_8-YC_6 line; for purine-purine mispairs, the *y*-axis is defined by the RC_8-RC_8 line; and for pyrimidine-pyrimidine mispairs, the *y*-axis is defined by the YC_6-YC_6 line. The identification of a base as purine or pyrimidine is based on the presence or absence of an N_9 atom, respectively. The other axes of the base and base-pair reference triads are defined as in the standard Watson-Crick base-pairs. It is clear that base-pair parameters for such mispaired bases will not have the same physical meaning as they would in standard Watson-Crick base-pairs.

The sugar-phosphate backbone conformation is described by means of a large number of internal torsion angles, groove-widths, etc. (Lu, 1996; Saenger, 1984). These are also calculated by SCHNAaP. The list of backbone parameters is given below.

(i) Main-chain torsion angles $\alpha \rightarrow \zeta$ and the glycosyl torsion angle χ.

(ii) Endocyclic torsion angles of the sugar, $\upsilon_0 \rightarrow \upsilon_5$, as well as the amplitude τ_m and the phase angle *P* of pseudorotation of the sugar ring (Altona & Sundaraligam, 1972).

(iii) Polar coordinates of the phosphorus atoms, C_1' atoms, and O_4' atoms with respect to the global reference triad.

(iv) The angles, defined as those between C_1'-YN_1/C_1'-RN_9 and the C_1'-C_1' line for each base-pair, as well as distances between C_1'-C_1' and RC_8-YC_6 within each pair.

(v) Groove widths are measured by the O_4'-O_4' and P-P distances reduced by 2.8 Å and 5.8 Å, respectively, to allow for the van der Waals radii of oxygen and phosphate.

(vi) The *xyz* coordinates of the phosphorus atoms in each base-step with respect to the MST.

In addition to short oligonucleotide fragments, SCHNAaP can be applied to DNA-protein complexes. Since DNAs in such complexes are normally strongly curved, or even reverse their helix axes (Rice *et al*., 1996; Rice, 1997), the global helical parameters based on an overall straight axis are not meaningful. Under such circumstances, the local CEHS parameters can be used. As an example, the local step parameters for a 35 base-pair DNA in the IHF-DNA complex (Rice *et al*., 1996; PDB code: 1IHF) calculated by SCHNAaP is shown in Figure 10. For comparison, the same set of local parameters from Curves (Lavery & Sklenar, 1988, 1989; version 5.1) and the program of Babcock *et al*. (1994) are also given. It is clear that for Twist and Roll, these three methods give quite similar results. There are, however, noticeable differences for Rise in the AA kink region (TCAATT). SCHNAaP distributes the larger rise over the constituent steps. Curves, however, shows extremely large fluctuations. It is worth noting that for the AA step (the kink), Curves gives an unreasonably large value of Rise of 8.49 Å, and for TC it is as low as 2.38 Å. The Babcock *et al*. (1994) method gives a result in between. As regard to Slide, SCHNAaP and the Babcock *et al*. (1994) method match quite well, while Curves gives a uniformly smaller value. For the six A·T base-pairs (A-tract) region, which is characteristic of *B*-form DNA, SCHNAaP, the Babcock *et al*. (1994) method and Curves give an average Slide value of −0.41 Å, −0.38 Å and −0.84 Å, respectively.

Input and output formats

On entry, SCHNAaP needs an input file (filename.inp) containing the atomic coordinates of the structure to be analyzed. The input file is in ''standard'' PDB format (Bernstein *et al*., 1977). The structure must be an antiparallel duplex and each strand must have the same number of bases, which should be numbered in the $5' \rightarrow 3'$ direction. The designation of the two strands is defined by their relative position in the input file: strand I comes before strand II. SCHNAaP can handle normal

PDB/NDB oligomers of the general form CpGp . . . pT (without a leading phosphate group), as well as structures of the form pCpG . . . pT (with a leading phosphate group):

On exit, SCHNAaP generates the following six output files.

(i) filename.out is the main output file and contains a detailed listing of all the parameters discussed here.

(ii) filename.all is the double-helical structure oriented with respect to the global reference frame (PDB format).

(iii) filename.bp is the same as filename.all, except that only the base-pair atoms are included (PDB format).

(iv) filename.mst is a multi-structure file that contains each dinucleotide step oriented with reference to its MST (PDB format).

(v) filename.ceh is a listing of the CEHS base-pair and base-step parameters in a format that can be read directly by SCHNArP (see accompanying paper, Lu *et al.*, 1997).

(vi) filename.glh is a listing of the global helical parameters together with the CEHS base-pair parameters, which can also be used as input to SCHNArP.

Conclusions

SCHNAaP is a mathematically rigorous program for analyzing double-helical nucleic acid structures.

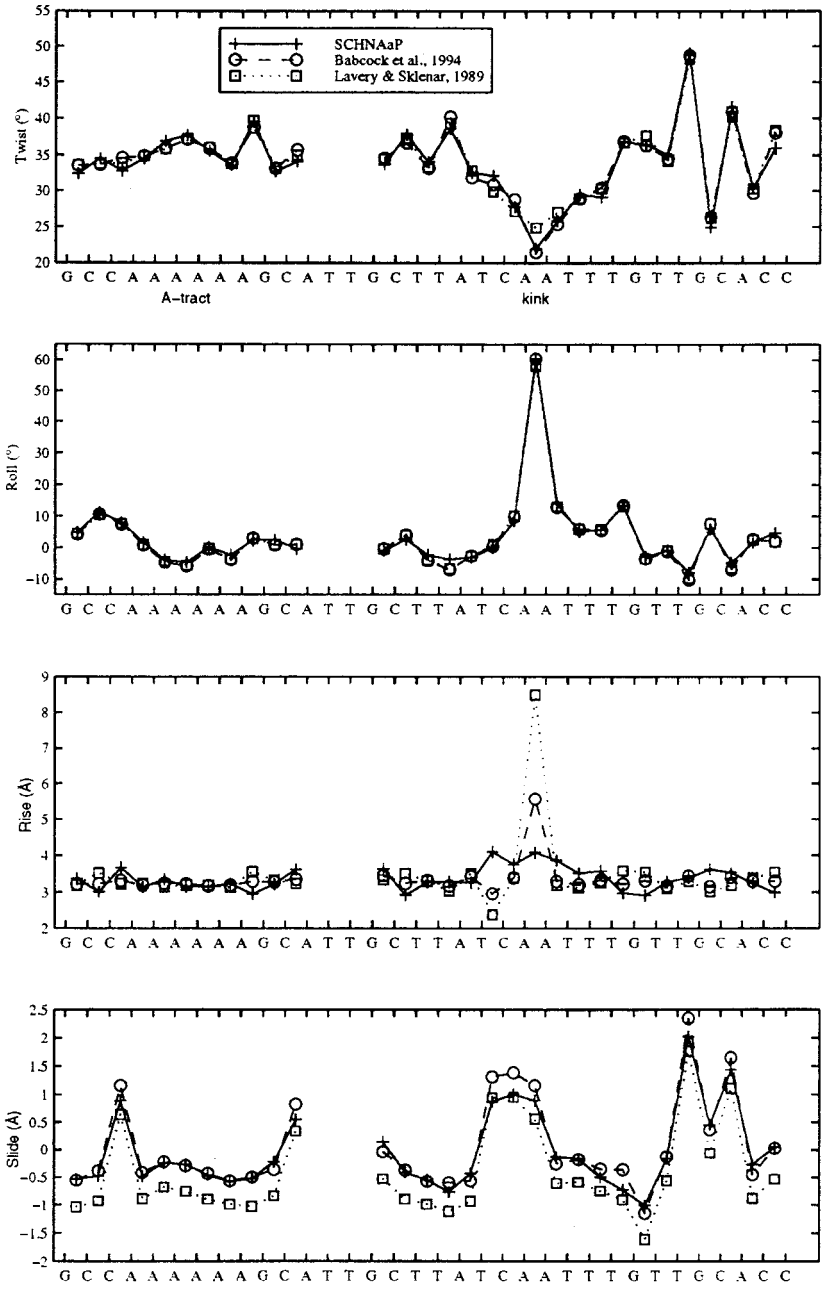

Figure 10. Comparisons of local step parameters Twist, Roll, Rise and Slide for DNA in the IHF-DNA complex of Rice *et al.* (1996) as given by SCHNAaP, the Babcock *et al.* (1994) method and Curves. Two T·A base-pairs in the middle-left were omitted from the calculations due to their non-Watson-Crick conformations.

In this scheme, both the local base-step and base-pair and the global helical parameters are handled in a very simple and consistent fashion. These parameters form a complete set that fully conforms with the Cambridge Accord. The automatic classification of the structure as an *A*, *B*, *Z*, *W* or *R* polymorph, and the generation of the ''standardized'' base-stacking diagrams are all unique to the program. Moreover, the reversibility of the scheme renders an exact reconstruction of a duplex structure given a set of either local or global parameters as demonstrated in the accompanying paper (Lu *et al*., 1997). This software will thus provide a useful tool for studying various structural features of helical nucleic acids. The program is coded in MATLAB and C, and is available from the authors upon request (X.LU1@sheffield.ac.uk, MAE@eng.cam.ac.uk, C.Hunter@sheffield.ac.uk).

Acknowledgements

We thank the CVCP and the University of Sheffield (X.J.L.), Peterhouse College, Cambridge (M.A.E.) and the Lister Institute (C.A.H.) for financial support. We also thank Chris Calladine for his many helpful suggestions and comments.

References

Altona, C. & Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212.

Ansevin, A. T. & Wang, A. H. (1990). Evidence for a new *Z*-type left-handed DNA helix: properties of *Z*(WC)-DNA. *Nucl. Acids Res.* **18**, 6119–6126.

Babcock, M. S. & Olson, W. K. (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., Pednault, E. P. D. & Olson, W. K. (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.

Bansal, M., Bhattacharyya, D. & Ravi, B. (1995). NUPARM and NUCGEN: software for analysis and generation of sequence nucleic acid structures. *Comput. Appl. Biosci.* **11**, 281–287.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database: a comprehensive relational database of three dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bhattacharyya, D. & Bansal, M. (1989). A self-consistent formulation for analysis and generation of non-uniform DNA structures. *J. Biomol. Struct. Dynam.* **6**, 635–653.

Blow, D. M. (1960). To fit a plane to a set of points by least squares. *Acta Crystallog.* **13**, 168.

Calladine, C. R. & Drew, H. R. (1984). A base centered explanation of the *B*- to *A*-transition in DNA. *J. Mol. Biol.* **178**, 773–781.

Calladine, C. R. & Drew, H. R. (1992). *Understanding DNA; The Molecule & How It Works*. Academic Press, London.

Dickerson, R. E. (1985). Helix Comparison Table. In *Biological Micromolecules and Assemblies* (Jurnak, F. A. & McPherson, A., eds), vol. 2, Appendix, pp. 471–494, Wiley-Interscience, New York.

Dickerson, R. E. (1988). Usual and unusual DNA structure: a summing up. In *Unusual DNA Structures* (Wells, R. D. & Harvey, S. C., eds), pp. 287–306, Springer-Verlag, New York.

Dickerson, R. E. (1992). DNA structure from *A* to *Z*. *Methods Enzymol.* **211**, 67–111.

Diekmann, S. (1989). Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* **205**, 787–791.

El Hassan, M. A. (1995). The geometry and structure of DNA and its role in DNA/protein recognition. PhD thesis, University of Cambridge, UK.

El Hassan, M. A. & Calladine, C. R. (1995). The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *J. Mol. Biol.* **251**, 648–664.

El Hassan, M. A. & Calladine, C. R. (1996). Propeller-twisting of case-pairs and conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.* **259**, 95–103.

Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a *B*-DNA dodecamer—CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 4686–4707.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34–48.

Grzeslowiak, K., Yanagi, K., Prive, G. G. & Dickerson, R. E. (1991). The structure of *B*-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.* **266**, 8861–8883.

Hopkins, R. C. (1981). Deoxyribonucleic acid structure: a new model. *Science,* **211**, 289–291.

Hopkins, R. C. (1983). Transitions between *B*-DNA and *Z*-DNA: a dilemma. *J. Theoret. Biol.* **101**, 327–333.

Hunter, W. N., Destaintot, B. L. & Kennard, O. (1989). Structural variation in d(CTCTAGAG). implications for protein-DNA interactions. *Biochemistry,* **28**, 2444–2451.

Jursa, J. (1994). DNA modeler: an interactive program for modeling stacks of DNA base pairs on a microcomputer. *Comput. Appl. Biosci.* **10**, 61–65.

Kennard, O. & Hunter, W. N. (1991). Single-crystal X-ray diffraction studies of oligonucleotides and oligonucleotide-drug complexes. *Agnew. Chem. Int. Ed. Engl.* **30**, 1254–1277.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91.

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

Lu, X. J. (1996). Sequence-dependent DNA structure. PhD thesis, University of Sheffield, UK.
Lu, X. J., El Hassan, M. A. & Hunter, C. A. (1997). Structure and conformation of helical nucleic acids: rebuilding program (SCHNArP). *J. Mol. Biol.* **273**, 681–691.
Malinina, L., Urpi, L., Salas, X., Huynh-Dinh, T. & Subirana, J. A. (1994). Recombination-like structure of d(CCGCGG). *J. Mol. Biol.* **243**, 484–493.
Mazur, J. & Jernigan, R. L. (1995). Comparison of rotation models for describing DNA conformations: application to static and polymorphic forms. *Biophys. J.* **68**, 1472–1489.
McCall, M., Brown, T. & Kennard, O. (1985). The crystal structure of d(G-G-G-G-C-C-C-C): a model for poly(dG)·poly(dC). *J. Mol. Biol.* **183**, 385–396.
Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature,* **330**, 221–226.
Nunn, C. M. & Neidle, S. (1996). The high resolution crystal structure of the DNA decamer d(AGGCATGCCT). *J. Mol. Biol.* **256**, 340–351.
Olson, W. K. (1996). Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* **6**, 242–256.
Olson, W. K., Srinivasan, A. R., Marky, N. L. & Balaji, V. N. (1982). Theoretical probes of DNA conformation examining the $B \rightarrow Z$ conformational transition. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 229–241.
Portmann, S., Usman, N. & Egli, M. (1995). The crystal structure of r(CCCCGGGG) in two distinct lattices. *Biochemistry*, **34**, 7569–7575.
Prive, G. G., Yanagi, K. & Dickerson, R. E. (1991). Structure of the *B*-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.* **217**, 177–199.
Rice, P. A. (1997). Making DNA do a U-turn: IHF and related proteins. *Curr. Opin. Struct. Biol.* **7**, 86–93.
Rice, P. A., Yang, S. W., Mizuuchi, K. & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell,* **87**, 1295–1306.
Rosenberg, J. M., Seeman., N. C., Day, R. O. & Rich, A. (1976). RNA double helices generated from crystal structures of double helical dinucleotide phosphates. *Biochem. Biophys. Res. Commun.* **69**, 979–987.
Saenger, W. (1984). *Principles of Nucleic Acids Structure.* Springer-Verlag, New York.
Schomaker, V., Waser, J., Marsh, R. E. & Bergman, G. (1959). To fit a plane or a line to a set of points by least squares. *Acta Crystallog.* **12**, 600–604.
Stephenson, G. (1973). *Mathematical Methods for Science Students,* 2nd Edit., Longman Scientific & Technical, Harlow.
Tung, C. S., Soumpasis, D. M. & Hummer, G. (1994). An extension of the rigorous base-unit oriented description of the nucleic-acid structures. *J. Biomol. Struct. Dynam.* **11**, 1327–1344.
von Kitzing, E. & Diekmann, S. (1987). Molecular mechanics calculations of $dA_{12}\cdot dT_{12}$ and of the curved molecule $d(GCTCGAAAAA)_4\cdot d(TTTTTCGAGC)_4$. *Eur. Biophys. J.* **15**, 13–26.
Watson, J. D. & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature*, 171–738.
Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. (1978). Conformations of DNA as revealed by computer calculations. *Biopolymers,* **17**, 377–412.

Edited by K. Nagai

(Received 28 November 1996; received in revised form 16 July 1997; accepted 25 July 1997)

http://www.hbuk.co.uk/jmb

Supplementary material for this paper comprising the main output file from SCHNAaP for the dodecamer d(CGCAAAAAAGCG) (Nelson *et al.*, 1987) is available from JMB Online.

Appendix

Tables A1 and A2 give the local CEHS base-pair and base-step parameters, and the new set of global helical parameters from SCHNAaP for the dodecamer d(CGCAAAAAAGCG) (Nelson *et al.*, 1987).

Table A1. The local CEHS base-pair and base-step parameters for the dodecamer d(CGCAAAAAAGCG) (Nelson *et al.*, 1987)

	bp	Propeller	Opening	Buckle	Stretch	Stagger	Shear
1	C-G	−12.27	−6.75	17.81	5.52	−0.30	−1.01
2	G-C	−11.69	2.26	−4.91	5.34	0.01	−0.54
3	C-G	−3.17	0.19	−10.39	5.35	−0.13	−0.21
4	A-T	−15.07	5.89	5.62	5.15	−0.20	0.89
5	A-T	−23.46	7.15	3.82	5.43	−0.04	0.34
6	A-T	−26.14	7.50	0.51	5.14	−0.11	1.22
7	A-T	−20.96	3.75	−9.90	5.32	−0.51	−0.28
8	A-T	−16.39	6.29	−7.09	5.29	0.25	0.26
9	A-T	−17.13	1.09	−7.55	5.15	−0.47	1.07
10	G-C	−10.46	−3.18	−0.24	5.43	0.18	0.68
11	C-G	−14.54	−1.52	1.69	5.62	0.25	0.13
12	G-C	−3.91	−6.09	−5.72	5.35	0.04	−0.85
	Average	−14.60	1.38	−1.36	5.34	−0.09	0.14
	s.d.	6.99	4.99	8.06	0.15	0.25	0.74
	Step	Twist	Roll	Tilt	Rise	Slide	Shift
1	CG/CG	41.51	−0.49	−3.32	3.23	0.18	−0.18
2	GC/GC	38.12	−2.73	2.48	3.22	0.46	0.90
3	CA/TG	28.06	10.41	2.16	3.76	0.81	−0.54
4	AA/TT	36.19	1.23	−1.51	3.25	−0.09	−0.28
5	AA/TT	35.50	−0.60	−0.45	3.29	−0.04	−0.40
6	AA/TT	38.77	2.23	2.14	2.97	−0.76	0.12
7	AA/TT	32.97	0.37	−1.88	3.25	0.21	−0.14
8	AA/TT	37.89	−4.79	−0.37	3.19	−0.14	−0.23
9	AG/CT	33.50	4.29	−1.08	3.52	0.85	0.67
10	GC/GC	38.29	−10.09	−4.14	3.58	0.76	−1.13
11	CG/CG	41.39	8.91	2.73	3.21	0.48	0.16
	Average	36.56	0.79	−0.29	3.32	0.25	−0.09
	s.d.	3.95	5.82	2.40	0.22	0.49	0.56

Translational parameters are in Å and rotational parameters in degrees. s.d. is the standard deviation.

Table A2. The global helical parameters for the dodecamer d(CGCAAAAAAGCG) (Nelson *et al.*, 1987)

	bp	$Twist_g$	Tip	Inclination	$Rise_g$	*y*-disp.	*x*-disp.
1	C-G	41.18	2.20	27.11	2.97	4.98	−0.79
2	G-C	38.28	−15.52	18.54	3.33	3.39	1.71
3	C-G	28.56	−27.06	6.42	3.27	1.41	2.96
4	A-T	35.90	−17.29	−2.57	3.04	1.05	1.58
5	A-T	35.56	−10.83	−13.30	3.10	0.63	0.96
6	A-T	38.35	−1.52	−17.73	3.04	1.02	0.68
7	A-T	33.18	11.24	−12.07	3.11	0.33	1.59
8	A-T	37.95	16.90	−5.67	3.13	−0.09	2.35
9	A-T	33.88	12.41	4.01	3.36	−1.85	2.35
10	G-C	37.36	12.50	10.44	3.72	−2.85	2.43
11	C-G	42.08	−4.70	8.73	3.25	−3.45	−0.66
12	G-C	–	−1.97	9.10	–	−2.09	−2.79
	Average	36.57	−1.97	2.75	3.21	0.21	1.03
	s.d.	3.80	13.81	13.45	0.21	2.49	1.68

s.d. is the standard deviation, disp. the displacement.