Article No. jmbi.1998.2390 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 1563–1575

JMB

Resolving the Discrepancies Among Nucleic Acid Conformational Analyses

Xiang-Jun Lu and Wilma K. Olson*

Department of Chemistry Rutgers, the State University of New Jersey, Wright-Rieman Laboratories, 610 Taylor Road Piscataway, NJ 08854-8087 USA

*Corresponding author

Growing interest in understanding the relationship between the global folding of nucleic acids and the sequence-dependent structure of individual base-pair steps has stimulated the development of new mathematical methods to define the geometry of the constituent base-pairs. Several approaches, designed to meet guidelines set by the nucleic acid community, permit rigorous comparative analyses of different three-dimensional structures, as well as allow for reconstruction of chain molecules at the base-pair level. The different computer programs, however, yield inconsistent descriptions of chain conformation. Here we report our own implementation of seven algorithms used to determine base-pair and dimer step parameters. Aside from reproducing the results of individual programs, we uncover the reasons why the different algorithms come to conflicting structural interpretations. The choice of mathematics has only a limited effect on the computed parameters, even in highly deformed duplexes. The results are much more sensitive to the choice of reference frame. The disparate schemes yield very similar conformational descriptions if the calculations are based on a common reference frame. The current positioning of reference frames at the inner and outer edges of complementary bases exaggerates the rise at distorted dimer steps, and points to the need for a carefully defined conformational standard.

© 1999 Academic Press

Keywords: nucleic acid conformational analysis; base-pair geometry; dimer step parameters; nucleic acid structure; conflicting rise values

Introduction

A rigorous scheme for calculating parameters that describe nucleic acid structures is a prerequisite for deciphering the sequence-dependent conformations and interactions of DNA and RNA. The importance of a common set of structural variables was recognized more than a decade ago at an EMBO Workshop, which established the conceptual framework now used to specify the arrangements of bases and base-pairs in the double helix (Dickerson *et al.*, 1989; Figure 1). The software which has been created to meet the guidelines of the workshop, however, differs subtly in the choice of mathematics and numerical methodologies. Individual interpretation and understanding of the loosely written and illustrated definitions have led to programmatic differences that confuse researchers not involved in the development of the routines, but who wish to characterize the conformations of nucleic acids. Further confounding the issue are the inconsistent descriptions of base-pair geometry that stem from the application of different programs to the same structure (Werner *et al.*, 1996; Fernandez *et al.*, 1997; Lu *et al.*, 1997a). As illustrated below, parts of some structures may appear to be quite ''normal'' according to one computational scheme, but are highly unusual according to another. Furthermore, conformational patterns extracted with one set of routines may be opposite from those collected with a different program. Because of these discrepancies, new structures are not easily compared with published data derived from different programs, and progress in understanding nucleic acid conformational principles and the response of DNA and RNA to protein and ligand binding is hindered.

Due to the complexities of nucleic acid structure, it is difficult, if not impossible, to understand the various analytical methods by simply applying the freely distributed software to selected test cases. Key to this issue is a thorough understanding from

E-mail address of the corresponding author: olson@rutchem.rutgers.edu

0022-2836/99/041563–13 $30.00/0

© 1999 Academic Press

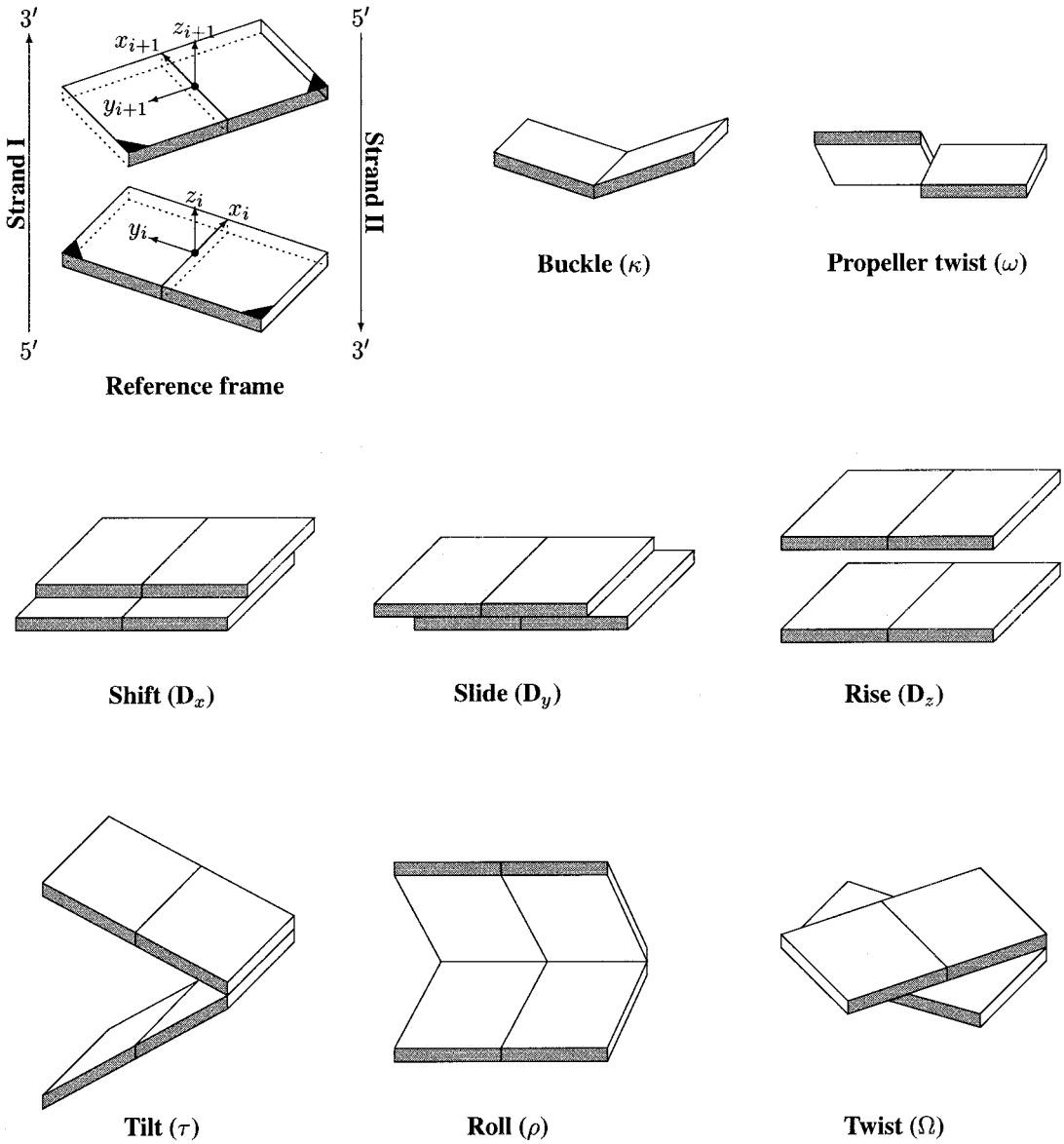

Figure 1. Pictorial definitions of reference frames and step parameters that relate sequential base-pairs and the two complementary base-pair parameters, buckle and propeller twist, common to the computer programs analyzed in this work. The base-pair reference frame is constructed such that the *x*-axis points away from the (shaded) minor groove edge and the N9(R)/N1(Y) atoms linked to the sugar-phosphate backbone (filled corners in reference frame). Images illustrate positive values of the designated parameters (Dickerson *et al.*, 1989).

both a mathematical and programmatic viewpoint of the different methods. As a first step to resolving the discrepancies among existing analysis routines, we have carried out a comparative study of local base-pair parameters with seven computational approaches. In order to understand the various methods, we have developed our own implementation of each program in a single computer package. Aside from reproducing the results of individual programs, we can control the choice of mathematics, reference frames, coordinate fitting, etc., and thereby uncover the reasons why the algorithms lead to different interpretations of nucleic acid structure. Each approach, which is simply a different way of describing double helical structure, has its own merits. No one method is superior to any other from a mathematical perspective or based on the description of a particular structure. Indeed, the different mathematics have only limited effects on conformational variables. All algorithms produce qualitatively similar conformational pictures when carried out with the same reference frame. The arbitrary choices of reference frames, however, have a profound influence on the computed data and point to the need for a common standard that fulfills our expectations and insight about nucleic acid conformational changes.

Results and Discussion

Scope and strategy

The seven schemes considered here are illustrative of the many possible mathematical approaches used in base and base-pair structural analyses and include some of the currently most influential computer packages: CEHS (El Hassan & Calladine, 1995), CompDNA (Gorin *et al.*, 1995), Curves (Lavery & Sklenar, 1988, 1989), FREEHELIX (Dickerson, 1998), NGEOM (Soumpasis & Tung, 1988; Tung *et al.*, 1994), NUPARM (Bansal *et al.*, 1995; Bhattacharyya & Bansal, 1989), and RNA (Pednault *et al.*, 1993; Babcock *et al.*, 1994; Babcock & Olson, 1994). Each of these programs conforms to the established conformational guidelines (Dickerson *et al.*, 1989), and produces parameters that are suitable for rigorous reconstruction of nucleic acid structure at the base-pair level. Closely related programs with similar functionalities are not included in the survey, e.g. Mazur & Jernigan (1995) define rotational variables in much the same way as Babcock *et al.* (1994) in RNA, Lu *et al.* (1997a) incorporate the rotational scheme developed by El Hassan & Calladine (1995) in CEHS and most of the functionalities of the NewHelix/FREEHELIX software in their SCHNAaP package, and Shpigelman *et al.* (1993) generate curved DNA structures using angles consistent with both CEHS and NGEOM.

We have thoroughly analyzed the source codes, examples, and publications available for the listed programs and have reproduced the output of each exactly with our own implementations in MATLAB (Math-Works, Inc., Natick, MA). Here, we concentrate on the six local step parameters (shift, slide, rise, tilt, roll, twist; Figure 1) that are common to the seven computational schemes. Only a subset of the programs compute all six local base-pair parameters, i.e. shear, stretch, stagger, buckle, propeller twist, opening. FREEHELIX and NUPARM determine propeller twist and buckle only, while Curves defines complementary base-pair geometry with respect to a global coordinate frame (see below). In principle, local base-pair parameters could be computed in Curves by using the base-pair reference frame that arises naturally as the ''middle frame'' between complementary bases (see below). Here, we illustrate and account for differences among algorithms using representative crystal structures from the Nucleic Acid Database (NDB; Berman *et al.*, 1992).

Conflicting structural descriptors

There are three fundamental differences among the various base and base-pair analysis packages: (1) the algorithm chosen to calculate parameters; (2) the reference frames used to position bases and base-pairs in three-dimensional space; and (3) the middle frame constructed so that the same parameters are obtained regardless of the direction from which a DNA or RNA structure is analyzed. The middle frame corresponds with what is also called the ''half-way rotation'' (Babcock *et al.*, 1994), ''mean plane'' (Lavery & Sklenar, 1989), or ''middle-step triad'' (El Hassan & Calladine, 1995).

Projection versus matrix-based algorithms

As detailed elsewhere (Lu *et al.*, 1999), the algorithms fall into two principal groups: (1) projection-based schemes that build upon the NewHelix algorithm developed by Dickerson and co-workers to characterize base-pairs in the first DNA crystal structures relative to a single global helical axis (Fratini *et al.*, 1982; Dickerson, 1985), and (2) matrix-based methods in the spirit of the rotational scheme adopted by Zhurkin *et al.* (1979) to analyze the anisotropic bending of DNA dimers. Bending and twisting angles are extracted in the former programs from various projections of base-pair coordinate axes, and in the latter from elements of the rotation matrix that relates neighboring base-pair frames (Lu *et al.*, 1999). The Curves package, while primarily projection-based, incorporates matrix operations in the definition of twist reminiscent of the coordinate transformations used in CEHS and NGEOM. This program is unique in determining a curvilinear global axis as well as a local frame to relate neighboring base-pairs. Both the projection and the matrix-based schemes offer clear illustrations of relative base-pair positioning. The matrix operations, however, are more readily adaptable to analytical formalisms (Flory, 1969) and large-scale molecular simulations (Levitt, 1983), and also provide an automatic definition of the orthogonal middle frame needed to define translational parameters. In contrast, the projection-based schemes introduce different averaging protocols and orthogonality corrections to locate the middle frame of a dimer step. Other distinctions among programs arise in the choice of projections or rotation matrices used to define angular parameters (Lu *et al.*, 1999). For example, Babcock *et al.* (1994) take a single space-fixed rotation axis to obtain direction-independent rotational parameters, while El Hassan & Calladine (1995) and Tung *et al.* (1994) express the same transformation in terms of a sequence of ''symmetrized'' Euler rotations.

Conflicting definitions of Rise

All programs define shift and slide consistently as the x and y-components of the vector connecting sequential base-pair origins ($o_{i+1} - o_i$) expressed in the middle frame (broken line in Figure 2). Rise is similarly described in all programs, except for Curves, as the z-component of this translational vector. In Curves, rise is the sum of two equal distances, $|p_i - o_i|$ and $|p_{i+1} - o_{i+1}|$ in Figure 2, where p_i and p_{i+1} are the points of intersection of the base-pair normals with the xy-plane of the middle frame. This unusual definition leads to

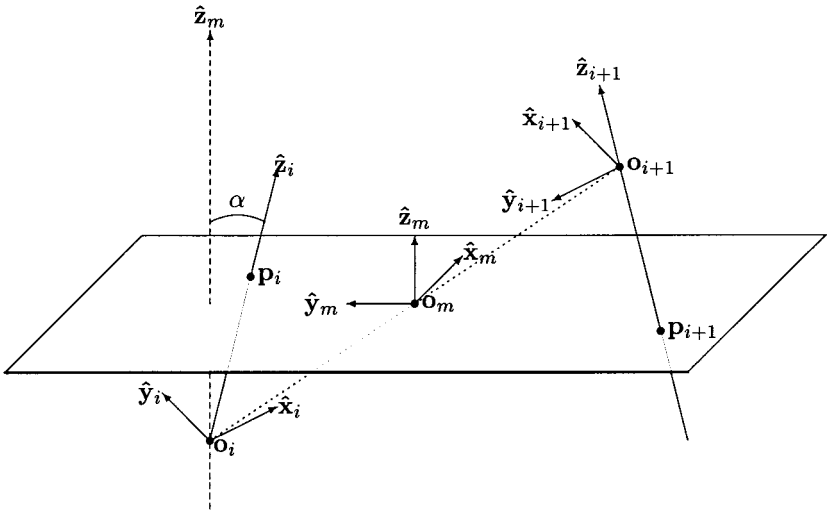

Figure 2. Schematic of base-pair frames i and $i+1$ and the middle frame m used to calculate local step parameters.

values of rise that are consistently larger than those determined by other methods, with the two lengths related by the angle α in Figure 2, i.e. half the bending angle between successive base-pair normals: $\text{rise}_{\text{others}}/\text{rise}_{\texttt{Curves}} = \cos\ \alpha$, $\alpha = \frac{1}{2}\cos^{-1}(\hat{z}_i\ \hat{z}_{i+1})$, where $\hat{z}_i$ and $\hat{z}_{i+1}$ are unit vectors along the designated normals. The differences in rise are thus exaggerated at highly-kinked steps. Furthermore, since rolling is generally favored over tilting at most base-pair steps (Zhurkin *et al.*, 1979), and tilting is rarely more than 15° (Gorin *et al.*, 1995; Hunter & Lu, 1997), the bend angle can be approximated by roll (ρ), i.e. $\alpha = \rho/2$ in the preceding formula. The combined effects of large tilt and roll on rise depend upon the assumed decomposition of the two bending components in individual programs, e.g. α is the Pythagorean sum of tilt and roll in `CEHS`.

Reference frames

Base-pair coordinate frames are defined in terms of the coordinate axes of complementary bases and/or specific atoms of the bases. Further differences lie in the fitting of coordinate axes to the bases, the assumed internal coordinates of bases (i.e. bond lengths, valence angles, torsions), and the origins of the base reference frames. The bases generated in crystal structures and computer models are not necessarily planar so that various procedures (presented in detail at http://rutchem.rutgers.edu/~olson/jmb/prog_comp.html) are used to find the base normals or to superimpose standard chemical structures onto the experimental coordinates before constructing the base-pair and middle frames. Except for `Curves`, which defines the local frame in terms of the canonical *B*-DNA fiber structure (Leslie *et al.*, 1980), the base origins are roughly coincident in the different schemes, but are significantly displaced (~0.8 Å along the positive *x*-axis) from the `Curves` reference. As illustrated below, this offset gives rise to systematic discrepancies of ~0.5 Å in slide and ~0.8 Å in global *x*-displacement in `Curves` compared with other programs, and also contributes to differences in rise at kinked steps. Babcock *et al.* (1994), by contrast, introduce a pivot point in each base to minimize built-in correlations involving translational parameters. While this pivot point has no influence on rotational parameters, it affects the origin of the base-pair and the translational parameters defined with respect to it (see below). None of the analysis packages as yet takes advantage of the idealized base geometries recently compiled by Clowney *et al.* (1996) from the Cambridge Structural Database (Allen *et al.*, 1994).

Influence of reference frame

Figure 3 illustrates the differences among six of the analysis packages when run respectively in their original form and in a common reference frame on the DNA base-pair steps of the yeast TATA-box (TBP) complex (Kim *et al.*, 1993), NDB code pdt012. Here, we omit the `NGEOM` analysis which makes ''adjustments'' (Tung *et al.*, 1994) to correct for the intrinsic sequence-specific differences in reference frames defined along the principal axes of bases and base-pairs, and reproduces the angles from `CEHS` and the distances from `RNA` when calculations are performed with respect to a common reference frame (see Table 1).

As evident from Figure 3 (left-hand side), the different approaches yield similar patterns of slide and roll (plus twist and shift which are not shown here) along the highly deformed DNA structure. The slide values computed with `Curves`, however, are consistently lower (~0.5 Å) than the data obtained with other methods, the differences directly linked to the unusual (*B*-DNA fiber) location of base frames noted above. The computed variations in rise, and to a lesser extent, those in tilt (data not shown), are noticeably different among the various schemes. The rise patterns cluster into two groups in the Figure: the `CompDNA`, `Curves`, and `RNA` algorithms yield zig-zag plots of

Table 1. RMS deviations of rotational and translational step parameters in randomly generated dimer configurations for designated pairs of analysis schemes

	CEHS	CompDNA	Curves	FREEHELIX	NGEOM	NUPARM	RNA	Average
CEHS[a]	-	0.129	0.053	0.126	0	0.550	0.216	0.179
CompDNA[b]	0.002	-	0.153	0.027	0.129	0.521	0.178	0.189
Curves[c]	0.020	0.020	-	0.146	0.053	0.552	0.222	0.197
FREEHELIX[d]	0.003	0.001	0.020	-	0.126	0.498	0.153	0.179
NGEOM[e]	0.051	0.051	0.055	0.051	-	0.550	0.216	0.179
NUPARM[f]	0.106	0.106	0.109	0.106	0.055	-	0.348	0.503
RNA[g]	0.051	0.051	0.055	0.051	0	0.055	-	0.222
Average	0.039	0.038	0.047	0.039	0.044	0.089	0.044	

The angular values, in degrees (upper right triangle), and the translational differences, in Å (lower left triangle), are stable after several hundred cases. See our website, http://rutchem.rutgers.edu/~olson/jmb/prog_comp.html, for further information.

[a] El Hassan & Calladine (1995); Lu *et al.* (1997a).
[b] Gorin *et al.* (1995).
[c] Lavery & Sklenar (1988, 1989).
[d] Dickerson (1998).
[e] Soumpasis & Tung (1998); Tung *et al.* (1994).
[f] Bansal *et al.* (1995); Bhattacharyya & Bansal (1989).
[g] Pedault *et al.* (1993); Babcock *et al.* (1994); Babcock & Olson (1994).

rise *versus* chain sequence and show noticeable increases in rise at the highly rolled TA·TA and AA·TT steps, whereas the CEHS, FREEHELIX, and NUPARM routines generate smoother and less pronounced changes in rise with base-pair position. The former programs also produce high values of rise at steps where the latter analyses give low values, and *vice versa*. For example, the rise at the kinked sites, where protein side-chains partially intercalate or pack against the DNA, is significantly increased according to the CompDNA, Curves, and RNA routines, but close to the 3.4 Å *B*-DNA norm in the CEHS, FREEHELIX, and NUPARM analyses. On the other hand, the former packages report standard rise values for the DNA between the sites of amino acid interaction, whereas the latter programs imply partial separation of base-pairs at the same steps. As expected from the unique definition of rise noted above, the Curves algorithm yields consistently higher

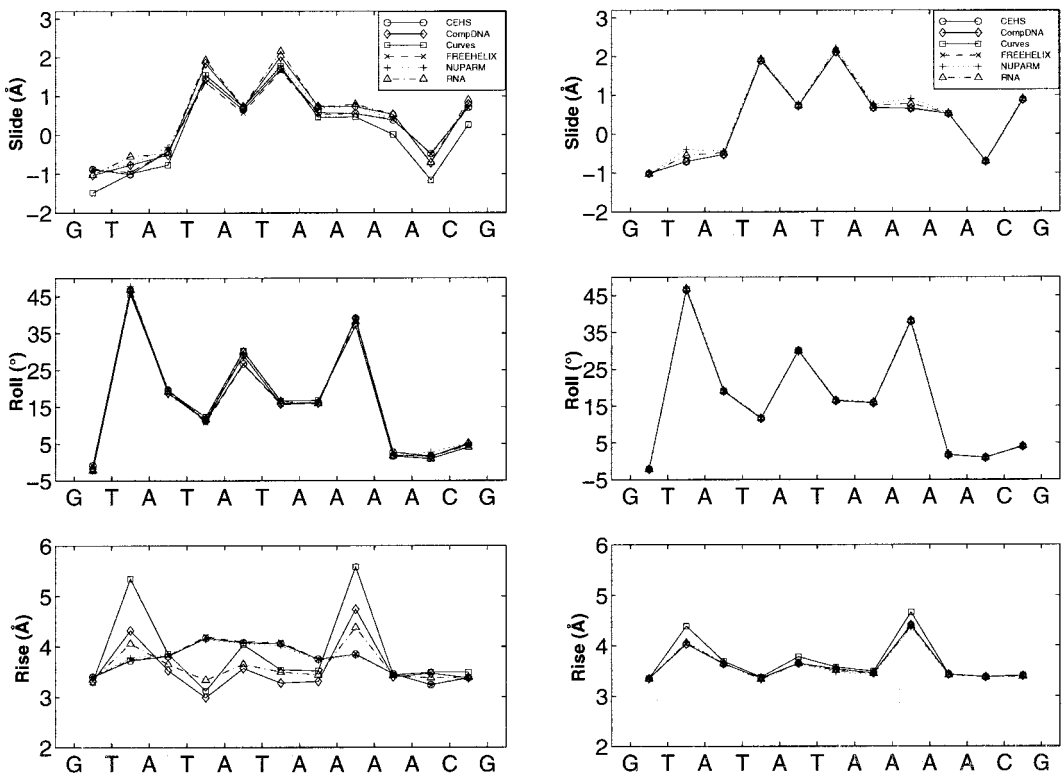

Figure 3. Comparison of the three local step parameters of DNA in the yeast TBP·TATA-box complex (Kim *et al.*, 1993), calculated using the designated schemes in their original form (left-hand side) and with respect to a common reference frame, here the local base-pair coordinate frames defined by the RNA software package (right-hand side).

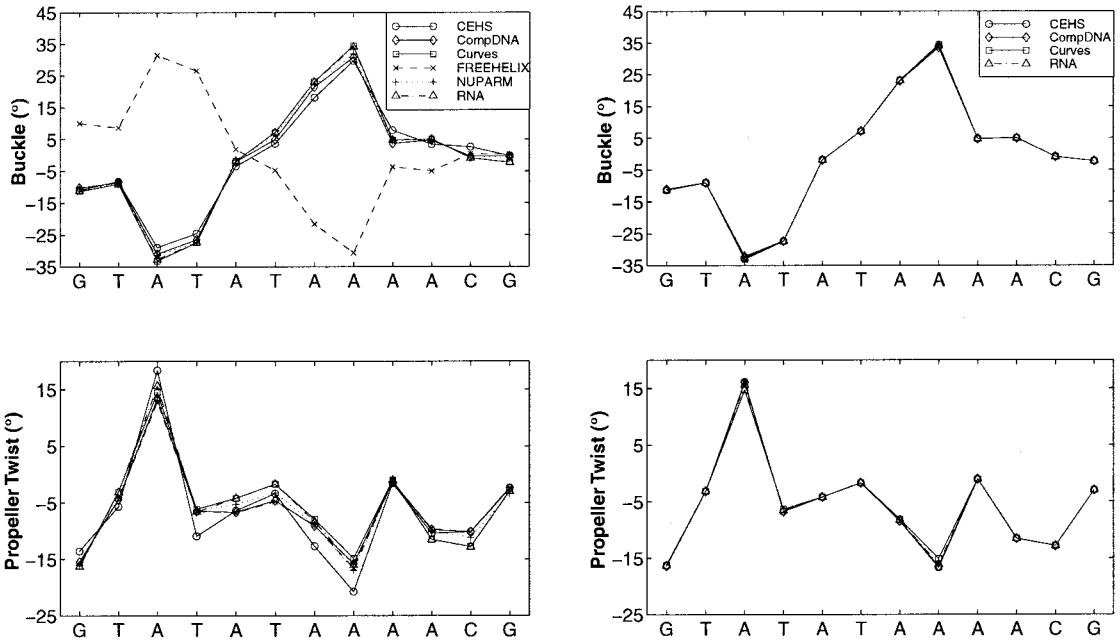

Figure 4. Buckle and propeller twist of DNA base-pairs in the yeast TBP·TATA-box complex (Kim *et al.*, 1993) calculated using the designated schemes in their original form (left column) and relative to the RNA base reference frame (right column). Note the different sign, but comparable magnitude of buckle reported by FREEHELIX. The cup, defined by Yanagi *et al.* (1991) as the difference between successive base-pair buckles, i.e. $\kappa_{i+1} - \kappa_i$, will also be reversed with this notation.

values of this parameter than all other approaches at unusual, i.e. highly rolled, base-pair steps.

The disparate mathematical schemes, however, yield nearly identical conformational descriptions if the calculations are performed with respect to a common reference, such as the base-pair axes generated with the RNA package (Figure 3, right-hand side). The major differences among nucleic acid structure analysis programs clearly stem from the choice of base and base-pair coordinate frames rather than mathematical approach. That is, the step parameters converge to values characteristic of the chosen coordinate frame. For example, the calculations based on the RNA base-pair frame in the right-hand side of Figure 3 follow the sequence of open triangles in the left-hand side of the Figure regardless of mathematics, whereas parameters determined in the CEHS frame would approach the open circles. With the exception of the Curves data at highly kinked dimer steps, the differences among rise values are almost indistinguishable when the TBP steps are analyzed from a common vantage point. The deviations in Curves rise at kinked steps again stem from its unusual definition. The NUPARM and RNA slide values also stand out at the extreme steps. The ~0.3 Å differences in slide (Figure 3, top right) arise from the slightly different construction of the middle frames in these programs compared with other methods (Lu *et al.*, 1999).

The corresponding comparison of complementary base parameters with the same six approaches in their original form and with respect to a common reference yields similar results (Figure 4). Differences in buckle and propeller twist, the only base-pair parameters common to all six methods, disappear when the calculations on the aforementioned DNA-TBP complex are performed using the same base coordinate frames. The FREEHELIX and NUPARM angles, which cannot be calculated from a set of arbitrary base frames (since these parameters are defined by the base normals and the C8(R)-C6(Y) axis), are omitted from the right-hand side of the Figure. The Curves data in this example are local parameters defined in terms of the base and base-pair reference frames that arise in the calculations, rather than the global parameters reported in the original program. Both complementary base-pair and dimer step parameters are highly sensitive to geometric distortions of individual bases and may differ substantially in the same program if a standard base is or is not fitted to the experimentally derived coordinates (see the following URL for numerical details: http://rutchem.rutgers.edu/~olson/jmb/prog_comp.html).

The local base-pair parameters calculated here closely match the global base-pair parameters determined with Curves in representative double helical structures: $d(GGGGCCCC)_2$, NDB code adh006 (McCall *et al.*, 1985); $d(CGCAATTGCG)_2$, NDB code bd1001 (Drew *et al.*, 1981); $d(CGCAAAAAAGCG)_2$, NDB code bd1006 (Nelson *et al.*, 1987). The root-mean-square (RMS) deviations of angular parameters are less than 0.5°. Discrepancies arise, however, at abnormal steps such as the extreme AA·TT kink in the conserved CAAT·ATTG sequence in the crystal complex of integration host factor (IHF) with DNA

Table 2. Method-dependent conformational correlation coefficients (r) and mean values of twist (deg.) and rise (Å) of high-resolution (better than 2 Å) *A*-DNA, *B*-DNA, and drug-intercalated dimer steps

	A-DNA (254)			*B*-DNA (173)			Drug-DNA (64) (Rise >5 Å, Twist >26°)			Drug-DNA (70) (Rise >5 Å)		
	Twist	Rise	r	Twist	Rise	r	Twist	Rise	r	Twist	Rise	r
CEHS	31.50	3.32	−0.30	35.64	3.33	−0.51	35.88	5.99	−0.07	34.32	6.07	−0.79
FREEHELIX	31.43	3.31	−0.29	35.62	3.33	−0.50	35.88	5.99	−0.09	34.32	6.07	−0.79
NUPARM	31.41	3.34	−0.32	35.61	3.33	−0.46	35.84	5.99	−0.01	34.28	6.07	−0.76
ComDNA	31.62	3.29	0.26	35.78	3.32	0.40	36.18	7.05	0.04	34.65	7.01	0.82
Curves	31.74	3.39	0.20	35.80	3.33	0.28	35.88	7.05	−0.17	34.37	7.01	0.76
RNA	31.69	3.29	0.14	35.79	3.32	0.24	36.23	6.66	0.10	34.69	6.67	−0.01

Analysis excludes structures with base mismatches or chemical modifications. The number of dimer steps within each class of structures is noted in parentheses.

Dimer steps are taken from the following files in the NDB (Berman *et al*., 1992): 32 *A*-DNA structures: 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, adj069, adj075, ad1025, udj032; 17 *B*-DNA structures: bdf068, bdj017, bdj019, bdj025, bdj031, bdj036, bdj037, bdj051, bdj052, bdj060, bdj061, bdj081, bd1001, bd1005, bd1020, bd1084, udj049; 36 Drug-DNA intercalation complexes: **ddb008, ddb009, ddb033, ddb034, ddd030,** ddf001, ddf018, ddf019, ddf020, ddf022, ddf026, ddf028, ddf029 , ddf031, ddf032, ddf035, ddf036, ddf038, ddf039, ddf040, ddf041, ddf044, ddf045, ddf049, ddf050, ddf052, ddf053, ddf054, ddf055, ddf056, ddf061, ddf062, ddf063, ddf065, ddf066, ddf079. The structures containing the six intercalation sites with Twist <26° are listed in boldface. See our website, http://rutchem.rutgers.edu/~olson/jmb/prog_comp. html, for full literature citations.

(Rice *et al*, 1996), NDB code pdt040. Here the global description exaggerates the propeller twist (−26.9°) and opening (19.7°) of the underlined A·T base-pair compared with the local perspective (−16.9° and 4.0°, respectively).

Algorithmic distinctions

As a further test of the different programs, we have compared the parameters of dimer steps simulated at random over the range of values observed in the *A* and *B*-DNA dinucleotide structure database compiled by El Hassan and Calladine (1997): Tilt [−10°, 10°], roll [−20°, 25°], twist [20°, 55°], shift [−1.0 Å, 2.0 Å], slide [−3.0 Å, 3.0 Å], rise [2.5 Å, 5.5 Å]. Table 1 lists the RMS deviations among computed step parameters for 10,000 dimer configurations within these parameter ranges. Each set of coordinate frames is generated from arbitrarily chosen rotational and translational states using CEHS construction routines (Lu *et al*., 1997b) and then reanalyzed with the seven computational approaches. That is, all calculations are carried out from a common vantage point, but with different mathematics. Angles used in the dimer building are limited to integral values and translations to 0.1 Å between the designated limits. The RMS values in the table reveal the mathematical distinctions between specific pairs of analysis schemes. Overall, the algorithmic differences are small, with average RMS deviations of computed rotational and translational parameters generally less than 0.3° and 0.05 Å, respectively. The NUPARM rotations and distances, however, stand out as more unusual than other values. The ~0.5° angular deviations and ~0.09 Å translational disparities in this scheme arise from the unusual construction of dimer frames, i.e. the middle frame z-axis is defined in terms of the x and y-axes rather than the average of the constituent base-pair normals (Lu *et al*., 1999). Notably, the NGEOM package yields the same rotational parameters as the CEHS routines and the same translational parameters as the RNA software (note the zero RMS entries in Table 1). Similar analyses of individual step parameters (data not shown) reveal the CEHS/NGEOM and Curves definitions of twist to be identical. The numerical data also confirm the identical definitions of twist in CompDNA and FREEHELIX (Lu *et al*., 1999) and the identical descriptions of rise obtained with CEHS, CompDNA, and FREEHELIX. Once again, the Curves rise consistently exceeds the values obtained with other methods.

Frame-dependent conformational correlations

The choice of reference frame is critical to understanding the conformational principles that can be extracted from nucleic acid structures. Conformational trends as well as individual parameter values depend upon the assumed positioning of base-pair axes. For example, reference frames like those from CompDNA, Curves, and RNA that increase rise at sites of partial amino acid intercalation in the TBP-DNA complex, lead to unexpected correlations between rise and twist in other nucleic acid structures. In contrast to the unwinding of DNA and RNA brought about by the intercalative binding of planar ligands in solution (Bauer & Vinograd, 1968), local DNA untwisting in *A* and *B*-DNA crystal structures is not associated with an increase in base-pair separation with these three schemes. Indeed, the correlations between twist and rise in 254 *A*-DNA and 173 *B*-DNA dimer steps from structures of resolution better than 2.0 Å are positive when the data are analyzed with the CompDNA, Curves, and RNA packages, but negative when examined with other programs (Table 2). As evident from the table, the levels of significance,

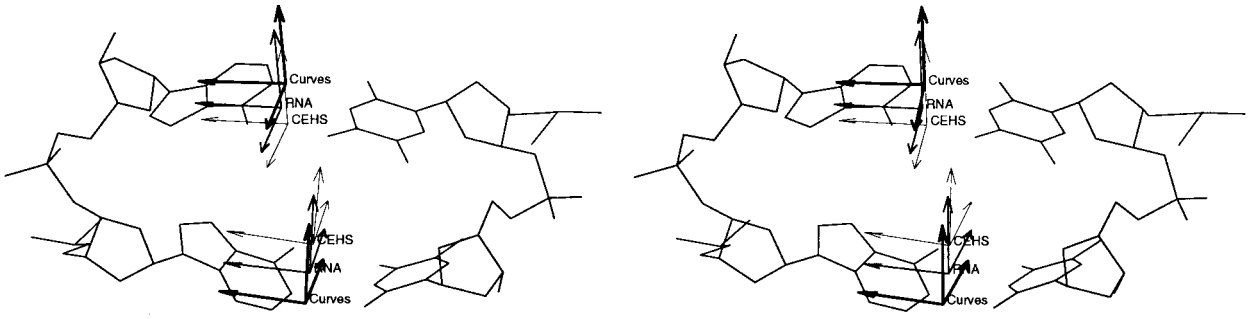

Figure 5. Stereo diagram of the superimposed local base-pair frames of the Curves, RNA, and CEHS schemes for the kinked AA·TT step in the IHF-DNA complex (Rice *et al.*, 1996). The view is towards the minor groove.

i.e. magnitudes, of the correlations also depend upon computational scheme and conformational sample. The significance levels are generally lower for the RNA data analysis which introduces a pivot point to minimize the interdependence of rotations and translations of complementary bases (Babcock & Olson, 1994). Correlations are also stronger for *B*-DNA compared with *A*-DNA. The sign of the rise-roll correlation, which is generally weaker than the rise-twist correlation, shows the same sort of reference frame dependence, i.e. negative for *B*-DNA analyzed with CompDNA and Curves, nil with RNA, and positive with all other routines (data not shown). Interestingly, these same correlations persist in 70 drug-intercalated dimer steps of 2.0 Å resolution or better, i.e. the twist-rise correlation is positive with CompDNA, Curves, and RNA, and negative with the other cases. Furthermore, the majority of currently available drug-intercalated sites show slight overwinding compared with *B*-DNA, and no significant twist-rise interdependence (Table 2). The six known dimer steps of large rise and low twist are the major contributors to the method-dependent correlations of twist and rise exhibited by the complete set of binding sites.

Contributions to Rise

Some extremes among base-pair reference frames are illustrated in Figure 5 for the severely kinked AA·TT step from the IHF-DNA complex (Rice *et al.*, 1996). The distortions of base-pairs in this example exaggerate the differences among coordinate frames obtained with programs like Curves and RNA, which define base-pair axes in terms of complementary base frames *versus* schemes like CEHS, which fix one of the base-pair axes along the C8(R)-C6(Y) line. As evident from the Figure, the large buckling of complementary bases displaces the base-pair origins defined by the different approaches. Here, a large positive (>40°) buckling in the lower base-pair combined with a large negative (< −25°) buckling in the upper residue substantially widens the distance between successive Curves and RNA origins compared with neighboring CEHS origins. This buckling pattern, also termed negative cup (see the legend to Figure 4), is seen at many sites of drug intercalation. The few unwound drug-intercalated steps noted above, however, exhibit a large positive cup with the inner edges of the bases in closer contact with bound drug than the C8(R)- and C6(Y)-edges. In these cases, the Curves rise is less than the CEHS and RNA values.

The Buckle-induced increment in Curves Rise over CEHS values is equal to $\Delta y[\sin(\kappa_i/2) - \sin(\kappa_{i+1}/2)]$, where Δy is the distance from the C8(R) or C6(Y) atom to the dyad axis of the undistorted base pair, i.e. 4.9 Å. The corresponding increment for RNA *versus* CEHS rise values is smaller, since the base origins used to define the base-pair coordinate frame, i.e. pivot points, are displaced 1.8 Å away from the dyad axis, i.e. $\Delta y = (4.9 - 1.8)$ Å.

The Curves rise in the IHF-DNA complex is further enhanced by the 60° Roll of the kinked AA·TT dimer in Figure 5. As noted in the above discussion of reference frames, the Curves base-pair origin is displaced ~0.8 Å toward the minor groove compared with the origins of other analysis schemes. This repositioning adds to the rise at positively rolled dimer steps, and reduces the rise when roll is negative (see below). The roll-induced increment in Curves rise over other schemes is equal to $2 \times 0.8 \sin(\rho/2)$ Å.

Rise: why the difference?

Rise is an important parameter for characterizing double helical structures (Sponer & Kypr, 1993b; Hunter & Lu, 1997). The separation of base-pairs is linked to both drug and protein-nucleic acid interactions. The correlation of rise with angular parameters, particularly cup, is well known (Bhattacharyya & Bansal, 1990; Yanagi *et al.*, 1991; Sponer & Kypr, 1993a; Babcock *et al.*, 1994), as are discrepancies in rise values obtained from Curves *versus* other routines (Werner *et al.*, 1996; Lu *et at.*, 1997a). The reasons for these differences and the reasons why bending affects base-pair displacement, however, have never been clear until now. Understanding how rise is defined in different nucleic acid analysis packages straightens out this confusion.

As is clear from the preceding illustrations, a number of factors influence the computed values of rise, complicating the comparison of different

Table 3. The observed differences in rise between published CompDNA and Curves values (Werner *et al.*, 1996) compared with corrections obtained with the approximate formula given in the text

Roll (°)	21	49	47	43	46	38	48	−52
$\text{Rise}_{\text{Curves}}$ (Å)	4.3	7.0	5.8	5.9	5.3	5.6	7.2	4.8
$\text{Rise}_{\text{CompDNA}}$ (Å)	3.9	5.8	4.7	5.0	4.3	4.7	5.9	5.0
$\text{Rise}_{\text{Curves}}$ corrected (Å)	3.9	5.7	4.7	4.9	4.2	4.8	5.9	5.0

nucleic acid conformational analyses. Furthermore, as first noticed by Babcock & Olson (1994), the correlations between rise and other base-pair parameters depend upon computational method. Both buckle and roll alter rise by repositioning the origins of neighboring base-pair frames. Their effects are related to the way in which the origins are defined and are dependent on the signs of these two parameters. The Curves rise is also increased over other values by its unusual decomposition into two terms. As noted above, all other programs define rise as a component of the displacement vector between base-pair origins.

These different contributions reinforce one another and account approximately for the enhancement of the Curves rise compared with values obtained with other methods:

$$\text{rise}_{\text{others}} \approx \text{rise}_{\text{Curves}} \cos(\rho/2) - 2 \times 0.8 \sin(\rho/2) - \Delta y(\sin(\kappa_i/2) - \sin(\kappa_{i+1}/2)).$$

Here, Δy depends upon the definition of the designated reference frames: 4.9 Å for CEHS, FREEHELIX, and NUPARM *versus* Curves; 1.8 Å for RNA *versus* Curves; 0 Å for CompDNA *versus* Curves. Because the roll and buckle values computed with different methods are normally different, the mean values from the schemes of interest can be inserted in this expression, e.g. the average of $\text{roll}_{\text{others}}$ and $\text{roll}_{\text{Curves}}$ for ρ. As seen above for the yeast TBP-DNA complex (Figures 3 and 4), roll and buckle values from different methods typically match quite closely.

Relation to other work

Comparative conformational analyses

Werner *et al.* (1996) were the first to report the large differences in CompDNA *versus* Curves rise values at kinked steps in several DNA-protein complexes. The reported discrepancies, however, vanish when the effects of roll on the Curves definition of rise are taken into consideration (Table 3). Curves rise values from Werner *et al.* (1996), which have been corrected for roll with the above expression, match the CompDNA data within 0.1 Å. Note that there is no buckling correction in this case, since the CompDNA and Curves origins lie approximately on the same short axis, i.e. $\Delta y = 0$. The negatively rolled example in Table 3 is a partially opened AT·AT step from the *Eco*RI endonuclease complex (Kim *et al.*, 1990), NDB code pde001. Here, in contrast to the positively kinked steps considered by Werner *et al.* (1996), the base-pairs roll into the minor groove and the CompDNA rise is larger than the Curves value. The Curves rise value can, nevertheless, be reconciled with the CompDNA definition with our simple formula.

Lu *et al.* (1997a) recently spotted even greater discrepancies among rise values for the AA·TT kinked step in the IHF-DNA complex (Figure 5). In this case the 8.5 Å Curves rise, which is more than double the CEHS value (4.1 Å) and ∼1.5 times the RNA value (5.6 Å), exceeds the maximum inter-atomic distance between adjacent bases. As noted in the above discussion, both roll and buckle contribute to the exaggerated value of the Curves rise in IHF-bound DNA. Consideration of the 60° roll and the unusually large negative cup of this dimer, i.e. $\kappa_{i+1} = 46°$ and $\kappa_i = -28°$, brings the Curves rise in closer agreement with the CEHS and RNA values (3.5 and 5.4 Å, respectively). The 0.6 Å difference between the corrected value and the computed CEHS rise is the worst approximation of our formula. Other subtle factors, such as the assumed value of Δy, affect base-pair displacement at this extreme dimer step. The best ''corrected'' Curves match is with CompDNA, $\text{rise}_{\text{CompDNA}} = 6.7$ Å *versus* corrected $\text{rise}_{\text{Curves}} = 6.6$ Å, the agreement tied to the common long axis reference point in the two programs, i.e. $\Delta y = 0$.

T. Elgavish & S. C. Harvey (unpublished data) have tested several base-pair analysis packages against a set of representative DNA structures and noticed the close correspondence between Curves and RNA local step parameters. As evident from the current work, the general agreement between these programs stems from the similar construction of base-pair reference frames. One of us has also noted the close correspondence of CompDNA and RNA dimer angles, despite the very different mathematics of these two approaches (Olson, 1996). The latter correspondence also reflects the similarity of base-pair coordinate frames. Indeed, the CompDNA parameters correlate even better with Curves data than with RNA findings (Table 3), because of the closer similarity of base-pair axes. CEHS and FREEHELIX step parameters also match one another because of the close correspondence of their respective base-pair reference frames (Lu *et al.*, 1999).

T. Elgavish & S.C. Harvey (unpublished data) and Lu *et al.* (1997a) have observed the 0.5 Å offset in slide values between Curves and RNA/CEHS (Figure 3, top left). We now understand that these difference reflect the unusual location of the base-pair origin in Curves compared with other methods. The same factor also accounts for the

−0.8 Å difference in global *x*-displacement, e.g. the displacement of base-pairs away from the helical axis, found with Curves *versus* other programs, and contributes as well to the discrepancies in Curves rise *versus* other values. The ∼5° difference in opening values between Curves and RNA observed by T.E. & S.C.H. (unpublished data) comes from slight distinctions between the *x* and *y*-axes of the different base reference frames.

Fernandez *et al*. (1997) have pointed out discrepancies in the global parameters of five *A*-DNA structures analyzed with the NewHelix (Fratini *et al*., 1982) and Curves packages. The different perspectives of the two programs, the former based on a best-fit linear global axis and the latter on a curved axial pathway, yield conflicting results even for the ''best cases'' when the double helix is nearly straight and base stacking is regular. The very different values of rise and the ∼0.5 Å offset in slide can, nevertheless, be understood from the current study of local base-pair step parameters. The different construction of reference frames influences the computed global parameters along the same lines reported here for local parameters. That is, the variable locations of base-pair origins alter translational parameters. The different signs of buckle in Curves *versus* NewHelix, also noticed by Lu *et al*. (1997a) and T.E. & S.C.H. (unpublished data), reflects inconsistent parameter definitions (see the legend to Figure 4).

Method-dependent structural differences

The recently reported Curves analysis of base-pair geometry (global parameters) in the wild-type PurR-DNA and mutant PurP L54 M-DNA crystal complexes gives an exaggerated (6.5 Å and 5.7 Å) rise at the CG·CG steps contacted by protein (Arvidson *et al*., 1998), NDBcode pdt063. Careful examination of the tabulated data additionally reveals significant roll (48° and 43°) and cup (−46° and −36°) that account for the abnormal displacement. The enhancement in rise in the DNA bound to the wild-type *versus* the mutant protein is thus method-related, the larger separation reflecting the greater roll and cup at the distorted steps in the wild-type complex. In contrast, CEHS analysis of the same protein-DNA complexes suggests only minor structural perturbations of rise at the CG·CG steps (3.8 Å and 4.0 Å, respectively), and RNA calculations yield intermediate and equivalent degrees of separation (4.8 Å and 4.8 Å, respectively). The >6 Å local Curves rise at the kinked TA·TA steps in the human TATA-binding protein complex (Juo *et al*., 1996) is similarly misleading and not necessarily indicative of ''the forcing apart of base-pairs''. CEHS and FREEHELIX rise values are normal at the kinked steps in this structure and slightly larger (∼4.0 Å) at the central base-pair steps, exactly as shown for the yeast TBP-DNA structure (Kim *et al*., 1993) in Figure 3.

Computer simulations and base-pair constraints

Finally, recent molecular dynamics simulations (Pardo *et al*., 1998) of the transformation of *A*-form DNA to the TA-conformation imposed by the TATA-binding protein (Guzikevich-Guerstein & Shakked, 1996) fail to account for the values of rise in the crystal complex (Kim *et al*., 1993). The simulations, however, constrain the buckle to different values from those observed in the crystal structure and only partially reproduce the pattern of roll. As noted above, both factors contribute to the large Curves rise in highly distorted DNA structures. On the other hand, Hunter & Lu (1997) have succeeded in reproducing (within a mean difference of ∼0.1 Å) the CEHS rise values of 400 dinucleotide steps from representative *A*- and *B*-DNA crystal structures (El Hassan & Calladine, 1997). The latter agreement may reflect the constraints on base-pair parameters, e.g. buckle, in these calculations (Hunter & Lu, 1997) as well as the limited range of distortions in *A* and *B*-DNA, i.e. cup < 15°.

A choice of reference frames: expectations *versus* observations

The reference frames for base-pair analysis should be chosen so that they fulfill our expectations of unusual as well as standard nucleic acid structures. For example, many protein side groups are non-planar and their insertion into double helical structures is only partial, kinking DNA or RNA at the sites of interaction. Distortions of this sort, long hypothesized (Tsai *et al*., 1975) to be intermediates in the intercalation of planar drugs and dyes, are likely to spread base-pairs apart to distances greater than the normal 3.4 Å van der Waals' separation, but less than the 6.8 Å spacing created by the ideal sandwiching of an aromatic ligand. Notions of partial intercalation and DNA/RNA kinking are clearly tied to the overall disposition of base-pair planes seen in molecular models.

Classic solution studies lead us to anticipate an untwisting in DNA bound to planar drugs and dyes (Bauer & Vinograd, 1968). The variation in rise with untwisting in *A*-DNA, *B*-DNA, and drug-intercalated base-pair steps, however, depends on the chosen reference frame, some programs yielding negative correlations and others positive or nil correlations. The crystallographic data also show that DNA/RNA unwinding is not necessarily concentrated at the sites of ligand intercalation (Table 2). Thus, one should not necessarily expect the anti-correlation of twist and rise at the dimer level.

The observed twist-rise correlations in *A*- and *B*-DNA helices and in drug-nucleic acid complexes ultimately depend on buckle, since roll is small in these unkinked structures. Indeed, we can reconcile the different correlations in Table 2 with our simple correction formula. When base-pairs are planar, the reference frame does not matter and all

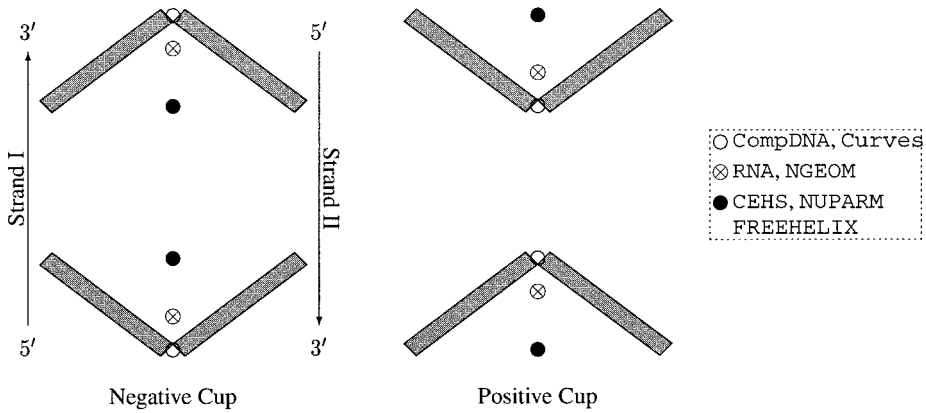

Figure 6. Influence of extreme base-pair buckling on the locations of reference frames used in seven analysis schemes. The view is along the *x*-axis with the (shaded) minor groove closer to the viewer. Cup is defined here as the difference in Buckle, i.e. $\kappa_{i+1} - \kappa_i$.

analysis packages yield comparable rise values. The negative cup of buckled base-pairs, however, brings the C8(R) and C6(Y) atoms at the outer edge of neighboring base-pairs near the van der Waals' contact limit (Figure 6). The average separation of the base-pairs is greater than this minimum contact distance, but is less than the distance between the open points in the Figure at the centers of the distorted base-pairs. Similarly, positive cup brings base-pair centers near the point of closest approach and enhances the separation of atoms like C8(R) and C6(Y) at the base-pair edges. The actual base-pair rise is again somewhere between these two extremes.

The base-pair frames in most available nucleic acid analysis packages are defined in terms of either base-pair outer edges or center points (Figure 6). `CEHS`, `FREEHELIX`, and `NUPARM` fall into the former category and accordingly underestimate the effects of negative cup on rise and exaggerate the influence of positive cup, whereas `CompDNA` and `Curves` constitute the second class, overestimating the rise at steps with negative cup and undervaluing rise when cup is positive. `NGEOM` and `RNA` attempt in different ways to incorporate features of chemical structure in the definition of base-pair frames, the former program by assigning a principal axis reference frame to individual bases, and the latter by choosing pivot points near the centers of the bases. The resulting rise values are of intermediate magnitude and the correlations of rise with other parameters are minimal.

In principle, the location of the base-pair origin also determines shift and slide. Extreme values of opening and large deviations of twist can affect these in-plane displacements in the same way that buckle and roll influence rise. The base-pair reference frame must also account for the anticipated behavior of laterally ''melted'' states implicated in important biological processes, but not yet captured crystallographically. Current structural examples (seen, for example, in the crystal complex of DNA with HhaI methyltransferase; Klimasauskas *et al.*, 1994) are limited to fully flipped-out bases, the shift-slide analog of drug-intercalated base-pair steps. The intermediates involved in base-pair disruption and transverse ''breathing'' are presumably less distorted than these known extremes.

Conclusions

This article is an attempt to clarify some of the confusion that has arisen from different analyses of nucleic acid three-dimensional structure. Our reconstruction of seven popular computational approaches shows the critical importance of the reference frame used to characterize base and base-pair geometry. The chosen standard affects the description of unusual dimers, such as the highly kinked steps of DNA in crystal complexes with proteins like TBP, IHF, and PurR, as well as the conformational trends of ordinary *A*- and *B*-DNA dimer steps and drug-intercalation sites. A given dimer step may appear to be quite normal according to one computational scheme, but may be highly unusual according to another. Furthermore, the conformational patterns extracted with one program may be opposite from those collected with another. These differences, in turn, affect our notions of molecular deformability (Olson *et al.*, 1998) and our expectations of how DNA and RNA will respond to external factors.

The mathematics used to interpret the loosely formulated guidelines for conformational analysis (Dickerson *et al.*, 1989), in contrast, have only a limited effect on computed parameters, even in highly deformed duplexes. Base-pair parameters from different schemes become virtually identical when determined with respect to the same base-pair axes. The numerical similarity, however, does not mean that one can mix and match parameters from different programs. Because small structural differences at the local level are magnified in long chains (Olson *et al.*, 1993), duplex reconstruction

must be based on the scheme from which the parameters are derived.

Only one mathematical descriptor, `Curves` rise, conflicts significantly with other programmatic definitions, leading to noticeably larger separation distances at highly deformed dimer steps. The different construction of middle coordinate frames also yields different representations of slide at these same steps.

A simple formula developed here to reconcile the conflicting values of rise found with different programs points to the extreme viewpoints now used in the analysis of base-pair geometry. Reference frames are based on the ideal *B*-DNA double helix, and do not generally anticipate common perceptions about deformed nucleic acid structures. The current construction of base-pair origins from points along the inner and outer edges of complementary bases also exaggerates the rise at distorted dimer steps, sometimes generating values greater than the separation of atom pairs in adjacent bases. While it may be impossible to find a reference frame which meets the expectations and needs of all researchers, the present work provides a numerical basis for achieving an ''optimized'' standard. Our analysis of the discrepancies in rise show that base reference frames midway between the base-pair outer edges and center points would avoid the exaggerated displacements reported in some programs at buckled base-pairs and minimize the correlation of rise with other parameters.

Acknowledgments

We are grateful to Dr Marla S. Babcock, Dr Manju Bansal, Dr Richard E. Dickerson, Dr Andrey A. Gorin, Dr Richard Lavery, and Dr Chang-Shung Tung for providing copies and documentation of their nucleic acid analysis programs, to Dr Stephen C. Harvey and Ms Tricia Elgavish for sharing their unpublished comparison of available software, to Dr Helen M. Berman, Dr Mustafa A. El Hassan, Dr Christopher A. Hunter, and Dr Chang-Shung Tung for valuable discussions on DNA structure and to Dr Suse Broyde, Dr Mustafa A. El Hassan, Dr Stephen C. Harvey, Dr Martin J. Packer, Dr Zippora Shakked, Dr Heinz Sklenar, and Dr Victor B. Zhurkin for carefully reading the manuscript and offering valuable comments. This research has been generously supported by the US Public Health Service under research grant GM20861. Computations were carried out at the Rutgers University Center for Computational Chemistry and through the facilities of the Nucleic Acid Database project (NSF grant DBI 9510703).

References

Allen, F. H., Kennard, O. & Watson, D. G. (1994). Crystallographic databases: search and retrieval of information from the Cambridge Structural Database. In *Structure Correlation* (Dunitz, J. D. & Bürgi, H.-B., eds), vol. 1, pp. 71-110, VCH Publisher, New York.

Arvidson, D. N., Lu, F., Faber, C., Zalkin, H. & Brennan, R. G. (1998). The structure of PurR mutant L54M shows an alternative route to DNA kinking. *Nature Struct. Biol.* **5**, 436-441.

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 dependent nucleic acid structures. *Comput. Appl. Biosci.* **11**, 281-287.

Bauer, W. & Vinograd, J. (1968). The interaction of closed circular DNA with intercalative dyes. I. The superhelix density of SV40 DNA in the presence and absence of dye. *J. Mol. Biol.* **33**, 141-171.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, L., 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.

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.

Bhattacharyya, D. & Bansal, M. (1990). Local variability and base sequence effects in DNA crystal structure. *J. Biomol. Struct. Dynam.* **8**, 539-572.

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.

Dickerson, R. E. (1985). Helix comparison table. In *Biological Macromolecules and Assemblies* (Jurnak, F. A. & McPherson, A., eds), vol. 2, pp. 471-494, Wiley-Interscience, New York.

Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucl. Acids Res.* **26**, 1906-1926.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. A., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S. & von Kitzing, E., *et al*. (1989). Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* **205**, 787-791.

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.

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. (1997). Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil. Trans. Roy. Soc. London*, **355**, 43-100.

Fernandez, L. G., Subirana, J. A., Verdaguer, N., Pyshnyi, D., Campos, L. & Malinina, L. (1997). Structural variability of the A-DNA in crystals of d(pCpCpCpGpCpGpGpG). *J. Biomol. Struct. Dynam.* **15**, 15l-163.

Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*, Wiley-Interscience Publishers, New York.

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**, 4686-470.

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

Guzikevich-Guerstein, G. & Shakked, Z. (1996). A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nature Struct. Biol.* **3**, 32-37.

Hunter, C. A. & Lu, X. J. (1997). DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.* **265**, 603-619.

Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). How proteins recognize the TATA box. *J. Mol. Biol.* **261**, 239-254.

Kim, Y., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990). Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307-1309.

Kim, Y., Geiger, J. H., Halm, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512-520.

Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357-369.

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.

Leslie, A. G. W., Arnott, S., Chandrasekaran, R., Birdsall, D. L. & Ratlif, R. L. (1980). Left-handed DNA helices. *Nature*, **283**, 743-746.

Levitt, M. (1983). Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.* **170**, 723-764.

Lu, X. J., El Hassan, M. A. & Hunter, C. A. (1997a). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J. Mol. Biol.* **273**, 668-680.

Lu, X. J., El Hassan, M. A. & Hunter, C. A. (1997b). Structure and conformation of helical nucleic acids: rebuilding program (SCHNArP). *J. Mol. Biol.* **273**, 681-691.

Lu, X. J., Babcock, M. S. & Olson, W. K. (1999). Mathematical overview of nucleic acid analysis programs. *J. Biomol. Struct. Dynam.* in press.

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. & Kermard, 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. E. & Klug, A. (1987). The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature*, **330**, 221-226.

Olson, W. K. (1996). Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* **6**, 242-256.

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.

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.

Pardo, L., Pastor, N. & Weinstein, H. (1998). Progressive DNA bending is made possible by gradual changes in the torsion angle of the glycosyl bond. *Biophys. J.* **74**, 2191-2198.

Pednault, E. P. D., Babcock, M. S. & Olson, W. K. (1993). Nucleic acids structure analysis: a users guide to a collection of new analysis programs. *J. Biomol. Struct. Dynam.* **11**, 597-628.

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.

Shpigelman, E. S., Trifonov, E. N. & Bolshoy, A. (1993). Curvature: software for the analysis of curved DNA. *Comput. Appl. Biosci.* **9**, 435-440.

Soumpasis, D. M. & Tung, C. S. (1988). A rigorous base-pair oriented description of DNA structures. *J. Biomol. Struct. Dynam.* **6**, 397-420.

Sponer, J. & Kypr, J. (1993a). Relationships among rise, cup, roll and stagger in DNA suggested by empirical potential studies of base stacking. *J. Biomol. Struct. Dynam.* **11**, 27-41.

Sponer, J. & Kypr, J. (1993b). Theoretical analysis of the base stacking in DNA: choice of the force field and a comparison with the oligonucleotide crystal structures. *J. Biomol. Struct. Dynam.* **11**, 277-292.

Tsai, C. C., Jain, S. C. & Sobell, H. M. (1975). X-ray crystallographic visualization of drug-nucleic acid intercalative binding: Structure of an ethidium-dinucleoside monophosphate crystalline complex ethidium: 5-iodouridylyl (3′-5′) adenosine. *Proc. Natl Acad. Sci. USA*, **72**, 628-632.

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.

Werner, M. E., Gronenborn, A. M. & Clore, G. M. (1996). Intercalation, DNA kinking, and the control of transcription. *Science*, **271**, 778-784.

Yanagi, K., Privé, 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.

Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. (1979). Anisotropic flexibility of DNA and the nucleosomal structure. *Nucl. Acids Res.* **6**, 1081-1096.

Edited by I. Tinoco

(Received 25 August 1998; received in revised form 3 November 1998; accepted 3 November 1998)