doi:10.1006/jmbi.2001.4987 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **313**, 229–237

JMB

NOMENCLATURE

A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry

**Wilma K. Olson, Manju Bansal, Stephen K. Burley
Richard E. Dickerson, Mark Gerstein, Stephen C. Harvey
Udo Heinemann, Xiang-Jun Lu, Stephen Neidle, Zippora Shakked
Heinz Sklenar, Masashi Suzuki, Chang-Shung Tung, Eric Westhof
Cynthia Wolberger and Helen M. Berman**

Keywords: nucleic acid conformation; base-pair geometry; standard reference frame

A common point of reference is needed to describe the three-dimensional arrangements of bases and base-pairs in nucleic acid structures. The different standards used in computer programs created for this purpose give rise to conflicting interpretations of the same structure.[1] For example, parts of a structure that appear ''normal'' according to one computational scheme may be highly unusual according to another and *vice versa*. It is thus difficult to carry out comprehensive comparisons of nucleic acid structures and to pinpoint unique conformational features in individual structures. In order to resolve these issues, a group of researchers who create and use the different software packages have proposed the standard base reference frames outlined below for nucleic acid conformational analysis. The definitions build upon qualitative guidelines established previously to specify the arrangements of bases and base-pairs in DNA and RNA structures.[2] Base coordinates are derived from a survey of high-resolution crystal structures of nucleic acid analogs stored in the Cambridge Structural Database (CSD).[3] The coordinate frames are chosen so that complementary bases form an ideal, planar Watson-Crick base-pair in the undistorted reference state with hydrogen bond donor-acceptor distances, C1′···C1′ virtual lengths, and purine N9-C1′···C1′ and pyrimidine N1-C1′···C1′ virtual angles consistent with values observed in the crystal structures of relevant small molecules. Conformational analyses performed in this reference frame lead to interpretations of local helical structure that are essentially independent of computational scheme. A compilation of base-pair parameters from representative *A*-DNA, *B*-DNA, and protein-bound DNA structures from the Nucleic Acid Database (NDB)[4] provides useful guidelines for understanding other nucleic acid structures.

Base coordinates

Models of the five common bases (A, C, G, T and U) were generated from searches of the crystal structures of small molecular mass analogs, e.g. free bases, nucleosides and nucleotides, in the most recent version of the CSD.[3] The internal geometries and associated uncertainties in this data set closely match numerical values reported in the recent survey of nucleic acid base analogs by Clowney *et al*.[5] Because the minor changes in chemical structure have essentially no effect on either the ideal base-pair frame or the computed rigid-body parameters, the Clowney *et al*. bases are retained as standards.

Coordinate frame

The right-handed coordinate frame attached to each base (Figure 1) follows established qualitative guidelines.[2] The *x*-axis points in the direction of the major groove along what would be the pseudo-dyad axis of an ideal Watson-Crick base-pair, i.e. the perpendicular bisector of the C1′···C1′ vector spanning the base-pair. The *y*-axis runs

This is a document of the Nomenclature Committee of IUBMB (NC-IUBMB)/IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN), whose members are R. Cammack (chairman), A. Bairoch, H.M. Berman, S. Boyce, C.R. Cantor, K. Elliott, D. Horton, M. Kanehisa, A. Kotyk, G.P. Moss, N. Sharon and K.F. Tipton.

0022-2836/01/010229–9 $35.00/0

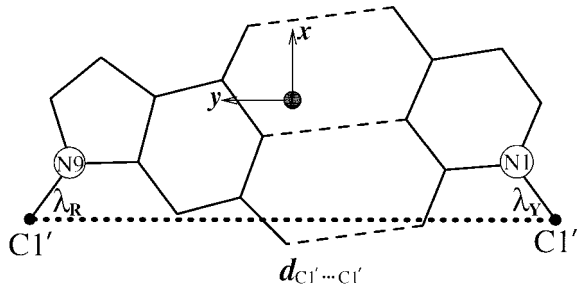

Figure 1. Illustration of idealized base-pair parameters, $d_{C1'\cdots C1'}$ and λ, used respectively to displace and pivot complementary bases in the optimization of the standard reference frame for right-handed *A* and *B*-DNA, with the origin at ● and the *x*- and *y*-axes pointing in the designated directions.

along the long axis of the idealized base-pair in the direction of the sequence strand, parallel with the C1′···C1′ vector, and displaced so as to pass through the intersection on the (pseudo-dyad) *x*-axis of the vector connecting the pyrimidine Y(C6) and purine R(C8) atoms. The *z*-axis is defined by the right-handed rule, i.e. $z = x \times y$. For right-handed *A* and *B*-DNA, the *z*-axis accordingly points along the 5′ to 3′ direction of the sequence strand.

The location of the origin depends upon the width of the idealized base-pair, i.e. the C1′···C1′ spacing, $d_{C1'\cdots C1'}$, and the pivoting of complementary bases, λ, in the base-pair plane (see Figure 1). The coordinates of the C1′ atoms establish the pseudo-dyad axis, i.e. the line in the base-pair plane where $y = 0$. The rotations of each base about a normal axis passing through the C1′ glycosyl atoms determine the Y(C6) and R(C8) positions used to define the line where $x = 0$.

Optimization

The atomic coordinates in Table 1 are expressed in the base-pair reference frames that optimize hydrogen-bond donor-acceptor distances, d_{HB}, and base ''pivot'' angles, λ_Y and λ_R, against corresponding standards ($d_0 = 3.0$ Å and $\lambda_0 = 54.5°$). The departures from ideality are measured by the sum of the absolute values of the relative deviations:

$$\left|\frac{\lambda_Y - \lambda_0}{\lambda_0}\right| + \left|\frac{\lambda_R - \lambda_0}{\lambda_0}\right| + \sum_{\{\text{H-bonds}\}} \left|\frac{d_{HB} - d_0}{d_0}\right|$$

where the last term runs over two (T·A) or three (C·G) hydrogen bonds. (Optimization in terms of the sum of the squares of the relative deviations of the γ_Y, γ_R, and d_{HB} yields similar results.)

Virtual distances and angles characterizing the optimized configurations are detailed in Table 2. The minor changes in chemical bonding between T *versus* C and A *versus* G in combination with the constraints of two or three hydrogen bonds, give rise to slightly different standard orientations of T·A and C·G base-pairs (compare $d_{C1'\cdots C1'}$, λ_Y, and λ_R values in Table 2). Notably, the hydrogen bonds closer to the minor groove edges of all base-pairs are shorter than those nearer the major groove edges, as is observed in high-resolution structures of Watson-Crick base-pair co-crystal complexes.[6,7] The hydrogen bonds are slightly shorter, on average, in the small molecule analogs, which are in turn distorted to a small degree from the perfectly planar base-pair geometry assumed here (see Wilson[8] and Table 2 for numerical values).

Minor changes in the imposed configurational constraints have almost no influence on the preferred base-pair arrangements, e.g. the increase of λ_0 from 54.5° to 55.5° shortens $d_{C1'\cdots C1'}$ by less than 0.1 Å and perturbs hydrogen bond lengths by less than 0.05 Å. The assignment of different rest states for N···H-N *versus* O···H-N hydrogen bonds consistent with the hydrogen bonding observed in the crystal structures of small organic compounds,[9–11] e.g. $d_{N\cdots H\text{-}N} = 3.0$ Å and $d_{O\cdots H\text{-}N} = 2.9$ Å, fails to reproduce the trends in hydrogen bond lengths noted above. These differences in standard configurations also have a slight effect on derived complementary base-pair parameters in representative oligonucleotide structures, but virtually no effect on base-pair step parameters.

Computational independence

Local complementary base-pair and dimer step parameters computed with respect to the standard reference frames are nearly independent of analytical treatment (Figure 2). The only significant discrepancies in derived values, illustrated here for the DNA complexed with the TATA box-binding protein (TBP),[12] involve the Rise at highly kinked base-pair steps, which, as noted previously,[1] reflects an inconsistency in definition. The small differences in Slide, Tilt, and Twist in this example stem from minor differences in definition and in the choice of ''middle frame.''

Base-pair geometry in high-resolution *A*-DNA and *B*-DNA crystal structures similarly shows limited dependence on computational methodology. The average values and dispersion of individual parameters in Table 3 are representative of numerical values obtained with the algorithms used in many nucleic acid analysis programs. A complete listing of local *A* and *B*-DNA parameters, expressed in terms of the standard reference frame and computed within `3DNA` (X.-J. Lu & W.K. Olson, unpublished results) using the mathematical definitions of several different programs, `CEHS/SCHNAaP`,[13,14] `CompDNA`,[15,16] `Curves`[17,18] `FREEHELIX`,[19] `NGEOM`,[20,21] `NUPARM`,[22,23] and `RNA`,.[24-26] is reported at the NDB Web site (see below). Since the angular parameters differ by no

Table 1. Cartesian coordinates of non-hydrogen atoms in the standard reference frames of the five common nitrogenous bases

		Atom	Base	x (Å)	y (Å)	z (Å)
Adenine						
ATOM	1	C1′	A	−2.479	5.346	0.000
ATOM	2	N9	A	−1.291	4.498	0.000
ATOM	3	C8	A	0.024	4.897	0.000
ATOM	4	N7	A	0.877	3.902	0.000
ATOM	5	C5	A	0.071	2.771	0.000
ATOM	6	C6	A	0.369	1.398	0.000
ATOM	7	N6	A	1.611	0.909	0.000
ATOM	8	N1	A	−0.668	0.532	0.000
ATOM	9	C2	A	−1.912	1.023	0.000
ATOM	10	N3	A	−2.320	2.290	0.000
ATOM	11	C4	A	−1.267	3.124	0.000
Cytosine						
ATOM	1	C1′	C	−2.477	5.402	0.000
ATOM	2	N1	C	−1.285	4.542	0.000
ATOM	3	C2	C	−1.472	3.158	0.000
ATOM	4	O2	C	−2.628	2.709	0.000
ATOM	5	N3	C	−0.391	2.344	0.000
ATOM	6	C4	C	0.837	2.868	0.000
ATOM	7	N4	C	1.875	2.027	0.000
ATOM	8	C5	C	1.056	4.275	0.000
ATOM	9	C6	C	−0.023	5.068	0.000
Guanine						
ATOM	1	C1′	G	−2.477	5.399	0.000
ATOM	2	N9	G	−1.289	4.551	0.000
ATOM	3	C8	G	0.023	4.962	0.000
ATOM	4	N7	G	0.870	3.969	0.000
ATOM	5	C5	G	0.071	2.833	0.000
ATOM	6	C6	G	0.424	1.460	0.000
ATOM	7	O6	G	1.554	0.955	0.000
ATOM	8	N1	G	−0.700	0.641	0.000
ATOM	9	C2	G	−1.999	1.087	0.000
ATOM	10	N2	G	−2.949	0.139	−0.001
ATOM	11	N3	G	−2.342	2.364	0.001
ATOM	12	C4	G	−1.265	3.177	0.000
Thymine						
ATOM	1	C1′	T	−2.481	5.354	0.000
ATOM	2	N1	T	−1.284	4.500	0.000
ATOM	3	C2	T	−1.462	3.135	0.000
ATOM	4	O2	T	−2.562	2.608	0.000
ATOM	5	N3	T	−0.298	2.407	0.000
ATOM	6	C4	T	0.994	2.897	0.000
ATOM	7	O4	T	1.944	2.119	0.000
ATOM	8	C5	T	1.106	4.338	0.000
ATOM	9	C5M	T	2.466	4.961	0.001
ATOM	10	C6	T	−0.024	5.057	0.000
Uracil						
ATOM	1	C1′	U	−2.481	5.354	0.000
ATOM	2	N1	U	−1.284	4.500	0.000
ATOM	3	C2	U	−1.462	3.131	0.000
ATOM	4	O2	U	−2.563	2.608	0.000
ATOM	5	N3	U	−0.302	2.397	0.000
ATOM	6	C4	U	0.989	2.884	0.000
ATOM	7	O4	U	1.935	2.094	−0.001
ATOM	8	C5	U	1.089	4.311	0.000
ATOM	9	C6	U	−0.024	5.053	0.000

Standard chemical structures are taken from Clowney *et al.*[5] Those data do not include C1′ atoms, which are placed here in the least-squares plane of the base atoms, with the purine C1′-N9 bond length and C1′-N9-C4 valence angle set to 1.46 Å and 126.5° and the pyrimidine C1′-N1 bond and C1′-N1-C2 angle to 1.47 Å and 118.1°, respectively. These distances and angles are based on the average glycosyl geometries of purines and pyrimidines in high-resolution crystal structures of nucleic acid analogs from the CSD (J. Westbrook & H. M. Berman, unpublished results).

more than 0.1° and most distances by 0.02 Å or less, the general trends in the Table can be used in combination with the characteristic patterns of *A* and *B*-DNA backbone and glycosyl torsion angles[27] to classify local, right-handed, double-helical conformations.

The subtle mathematical differences among nucleic acid analysis programs, however, become

Table 2. Comparative base-pair geometry of optimized coordinate frames and high resolution structural analogs

Base-pair[a]				Hydrogen-bond distances (Å)					Complementary base-pair parameters					
	$d_{C1'\cdots C1'}$ (Å)	λ_Y (deg.)	λ_R (deg.)	O2···H-N2	N3···H-N1	N4-H···O6	N3-H···N1	O4···H-N6	Propeller (deg.)	Buckle (deg.)	Opening (deg.)	Shear (Å)	Stretch (Å)	Stagger (Å)
Ideal Models														
C·G	10.8	54.2	54.5	2.87	3.00	3.00	-	-	0	0	0	0	0	0
T·A	10.7	54.5	54.5	-	-	-	2.96	3.05	0	0	0	0	0	0
U·A	10.7	54.6	54.5	-	-	-	2.95	3.02	0	0	0	0	0	0
X-ray analogs[b]														
G·C	10.7	55.4	53.6	2.80	2.94	2.98	-	-	−5.4	−2.3	1.2	0.13	−0.07	0.10
G·C	10.7	55.9	52.7	2.79	2.92	2.92	-	-	−0.9	1.0	−0.1	0.30	−0.12	−0.07

Base-pair configurations sampled at 0.1 Å increments of $d_{C1'\cdots C1'}$ between 9.5-11.5 Å and 0.1° intervals in λ_Y and λ_R between 50° and 58°, i.e. $21 \times 81 \times 81 = 137{,}781$ states.

[a] Based on the survey of high-resolution crystal structures of nucleic acid analogs by Clowney *et al*.[5] Values are unchanged in a survey of current structures.

[b] Average inter-base parameters of free C·G Watson-Crick base-paired co-crystal complexes.[6,7]

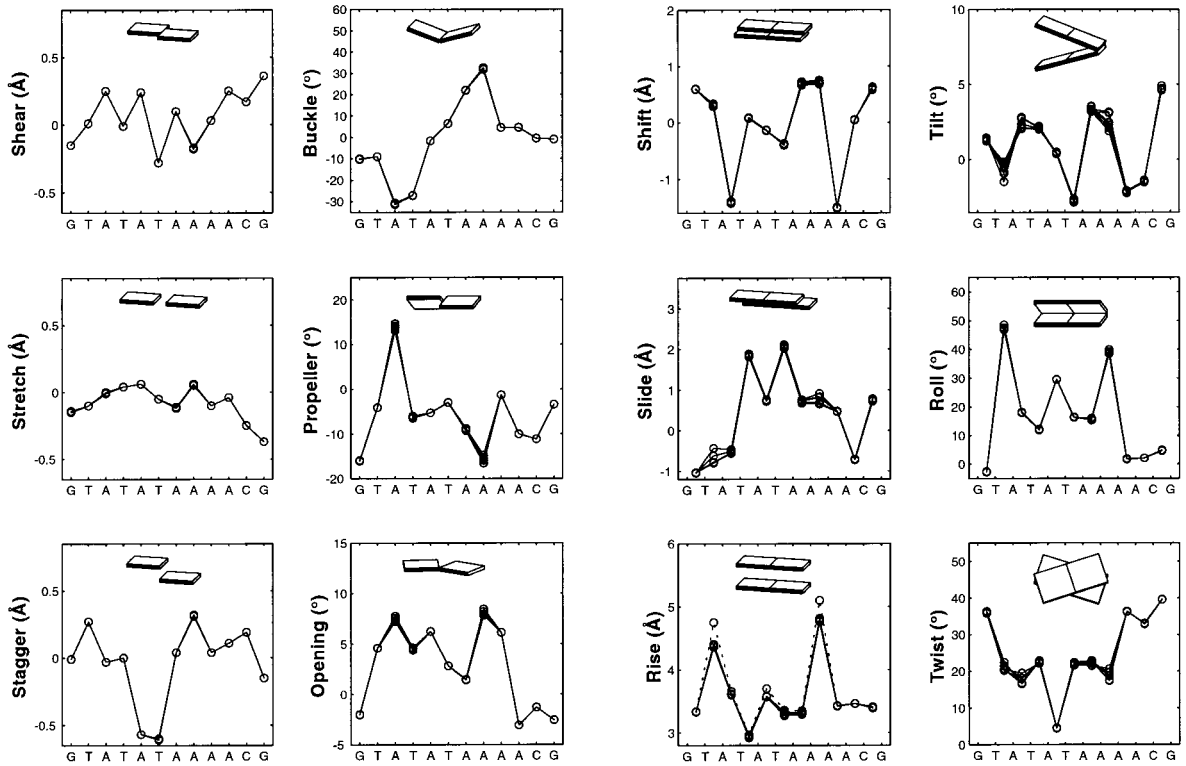

Figure 2. Comparative analysis of local base-pair (left) and dimer step (right) parameters (see schematic insets for definitions) of the DNA associated with the yeast TATA box-binding protein (TBP) in the 1.8 Å X-ray crystal complex (NDB entry: `pdt012`).[12] Parameters are calculated with the seven different analysis schemes within `3DNA` (X.-J. Lu & W. K. Olson, unpublished results) using the standard reference frame detailed in Tables 1 and 2. Agreement among programs is evident from the superposition of open circles. Dotted line connects Rise values computed using the `Curves` definition.[18] Numerical values are tabulated in the Supplementary Material available from `http://ndbserver.rutgers.edu/NDB/archives/`.

critical in the construction of DNA models. Seemingly minor numerical discrepancies can be magnified in polymeric chains[28] and in knowledge-based potentials[29] derived from the fluctuations and correlations of structural parameters. Duplex models and simulations must accordingly be based on the algorithm from which parameters are derived.

Conformational classification

The average values of Roll, Twist, and Slide in Table 3 confirm conformational distinctions known since the earliest studies of *A* and *B*-DNA crystal structures.[30,31] Namely, the transformation from *B* to *A*-DNA tends to decrease Twist, increase Roll, and reduce Slide. The standard deviations in recently accumulated crystallographic data, however, show that only Slide retains the discriminating power anticipated previously. Values of Slide below −0.8 Å are typical of most *A*-DNA dimer steps and those greater than −0.8 Å are found in the majority of *B*-forms. Slide is also more variable in *B*-DNA than in *A*-DNA dimer steps. The observed Twist and Roll angles, by contrast, show significant overlaps over a broad range of values. Specifically, Twist angles between 20° and 40° and Roll angles between 0° and 15° are found in both *A* and *B*-DNA structures. The values of Twist and Roll are coupled with changes in Slide, so that conformational assignments should be made in the context of all three parameters.[29]

The three remaining step parameters and the six complementary base-pair parameters are unaffected by helical conformation. The mean values and scatter of these values are roughly equivalent in high-resolution *A* and *B*-DNA structures (Table 3). The constraints of hydrogen bonding presumably give rise to the more limited variations in Opening and Stretch compared to other complementary base-pair angles and distances. Buckle, while fixed, on average, at zero, shows more pronounced fluctuations than Propeller, which is decidedly perturbed from ideal, i.e. 0°, planar geometry in all double-helical structures.

Helical parameters

Parameters relating consecutive residues with respect to a local helical axis can be computed using `CompDNA`,[15,16] `NUPARM`,[22,23] `RNA`,[24-26] and `3DNA` (X.-J. Lu & W. K. Olson, unpublished results), or in terms of a global axis with `CEHS`,[13]

Table 3. Average values and dispersion of base-pair parameters in high resolution *A* and *B*-DNA crystal structures

Parameter	Symbol[a]	*A*-DNA	*B*-DNA
Complementary base-pair parameters			
Buckle (deg.)	κ	−0.1 (7.8)	0.5 (6.7)
Propeller (deg.)	π	−11.8 (4.1)	−11.4 (5.3)
Opening (deg.)	σ	0.6 (2.8)	0.6 (3.1)
Shear (Å)	*Sx*	0.01 (0.23)	0.00 (0.21)
Stretch (Å)	*Sy*	−0.18 (0.10)	−0.15 (0.12)
Stagger (Å)	*Sz*	0.02 (0.25)	0.09 (0.19)
Base-pair step parameters			
Tilt (deg.)	τ	0.1 (2.8)	−0.1 (2.5)
Roll (deg.)	ρ	8.0 (3.9)	0.6 (5.2)
Twist (deg.)	ω	31.1 (3.7)	36.0 (6.8)
Shift (Å)	*Dx*	0.00 (0.54)	−0.02 (0.45)
Slide (Å)	*Dy*	−1.53 (0.34)	0.23 (0.81)
Rise (Å)	*Dz*	3.32 (0.20)	3.32 (0.19)
Local helical parameters			
Inclination (deg.)	η	14.7 (7.3)	2.1 (9.2)
Tip (deg.)	θ	−0.1 (5.2)	0.0 (4.3)
Helical twist (deg.)	Ω_h	32.5 (3.8)	36.5 (6.6)
x-displacement (Å)	*dx*	−4.17 (1.22)	0.05 (1.28)
y-displacement (Å)	*dy*	0.01 (0.89)	0.02 (0.87)
Helical rise (Å)	*h*	2.83 (0.36)	3.29 (0.21)

Data are based on the analysis within 3DNA (X.-J. Lu & W. K. Olson, unpublished results) of base-pairs and dimer steps in the following *A* and *B*-DNA crystal structures of 2.0 Å or better resolution without chemical modification, mismatches, drugs or proteins from the Nucleic Acid Database:[4] ad0002 (two molecules), ad0003, ad0004, adh008, adh010, adh0102, adh0103, adh0104, adh0105, adh014, adh026, adh027, adh029, adh033, adh034, adh038, adh039, adh047, adh070, adh078, adj0102, adj0103, adj0112, adj0113, adj022, adj049, adj050, adj051, adj065, adj066, adj067, adj075, bd0001, bd0005, bd0006, bd0014, bd0016, bd0018, bd0019, bdj017, bdj019, bdj025, bdj031, bdj036, bdj037, bdj051, bdj052, bdj060, bdj061, bdj081, bdl001, bdl005, bdl020, bdl084. Mean values and standard deviations (values in parentheses) exclude terminal dimer units, which may adopt alternate conformations. All calculations were performed with respect to the standard reference frame given in Tables 1 and 2. See the following URL for complete sequences and literature citations: http://ndbserver.rutgers.edu/NDB/archives/.

[a] Symbols follow guidelines established in conjunction with the conceptual framework used to define base-pair geometry[2] and subsequently modified by the IUPAC/IUB.[35] The symbols for helical twist and helical rise are those used by Gorin *et al.*[15]

(as implemented in the SCHNAaP software package[14]) NEWHELIX,[32] and Curves.[17,18] These angles and distances depend on how the helical axis is defined, particularly in deformed segments of the double-helical structure.[33] The local helical parameters of high-resolution *A* and *B*-DNA structures in Table 3 complement the dimeric descriptions of these structures. The *x*-displacement shares the same discriminating power as Slide in differentiating *A*-DNA from *B*-DNA, as anticipated from model building,[31] whereas Inclination and Helical Twist span overlapping ranges of values. The different mathematical definitions of local helical parameters yield numerical similarities equivalent to those found with dimer step parameters. Global helical parameters, which reflect a best-fit linear or overall curved molecular axis, are not necessarily comparable with these values (data not shown).

Intrinsic correlations

As is well known,[1,25] dimer step parameters depend on the choice of base-pair reference frame and can be perturbed significantly by distortions of complementary base-pair geometry. The base-pair reference frame in most nucleic acid analysis programs is an intermediate between the coordinate frames of the constituent bases.[33] The origin of this ''middle frame'' is shifted by significant distortions in Buckle and Opening, while the long *y*-axis is rotated by perturbations of base-pair Shear and Stagger (Figure 3). These changes, in turn, influence the step parameters describing the orientation and positions of neighboring base-pairs.

The effects of complementary base-pair deformations on dimer step parameters are most pronounced when perturbations of the same type, but of the opposite sense, occur in successive residues, i.e. Buckle, Opening, Shear, or Stagger is negative at base-pair i and positive at base-pair $i+1$ or *vice versa*. For example, a large negative difference in the buckle of consecutive base-pairs, $\Delta\text{Buckle} = \text{Buckle}(i+1) - \text{Buckle}(i)$, sometimes called Cup,[34] adds to the computed base-pair Rise of ''extreme'' dimer steps of high-resolution *A* and *B*-DNA crystal structures (Figure 3). Similarly, a large positive value of ΔOpening increases Shift, while large negative values of ΔStagger and large positive values of ΔShear enhance Tilt and Twist, respectively. Conversely, Rise, Shift, Tilt, and Twist can be depressed by large +ΔBuckle, −ΔOpening, +ΔStagger, and −ΔShear, respectively (Figure 3). On the other hand, Roll and Slide are not appreciably influenced by base-pair deformations.

Thus, extreme values of base-pair step parameters may simply reflect distorted or altered, i.e.

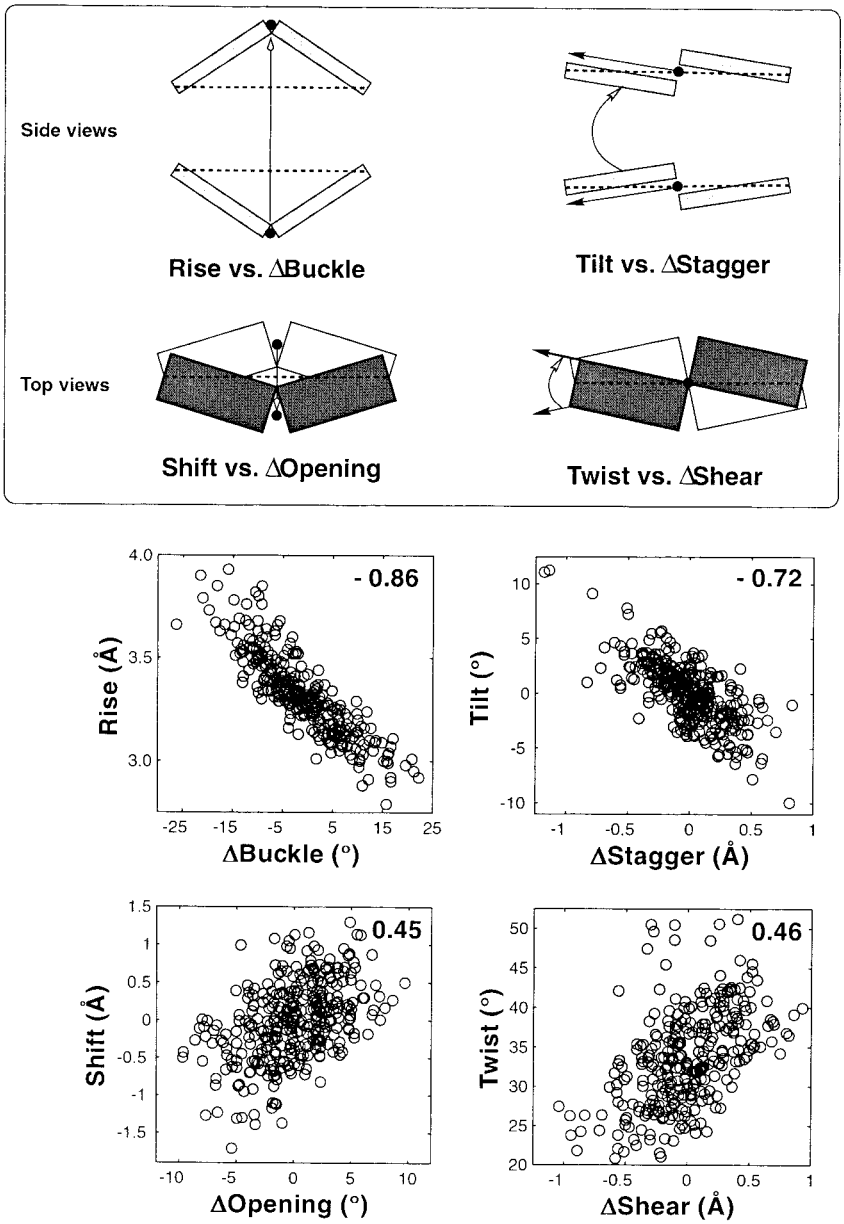

Figure 3. Illustrations and scatter plots of the intrinsic correlations of *A* and *B*-DNA base-pair and dimer step parameters associated with the standard reference frame. Large distortions of Buckle and Opening move the origin (●) of the base-pair reference frame, while significant changes in Shear and Stagger reposition the long *y*-axis (←) of the base-pair frame, leading to the changes in dimer step parameters.

non-Watson-Crick, base-pairing schemes. As a result, the computed Rise of a buckled dimer step with a partially intercalated amino acid side-chain in a protein-DNA complex such as TBP-DNA[12] may approach the base-pair separation found at a planar, fully drug-intercalated step. The dispersion of step parameters is similarly influenced by occasional deformations of complementary base-pair geometry. That is, Rise, Shift, Tilt, and Twist may appear intrinsically flexible in sets of structures with distorted base-pairing.

Non-Watson-Crick base-pairs

Direct application of the proposed reference frame to the analysis of non-Watson-Crick base-pairs yields numerical parameters characteristic of the particular hydrogen-bonding scheme. For example, 'wobble' G·T and A^{+}·C base-pairs are 'sheared' ~2 Å relative to the Watson-Crick configuration, the displacement being positive for the Y·R pair and negative for the R·Y association. These large displacements, in turn, affect Twist along the lines described in Figure 3. For example, the G·T mismatches in the d(CGCGAATTTGCG)$_2$ duplex structure (NDB entry: bd1009)[36] introduce ~15° under- and overtwisting in the associated CG and GA dimer steps since Shear is negative at the former step and positive at the latter step. The same principles apply in RNA structures where the G·U wobble assumes an important role.[37] On the other hand, Twist can be constrained to typical A or B-like values by proper choice of an intrinsic 'wobble' base-pair frame.[26] The latter approach necessitates a carefully chosen frame for each mode of base-pairing. In the future, it may be

necessary to define standards for common non-Watson-Crick base-pairing schemes.

Protein-DNA interactions

Characterizing the geometry of nucleic acids interacting with proteins, obviously, brings up a whole new host of geometrical issues. However, the standard description of base-pair geometry described here can be carried over, to a large degree, to this problem, and many of the geometrical issues involved in describing the protein are somewhat simpler than for the DNA, e.g. the description of helical geometry for an α-helix *versus* that for the DNA double helix.

A copy of this report is archived at the NDB Web site at `http://ndbserver.rutgers.edu/NDB/archives/`. Supplementary information is also available at this site, including Cartesian coordinates of non-hydrogen atoms in the standard reference frames of the five common nitrogenous bases; schematic representation of base-pair, dimer step, and helical parameters; comparative analysis of complementary base-pair and dimer step parameters in the DNA-TBP crystal complex (NDB entry: `pdt012`);[12] average values and dispersion of *A* and *B*-DNA complementary base-pair and dimer step parameters computed with representative nucleic acid analysis programs and the proposed standard reference frame; and intrinsic correlations of complementary base-pair and dimer step parameters.

Acknowledgements

The standardization of these parameters was the subject of the Tsukuba Workshop on Nucleic Acid Structure and Interactions that was organized by the Nucleic Acid Database (NDB; USA) and the Structural Biology Centre (Japan). The workshop was held on January 13-14, 1999 at the AIST-NIBHT Structural Biology Centre in Tsukuba, Japan and was sponsored by the COE program of the Science and Technology Agency (Japan) in association with the CREST program of the Japan Science and Technology Corporation and the National Science Foundation (USA; NSF BIR 95 10703). This document was prepared and made available from the NDB Web site (`http://ndbserver.rutgers.edu/NDB/archives/`) for public review and discussion. It was reviewed and approved by the IUBMB Nomenclature Committee at their May 2001 meeting. The survey of small molecule crystal structures was performed by John Westbrook with support from the NDB (NSF BIR 95 10703). The optimization of standard base-pair geometry and the calculation of derived parameters were carried out by Xiang-Jun Lu and Wilma K. Olson with support from U.S.P.H.S. grant GM20861.

The NC-IUBMB thanks a panel whose members were H.M. Berman (convenor, Rutgers University), M. Bansal (Indian Institute Science, Bangalore), S.K. Burley (Rockefeller University), R.E. Dickerson (University of California, Los Angeles), M. Gerstein (Yale University), S.C. Harvey (University of Alabama at Birmingham), U. Heinemann (Max-Delbrück-Centrum), S. Neidle (Institute of Cancer Research), W.K. Olson (Rutgers University), Z. Shakked (Weizmann Institute), H. Sklenar (Max-Delbrück-Centrum), M. Suzuki (AIST-NIBHT Structural Biology Centre), C.-S. Tung (Los Alamos National Laboratory), E. Westhof (Strasbourg), and C. Wolberger (Johns Hopkins University). Comments may be sent to any member of the panel, or to the convenor of the panel, H.M. Berman (berman@rcsb.rutgers.edu; Department of Chemistry, Rutgers, the State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854-8087 USA).

References

1. Lu, X.-J. & Olson, W. K. (1999). Resolving the discrepancies among nucleic acid conformation analyses. *J. Mol. Biol.* **285**, 1563-1575.
2. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O. *et al.* (1989). Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* **8**, 1-4.
3. Allen, F. H., Bellard, S., Brice, M. D., Cartright, B. A., Doubleday, A., Higgs, H. *et al.* (1979). The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallog. sect. B,* **35**, 2331-2339.
4. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T. *et al.* (1992). The Nucleic Acid Database - a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.
5. Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. & Berman, H. M. (1996). Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.* **118**, 509-518.
6. Fujita, S., Takenaka, A. & Sasada, Y. (1985). Model for interactions of amino acid side-chains with Watson-Crick base-pair of guanine and cytosine: crystal structrure of 9-(2-carboxyethyl)guanine and 1-methylcytosine complex. *Biochemistry,* **24**, 508-512.
7. Fujita, S., Takenaka, A. & Sasada, Y. (1984). A model for interactions of amino acid side-chains with Watson-Crick base-pair of guanine and cytosine. Crystal structure of 9-(2-carboxyethyl)guanine and its crystalline complex with 1-methylcytosine. *Bull. Chem. Soc. Jpn.* **57**, 1707-1712.
8. Wilson, C. C. (1988). Analysis of conformational parameters in nucleic acid fragments. II. Co-crystal complexes of nucleic acid bases. *Nucl. Acids Res.* **16**, 385-393.
9. Pirard, B., Baudoux, G. & Durant, F. (1995). A database study of intermolecular NH···O hydrogen bonds for carboxylates, sulfonates, and monohydrogen phosphonates. *Acta Crystallog. sect. B,* **51**, 103-107.
10. Llamas-Saiz, A. L. & Foces-Foces, C. (1990). N-H···N sp^2 hydrogen interactions in organic crystals. *J. Mol. Struct.* **238**, 367-382.
11. Gavezzotti, A. & Filippini, G. (1994). Geometry of the intermolecular X-H···Y (X, Y = N,O) hydrogen bond and the calibration of empirical hydrogen-bond potentials. *J. Phys. Chem.* **98**, 4831-4837.
12. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**, 512-520.
13. 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 with an appendix. *J. Mol. Biol.* **251**, 648-664.
14. Lu, X.-J., El Hassan, M. A. & Hunter, C. A. (1997). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J. Mol. Biol.* **273**, 668-680.
15. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34-48.
16. Kosikov, K. M., Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1999). DNA stretching and compression: large-scale simulations of double helical structures. *J. Mol. Biol.* **289**, 1301-1326.
17. 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.
18. Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655-667.
19. Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucl. Acids Res.* **26**, 1906-1926.
20. Soumpasis, D. M. & Tung, C. S. (1988). A rigorous basepair oriented description of DNA structures. *J. Biomol. Struct. Dynam.* **6**, 397-420.
21. Tung, C.-S., Soumpasis, D. M. & Hummer, G. (1994). An extension of the rigorous base-unit oriented description of nucleic acid structures. *J. Biomol. Struct. Dynam.* **11**, 1327-1344.
22. 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.
23. Bansal, M., Bhattacharyya, D. & Ravi, B. (1995). NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *CABIOS,* **11**, 281-287.
24. Babcock, M. S., Pednault, E. P. D. & Olson, W. K. (1993). Nucleic acid structure analysis: a users guide to a collection of new analysis programs. *J. Biomol. Struct. Dynam.* **11**, 597-628.
25. 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.
26. 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.
27. Schneider, B., Neidle, S. & Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers,* **42**, 113-124.
28. Olson, W. K., Marky, N. L., Jernigan, R. L. & Zhurkin, V. B. (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.
29. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA,* **95**, 11163-11168.
30. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981). Structure of a *B*-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA,* **78**, 2179-2183.
31. Calladine, C. R. & Drew, H. R. (1984). A base-centered explanation of the *B* to *A* transition in DNA. *J. Mol. Biol.* **178**, 773-782.
32. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a *B*-DNA dodecamer: CGCGAATT^{Br}CGCG. *J. Biol. Chem.* **257**, 14686-14707.
33. Lu, X.-J., Babcock, M. S. & Olson, W. K. (1999). Mathematical overview of nucleic acid analysis programs. *J. Biomol. Struct. Dynam.* **16**, 833-843.
34. Yanagi, K., Prive, G. G. & Dickerson, R. E. (1991). Analysis of local helix geometry in three *B*-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201-214.
35. Privé, 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.
36. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D. & Kennard, O. (1987). The structure of guanosine-thymidine mismatches in B-DNA at 2.5 Å resolution. *J. Biol. Chem.* **262**, 9962-9970.
37. Masquida, B. & Westhof, E. (2000). On the wobble G·U and related pairs. *RNA,* **6**, 9-15.

Edited by P. E. Wright

(Received 13 July 2001; accepted 17 July 2001)