JMB

A-Tract Bending: Insights into Experimental Structures by Computational Models

Daniel Strahs and Tamar Schlick*

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and Howard Hughes Medical Institute, 251 Mercer Street, New York NY 10012, USA

While solution structures of adenine tract (A-tract) oligomers have indicated a unique bend direction equivalent to negative global roll (commonly termed ''minor-groove bending''), crystallographic data have not unambiguously characterized the bend direction; nevertheless, many features are shared by all A-tract crystal and solution structures (e.g. propeller twisting, narrow minor grooves, and localized water spines). To examine the origin of bending and to relate findings to the crystallographic and solution data, we analyze molecular dynamics trajectories of two solvated A-tract dodecamers: 1D89, d(CGCGA$_6$CG), and 1D98, d(CGCA$_6$GCG), using a new general global bending framework for analyzing bent DNA and DNA/protein complexes. It is significant that the crystallographically-based initial structures are converted from dissimilar to similar bend directions equivalent to negative global roll, with the average helical-axis bend ranging from $10.5°$ to $14.1°$. The largest bend occurs as positive roll of $12°$ on the 5′ side of the A-tracts (supporting a junction model) and is reinforced by gradual curvature at each A-tract base-pair (bp) step (supporting a wedge model). The precise magnitude of the bend is subtly sequence dependent (consistent with a curved general sequence model). The conversion to negative global roll only requires small local changes at each bp, accumulated over flexible moieties both outside and inside the A-tract. In contrast, the control sequence 1BNA, d(CGCGA$_2$TTCGCG), bends marginally (only $6.9°$) with no preferred direction. The molecular features that stabilize the bend direction in the A-tract dodecamers include propeller twisting of AT base-pairs, puckering differences between A and T deoxyriboses, a narrow minor groove, and a stable water spine (that extends slightly beyond the A-tract, with lifetimes approaching 0.2 ns). The sugar conformations, in particular, are proposed as important factors that support bent DNA. It is significant that all these curvature-stabilizing features are also observed in the crystallographic structures, but yield overall different bending paths, largely due to the effects of sequences outside the A-tract. These results merge structural details reported for A-tract structures by experiment and theory and lead to structural and dynamic insights into sequence-dependent DNA flexibility, as highlighted by the effect of an A-tract variant of a TATA-box element on bending and flexibility required for TBP binding.

© 2000 Academic Press

Keywords: molecular dynamics; A-tracts; sequence-dependent bending; narrow minor groove; stable water spine

*Corresponding author

Introduction

Sequence-dependent nucleic acid structure influences numerous physiological events, from the recognition and modulation of regulatory proteins (such as the CAP/DNA and TBP/DNA transcription complexes) to the packaging of the genetic material in chromatin (Bossi & Smith, 1984;

Struhl, 1985; Ryder *et al.*, 1986; Snyder *et al.*, 1986; Gartenberg & Crothers, 1991; Kim *et al.*, 1995). Adenine tracts (A-tracts) are prominent examples of such a sequence-dependent structural property (Crothers *et al.*, 1990). These sequences of four to six consecutive adenine residues, first observed in fragments of the kinetoplast minicircles from *Leishmania tarentolae*, were found to generate DNA helix bends of ≈18° for each set of consecutive adenine residues, spaced between guanine/cytosine-rich regions, when phased with the DNA helical repeat.

The bent DNA induced by A-tract sequences continues to be the subject of intense focus (Shatzky-Schwartz *et al.*, 1997; de Souza & Ornstein, 1997; Young & Beveridge, 1998; Sprous *et al.*, 1999), as the origin, extent, and details of this bending remain somewhat controversial (Crothers *et al.*, 1990; Hagerman, 1992; Dickerson *et al.*, 1994, 1996; Harvey *et al.*, 1995). To discuss these bends, we introduce the angular quantities of global tilt (θ_T) and global roll (θ_R); see Methods and Figure 1.† These quantities, a function of all base-pair (bp) step tilt, roll, and twist values, describe the bend direction and magnitude as measured with respect to a reference plane. This plane can coincide with a bp or lie between two bp planes. A-tract systems, as modeled by Crothers *et al.* (1990) are believed to have an overall bend in the helical axis that is often termed minor-groove bending; we prefer the term negative global roll ($\theta_R < 0$), since groove narrowing and bending toward that groove do not necessarily occur together. For example, the CAP co-crystal (Schultz *et al.*, 1991) has negative global roll without significant narrowing, while the Dickerson dodecamer (Wing *et al.*, 1980) has a slightly positive global roll but a narrow minor groove (Figure 1).

Crystallographic A-tract oligonucleotide structures possess overall bent helical axes, primarily characterized by a variable global tilt (i.e., bending toward backbones). However, the helical axes often describe bend directions that differ from structure to structure. Specifically, the PDB entry 1D89, d(CGCGA$_6$CG) (DiGabriele & Steitz, 1993) (for brevity, called $\underline{S}$ for skewed A-tract position) studied here exhibits $\theta_T < 0$, while the structure 1D98, d(CGCA$_6$GCG) (Nelson *et al.*, 1987) (called $\underline{C}$ for central A-tract position), exhibits $\theta_T > 0$; both have near-zero θ_R (Figure 1). However, some crystallographic A-tract structures are disordered and as such are heterogenous with respect to the bend direction within a single crystal. For example, 1BDN, d(CGCA$_5$TGCG) (DiGabriele *et al.*, 1989), exhibits both negative and positive θ_T; the three forms of $\underline{S}$ display slightly different bends, all with $\theta_T < 0$ and $\theta_R \leqslant 0$. In contrast, *in vitro* gel mobility

studies predict a unique bend direction along $\theta_R < 0$ at the center of each A-tract (Zinkel & Crothers, 1987; Koo & Crothers 1988).

To explain the observed global tilt/roll discrepancy between solution and crystallographic experiments, it has been suggested that the intermolecular forces generated in the crystal lattice and/or the dehydrating agents used to form the crystal influence the resulting structures (Dickerson *et al.*, 1994, 1996; DiGabriele *et al.*, 1989; DiGabriele & Steitz, 1993). Yet, despite differences in the θ_T and θ_R values of the crystallographic structures, the A-tract units themselves are very similar and relatively straight (Crothers *et al.*, 1990; Dickerson *et al.*, 1994, 1996; DiGabriele & Steitz, 1993), with root mean square (RMS) differences of 0.9 Å. The different characteristics observed for A-tract crystallographic structures, as well as the discrepancies between these structures and predicted solution properties, have contributed to various theoretical models of A-tract bending (Crothers *et al.*, 1990; Dickerson *et al.*, 1994, 1996; Haran *et al.*, 1994).

With significant improvements in the past decade, simulation techniques are now recognized as valuable complementary tools to experimental data (Schlick, 1999), with the potential to generate insightful structural and thermodynamic detail. Molecular dynamics (MD) simulations, in particular, can offer extensive molecular details regarding the structures underlying A-tract bending, thereby clarifying the differences between the crystallographic and solution structures of oligonucleotides and evaluating theoretical bending models. Here, we analyze intrinsic bending from nanosecond simulations of solvated DNA dodecamers with explicit counterions for the six-membered A-tract systems $\underline{S}$ and $\underline{C}$, as well as a control oligonucleotide that possesses a structure close to the intrinsic *B*-DNA form (Wing *et al.*, 1980): 1BNA, d(CGCGA$_2$TTCGCG). We use the Cornell *et al.* (1995) force field implemented within the CHARMM molecular modeling package (Brooks *et al.*, 1983), with both Newtonian and Langevin dynamics protocols employed, the latter with the efficient multiple-timestep method LN (Barth & Schlick, 1998a,b; Sandu & Schlick, 1999; Schlick, 1998), as discussed in Methods.

Our analyses focus on both the structural deviations as well as the similarities of our A-tract oligonucleotide models with respect to the starting crystallographic conformations, and their agreement with other experimental data. We show that the global bend of our MD structures (≈10° global tilt, −6° global roll per A-tract), is achieved through the combined mechanisms of a large positive roll at the 5′ side of the A-tract and a small level of bending between each AA dinucleotide step; the latter bending can be characterized by mean values of ≈ − 1° tilt and 1° roll per AA step. In addition, large amounts of propeller twist at each AT bp and a stable deoxyribose pseudorotation phase-angle decrease of 15° for the T deoxyriboses relative to A are stabilized by a narrow

---

† Our procedure is more general than that described by Ulanovsky & Trifonov (1987). The method applied by Beveridge and co-workers only measures the angle between the first and last bp steps (Young & Beveridge, 1998; Sprous *et al.*, 1999).

minor groove hydrated by a regular water spine. Though many of these details are observed in the crystallographic structures (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993), the emergence of the bend in the MD simulations occurs as a response to the molecular environment and the inherent flexibility of the DNA.

The structural details of our simulated models thus consolidate various experimental observations on A-tract dodecamers (Crothers *et al.*, 1990; Dickerson *et al.*, 1994, 1996; DiGabriele *et al.*, 1989; DiGabriele & Steitz, 1993; Haran *et al.*, 1994) and can be extended (X. Qian, D. S. & T. S., unpublished results) to analyze the relationship between intrinsic DNA bending for TATA-box variants and TBP binding (Patikoglou *et al.*, 1999); see the last section of Discussion.

Results

Direction of bending

Each dodecamer undergoes motions involving bending of the global helical axis throughout the simulations. For example, Figure 2 shows the bent helical axis for the $\underline{C}$ dodecamer, taken from a snapshot at 1.5 ns. While the bends ($\approx 15°$) and the bending ranges ($\approx 16°$) are comparable for all dodecamers (Figure 3), the overall motion of the control sequence reveals no clear pattern. The control dodecamer ensemble is bent $6.9°$ on average and the A-tract dodecamers are bent $14.1°$ ($\underline{S}$) and $10.5°$ ($\underline{C}$); these averages are obtained by taking the mean of the global roll and tilt values for each snapshot

$$(\sqrt[+]{\langle\theta_T^2\rangle + \langle\theta_R^2\rangle}).$$

These differences have profound global effects when viewed on larger scales, as seen from the longer DNA models developed based on a series of phased repeats of our dodecamers (Figure 4).

The bends resolved from the A-tract dodecamer MD ensembles are clearly oriented as $\theta_R < 0$ when viewed from the geometric center of the A-tract (Figure 3). Figure 3 shows that both the snapshot averages (yellow) and the angular values calculated for the average structures (red) display $\theta_R < 0$. This reflects a departure from the starting crystallographic structures (teal circles, Figure 3). These results are in good general agreement with predictions from electrophoretic mobility studies of A-tract oligonucleotides (Crothers *et al.*, 1990; Zinkel & Crothers, 1987; Koo & Crothers 1988; Haran *et al.*, 1994).

It is interesting that the bends in both the $\underline{S}$ and $\underline{C}$ ensembles are oriented in a slightly ''south-easterly'' direction (i.e., $\theta_T > 0$, $\theta_R < 0$); Young *et al.* also observed this trend in simulations of A-tracts of length five and six adenine residues (Young & Beveridge, 1998). The 2 ns trajectories of the $\underline{S}$ and $\underline{C}$ dodecamers display bends that oscillate among various θ_T/θ_R regions. This flexibility is notable from our simulations of the $\underline{S}$ dodecamer with both Newtonian and LN Langevin integrators (for enhanced sampling and computational performance) (Figure 3, lower half). The greater sampling achieved by multiple independent trajectories (Caves *et al.*, 1998) and stochastic dynamics ensembles (Schlick, 1998) permit a closer examination of the $\underline{S}$ dodecamer behavior.

Bending details

As initially observed in the A-tract simulations reported by Young *et al.* (Young & Beveridge, 1998; Sprous *et al.*, 1999), the largest helical axis bend occurs at the bp step on the 5′ side of the A-tracts ($\underline{S}$:G$_4$, $\underline{C}$:C$_3$: boxed values in Figure 5) (Shatzky-Schwartz *et al.*, 1997; Dickerson *et al.*, 1994; Nelson *et al.*, 1987; DiGabriele & Steitz, 1993). The average roll at this bp step ($\underline{S}$:G$_4$A$_5$ and $\underline{C}$:C$_3$A$_4$ respectively) is $\approx 12°$ and the motion at this bp step occupies a range of ≈ 20-$30°$, principally oriented along the positive roll axis. This local bend is principally responsible for the observed south-easterly bending trend ($\theta_T > 0$, $\theta_R < 0$) (bold row, Table 1). No other bp step in the interior of

Table 1. Analysis of individual base-pair step contributions to global tilt (θ_T) and global roll (θ_R)

	$\underline{S}$			$\underline{C}$		
Step	θ_T	θ_R		Step	θ_T	θ_R
2	10.7	−15.9		2	9.3	−8.4
3	9.5	0.5		**3 (5′)**	**−3.2**	**−0.8**
4 (5′)	**2.1**	**−1.6**		4 (A)	7.8	−9.5
5 (A)	9.0	−11.5		5 (A)	7.1	−8.8
6 (A)	10.1	−9.4		6 (A)	9.5	−7.5
7 (A)	12.6	−8.3		7 (A)	8.2	−6.5
8 (A)	11.1	−7.9		8 (A)	8.5	−5.3
9 (A)	11.2	−6.8		9 (3′)	11.0	−4.2
10 (3′)	16.3	−4.2		10	7.5	−4.1
All (2-10)	11.6	−8.1		All (2-10)	8.3	−6.4

In each row, we show the resulting global angular values in degrees produced by eliminating the corresponding tilt and roll values from Equations (1) or (2a,b). The bp steps 1 and 11 are omitted from our calculations. Note the dramatic reduction in angular values when the 5′ flanking A-tract bp step is omitted (bold face). The reference values for global tilt and global roll (using steps 2-10) are shown in the last row.
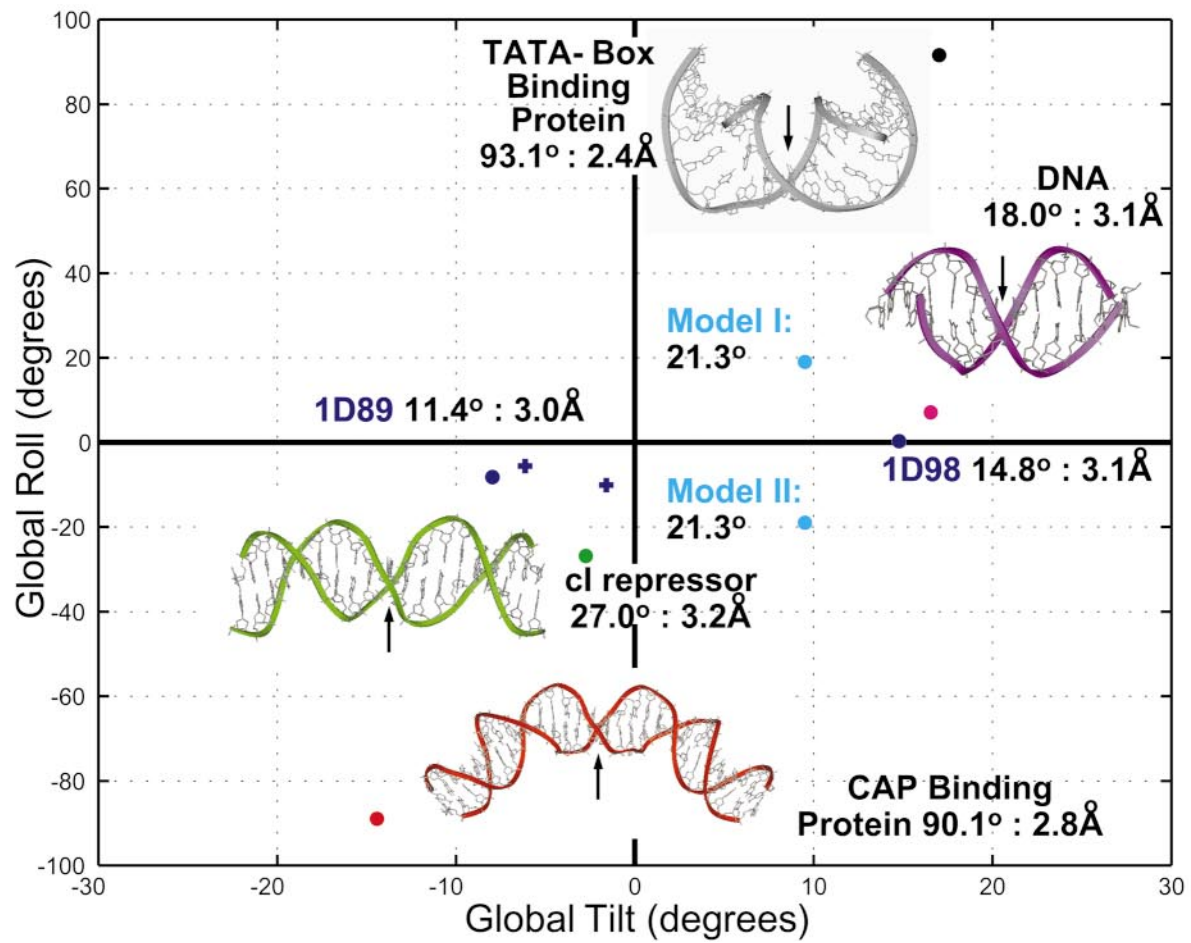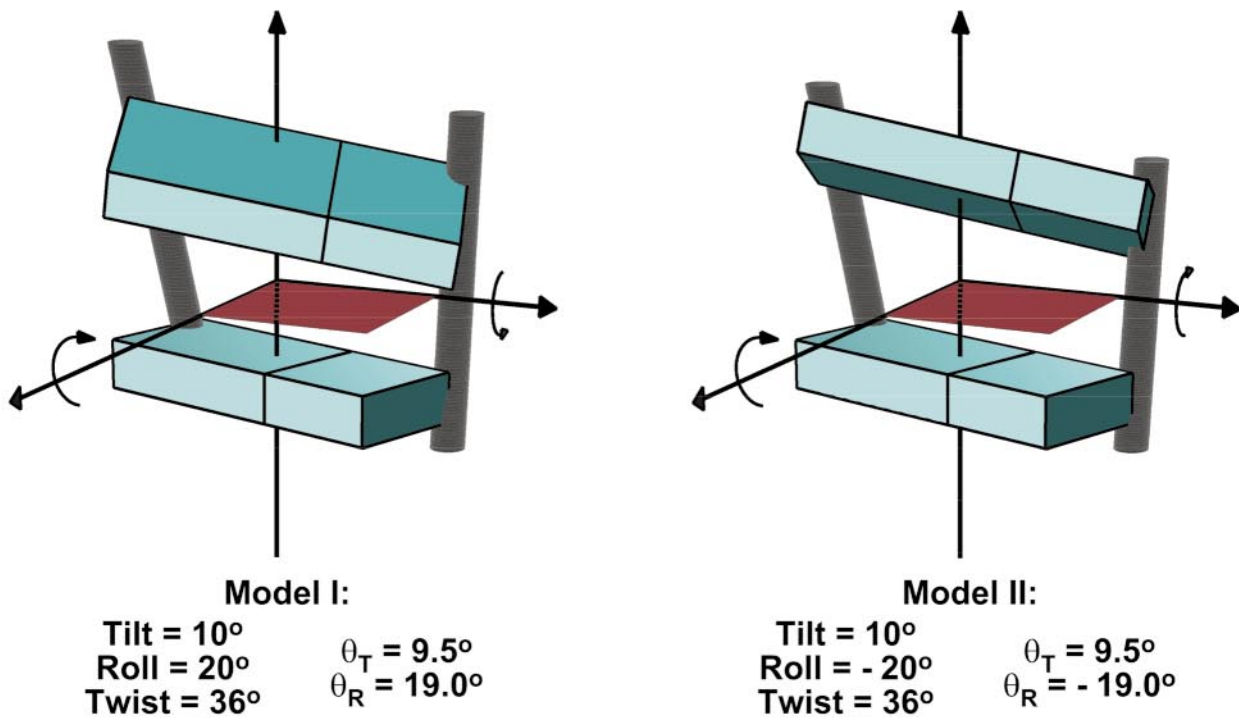
Figure 1 (*legend opposite*)

the $\underline{S}$ and $\underline{C}$ dodecamers displays a comparably large average bend, although the motion ranges are similar. We note a strong positive roll in the $\underline{C}$ crystal structure at step C_3A_4, but this is not observed in the $\underline{S}$ crystallographic structure.

The second largest bend results from cumulative internal features and the curvature of the A-tract. The A-tract region is characterized by systematic negative tilt and positive roll values at each bp step within the A-tract (Figure 5). Although observed in previous MD simulations of A-tracts (Young & Beveridge, 1998), the significance of this trend was not discussed. Below, we discuss the structural underpinning of this shallow bend that results from sugar pseudorotation differences observed between A and T deoxyriboses. The presence of this curvature in both crystallographic and theoretical models of A-tracts supports the ''wedge'' model structure of AA-dinucleotide bp steps.

While the roll at the 5′ end of both A-tracts is approximately 12°, a smaller roll of 8° just above this bend occurs at the $\underline{S}$ bp step C_3G_4. This bend is qualitatively similar, both in position and magnitude, to the largest bend observed in our A-tract dodecamers. The larger overall bend angle observed in our simulations of the $\underline{S}$ dodecamer depends on the presence of this bend (Table 1) and can be attributed to flanking sequence dependencies.

Structural features that support bent DNA

Analyses of the deoxyribose conformations reveal that a systematic 15° pseudorotation angle difference between A and T deoxyriboses leads to an intrinsic curvature of the A-tract region, and that the preference of the sugar on the 5′ side of the first A-tract adenine for the O4′-endo conformation correlates to the major bend at this junction. This consistent 15° sugar difference appears structurally related to the large amounts of propeller twist, narrow minor grooves, and a water spine, as we discuss.

Sugar conformation and curved A-tracts

The deoxyriboses in the three crystallographic structures, as well as our simulations, adopt primarily ''southern'' C2′-endo conformations (which is defined by $P = 162°$, where P is the pseudorotation angle (Altona & Sundaralingam, 1972) (Figure 6). However, in both cases, a 15° pseudorotation difference is present: $P = 115°$ for thymine residues, north of C1′-exo (C1′-exo $\equiv P = 126°$), and $P = 131°$ for adenine residues, just south of C1′-exo, on average. The control dodecamer's central AATT section also exhibits this trend (data not shown): the MD-ensemble averages of the P angles for A and T are $122.1(\pm39.1)°$ and $113.7(\pm31.5)°$, respectively.

This puckering difference is related to the intrinsic curvature of the A-tract (Figure 5). The distance between the adenine sugar's O3′ and C5′ atoms is increased relative to thymine sugars (Figure 6, model at bottom right). When this slightly increased backbone distance accumulates over consecutive adenine residues, it leads to a longer path along the helical backbone. This in turn introduces a net curvature to the A-tract helical axis, producing an intrinsic bend characterized by an average tilt of $\approx -1°$ and average roll of $\approx 1°$ at each AA-dinucleotide step.

The lower average P value of T sugars also displaces C1′ atoms towards the 5′ end of the strand relative to C1′ atoms of A on the opposite strand (Figure 6, sugar model). This backbone structural mechanism also supports propeller twisting and a narrow minor groove (Fratini *et al.*, 1982).

Sugar conformations at the A-tract junction bend

The largest single bend in our modeled A-tract structures is observed at the bp step on the 5′ side of the first A-tract adenine residue (Figure 5). The $\underline{S}$ and $\underline{C}$ deoxyriboses at this step, although flexible, adopt an average conformation between O4′-

Figure 1. Macromolecular examples of global tilt (θ_T) and global roll (θ_R) combinations. Top: Schematic models of a single bp step, with bps (cyan blocks) and deoxyribose-phosphate backbones (gray cylinders) shown. Two examples ($\theta_R > 0$, left; $\theta_R < 0$, right) are illustrated, oriented from the major groove. The angles were calculated according to Equation (2a, b), and the tilt, roll and twist values for each example are indicated. An ''average'' reference bp plane (brown) is positioned midway in the step; the axes defining θ_T and θ_R lie in this plane; the rotation directions of θ_T and θ_R, defined by equation (2a and b), are illustrated with arrows. Bottom: Examples of θ_T and θ_R for various molecular structures. Models are colored to identify the corresponding points. The bending angle magnitude

$$(\sqrt{\theta_T^2 + \theta_R^2})$$

and the end-to-end distance (normalized per bp) are indicated for each system. The arrows indicate the position of the reference plane and also the groove towards which bending occurs. Regions of positive and negative θ_R are often termed ''major-groove compression'' or ''minor-groove compression'' respectively. Co-crystals of dimeric major-groove binding proteins like CAP (Schultz *et al.*, 1991) and cI repressor (Beamer & Pabo, 1992) tend to wrap the DNA around the protein with the minor groove at the center of curvature; $\theta_R < 0$ for these structures. Structures with bends towards the major groove such as 1BNA (Wing *et al.*, 1980) or co-crystals of minor-groove binding proteins like TBP (Nikolov *et al.*, 1996) have $\theta_R > 0$. The θ_T and θ_R values of the two structures examined in this work (1D89 (or $\underline{S}$) and 1D98 (or $\underline{C}$)) are indicated in blue; the two forms of 1D89 not examined in this work are indicated by blue crosses.
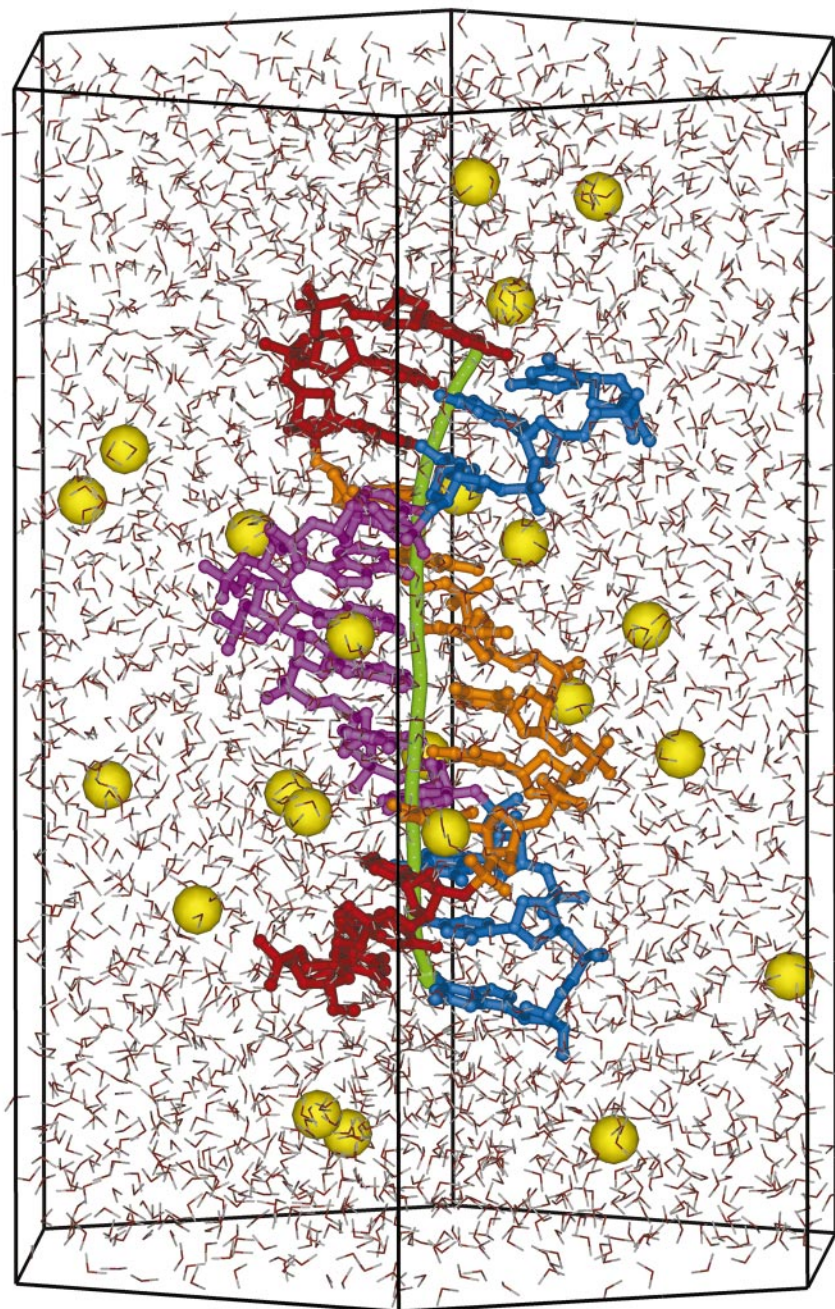
Figure 2. Representative snapshot from the C̲ MD simulation at time 1.546 ns. The A-tract region of the DNA is distinguished from the other nucleotides by color; the sodium ions are yellow and the water molecules are shown as stick images. The edges of the regular hexagonal prism periodic cell used are indicated; the cell height is 65.5 Å and each hexagonal side is 25.1 Å. The global helical axis produced by the Curves 5.2 program is shown in green (Lavery & Sklenar, 1989, 1997). The primary helical bend at the 5′ end of the A-tract is visible approximately three-fourths of the distance towards the top of the helical axis; a smaller, transient bend is visible near the bottom of the helical axis.

endo and C1′-exo: $P = 122.7°$ for G_4 of the S̲ dodecamer and $91.7°$ for C_3 of the C̲ dodecamer. These sugars are very flexible, spending approximately equal time in northern and southern puckers (Figure 6). The northern sugar conformations extend the C1′ proton towards the preceding base on the strand (C_3 and G_2, respectively), resulting in the upfield chemical shifts observed in NMR experiments (Kintanar *et al.*, 1987; Katahira *et al.*, 1988).

Thus, the northern conformational extensions and the flexibility of this sugar appear to be a common A-tract feature and may be correlated with the bending observed at this dinucleotide step.

Propeller twisting in the A-tract

Substantial propeller twisting has been observed in both crystallographic and solution NMR studies
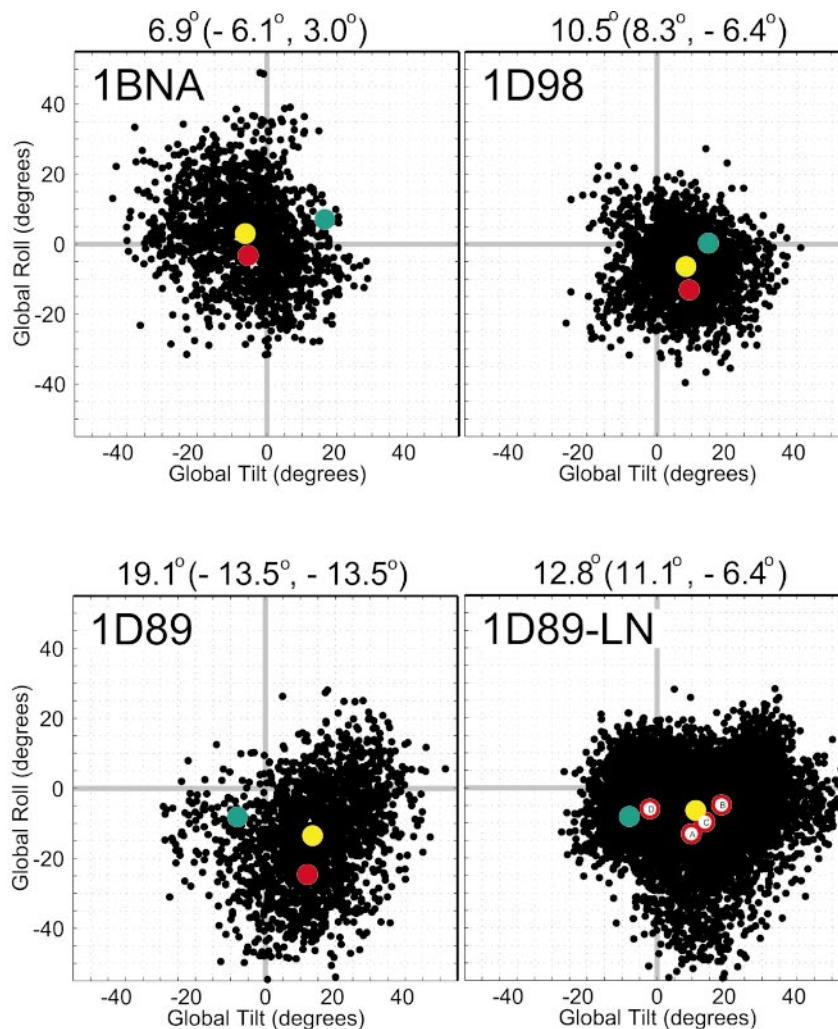
Figure 3. Bending of the dodecamer structures as collected over all analyzed trajectory snapshots compared to the starting crystallographic conformation (teal circle). Each snapshot, sampled at a 1 ps frequency, is indicated by one black point. The global tilt, θ_T, and global roll, θ_R, bend angles were calculated with respect to the plane at the center of the A-tract or the AT bp step in 1BNA (Equation (2)). Red circles indicate the average structures computed from the trajectories (as described below). Yellow circles indicate the average global roll and global tilt corresponding to all snapshots; these should not be confused with the values corresponding to the average structures. The average bending magnitude, the average global roll, and the average global tilt are indicated at the top of each panel in the format

$$\sqrt{\langle\theta_T\rangle^2 + \langle\theta_R\rangle^2}\ (\langle\theta_T\rangle,\ \langle\theta_R\rangle).$$

The average structures computed over four LN trajectories of the $\underline{S}$ dodecamer are indicated by the letters A-D. The 1BNA average structure was computed over the 500-2000 ps interval; the $\underline{C}$ dodecamer average structures were computed from the region 290-2000 ps; the regions used in computing the Newtonian dynamics $\underline{S}$ dodecamer average structure are 309-2000 ps. The abscissa and ordinate of each plot are θ_T and θ_R, respectively, with grid lines indicating 5° increments. The terminal bp steps in all three dodecamers were excluded from the global tilt/roll sums.

of A-tracts (Nelson *et al.*, 1987; DiGabriele *et al.*, 1989; DiGabriele & Steitz, 1993; Katahira *et al.*, 1988). This feature is also associated with AA-dinucleotide steps in statistical studies of nucleic acid structure (El Hassan & Calladine, 1996) Such geometries favor the formation of bifurcated hydrogen bonds within the A-tract (in addition to the Watson-Crick hydrogen bonds), linking the adenine residue N6 amino group to the next thymine's O4 group.

Initially, our crystal-based A-tract dodecamer structures are highly propeller-twisted within the

A-tract region and less so outside it (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993). Although decreasing from the initial crystallographic values ($\approx -23°$), the A-tract propeller twisting persists throughout our MD simulations, reaching an average of $-16°$ (Figure 7). This value agrees well with the estimated figure of $-17°$ from NMR studies (Katahira *et al.*, 1988). The MD average distances between the bifurcated atom functional groups adenine N6 and thymine O4 are larger than the crystallographic lengths (4.2 *versus* 3.2 Å) (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993), perhaps due
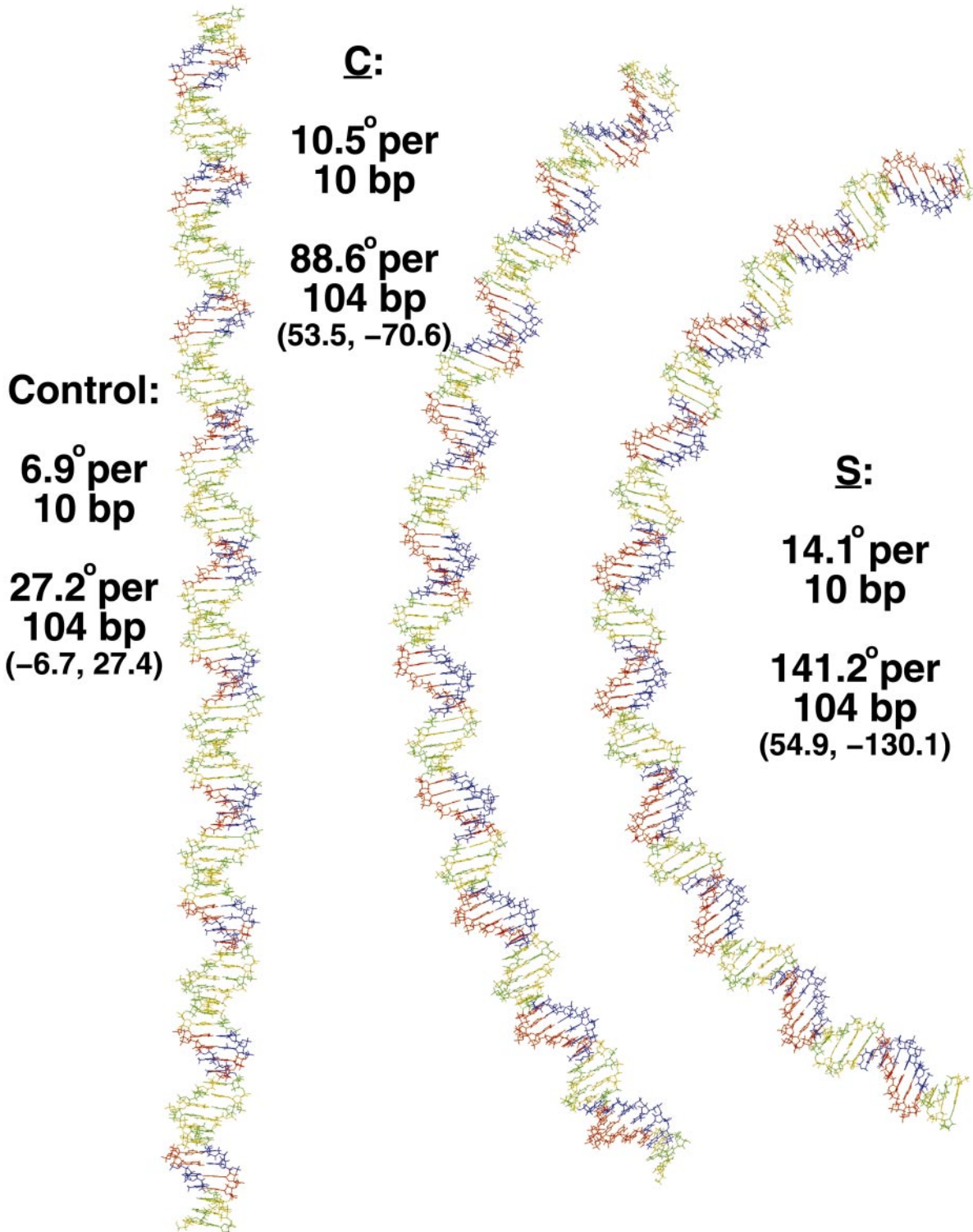
C:

10.5° per
10 bp

88.6° per
104 bp
(53.5, −70.6)

Control:

6.9° per
10 bp

27.2° per
104 bp
(−6.7, 27.4)

S:

14.1° per
10 bp

141.2° per
104 bp
(54.9, −130.1)

Figure 4. Large DNA models (104 bps) constructed from the dodecamer average structures to show the global effects of intrinsic curvature. Ten base-pair-segments of the average structures (base-pairs 2 to 11 and 14 to 23 in Figure 12) were assembled end to end and repeated ten times in phase with the helical repeat. Four additional nucleotides were added (at bp positions 21, 42, 63, and 84) to achieve a 10.5 bp helical repeat. *B*-DNA geometric parameters were assumed between repeats at each step of the assembly process by aligning the terminal bp of each repeat with idealized *B*-DNA coordinates. Adenosine nucleotides are red; thymidine, blue; guanosine, yellow; cytosine, green.
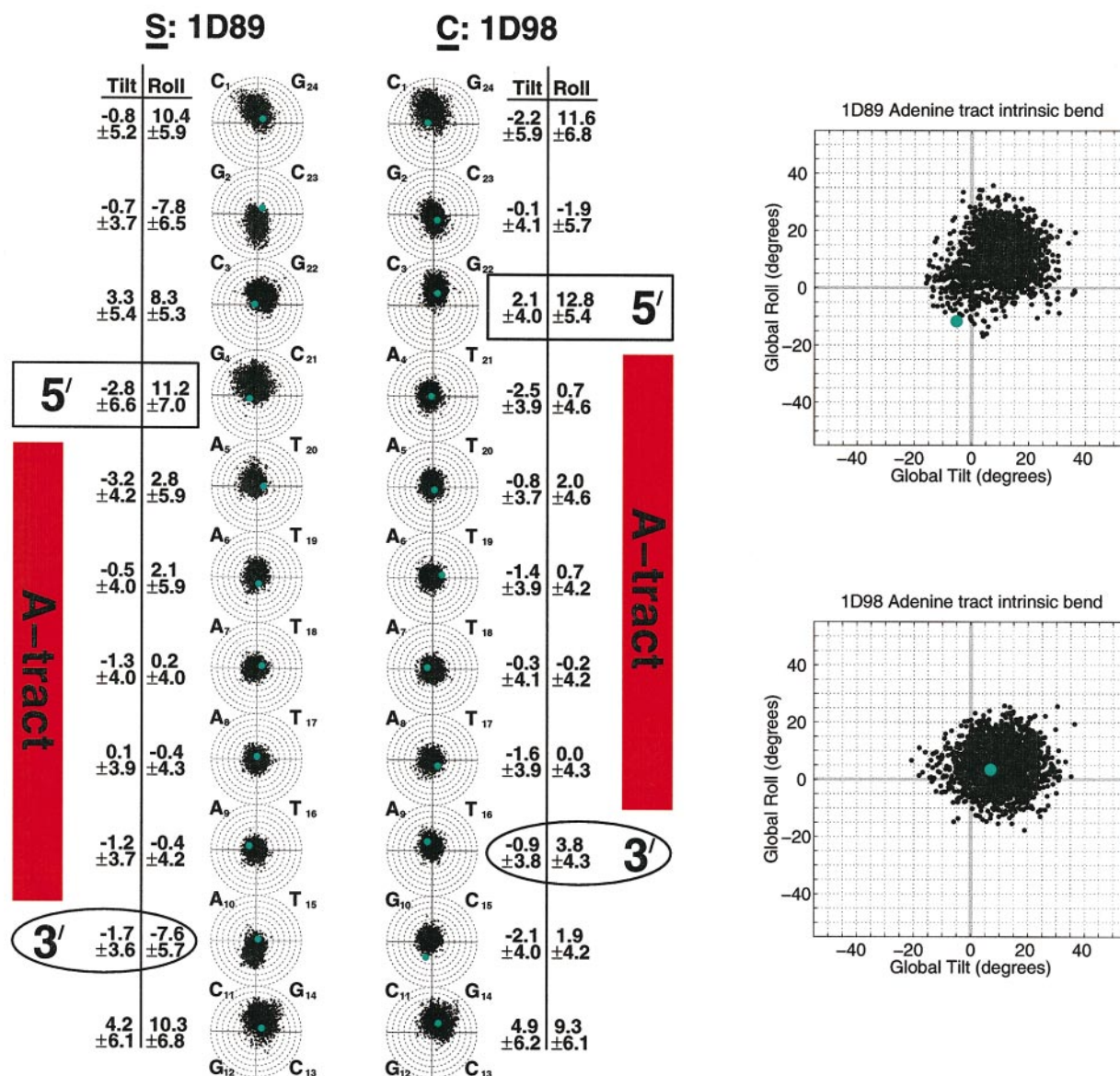
Figure 5. Local tilt and roll values (per bp step) in the A-tract dodecamers and the intrinsic A-tract bend angle. The MD ensembles and average values with one standard deviation are shown; boxed and circled values indicate the bp steps at the 5′ and 3′ ends of the A-tract, respectively. The local tilt and roll values for the crystallographic structures are indicated by teal circles; trajectory snapshots (black circles) were analyzed at a frequency of 1 per ps. The global tilt (θ_T) and global roll (θ_R) bend angles intrinsic to each A-tract were calculated as described in Equations (1) and (2a, b); only AA bp dinucleotide steps were included in calculating θ_T and θ_R. The bend angle of the crystallographic structures are indicated by teal circles.

to the explicit absence of π electron systems in the force field parameters (Cornell *et al.*, 1995).

Narrow A-tract minor groove

A-tract crystallographic structures have notably narrowed minor grooves (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993), which are also associated with propeller twisting (Fratini *et al.*, 1982). However, the multiple A-tract conformations present in crystallographic lattices led Steitz and co-workers to suggest that A-tract dodecamer structures were somewhat distorted (DiGabriele &

Steitz, 1993). Other experimental studies of solvated A-tract sequences support the presence of a narrowed A-tract groove region (Katahira *et al.*, 1988; Burkhoff & Tullius, 1987; Nadeau & Crothers, 1989). Harvey, Tullius, and co-workers have confirmed by gel electrophoresis and hydroxyl radical footprinting that the crystallographic dehydrating agent 2-methyl-2,4-pentane-diol (MPD) alters A-tract structures (Sprous *et al.*, 1995; Dlakic *et al.*, 1996; Ganunis *et al.*, 1996), but the nature of those modifications is not entirely understood.
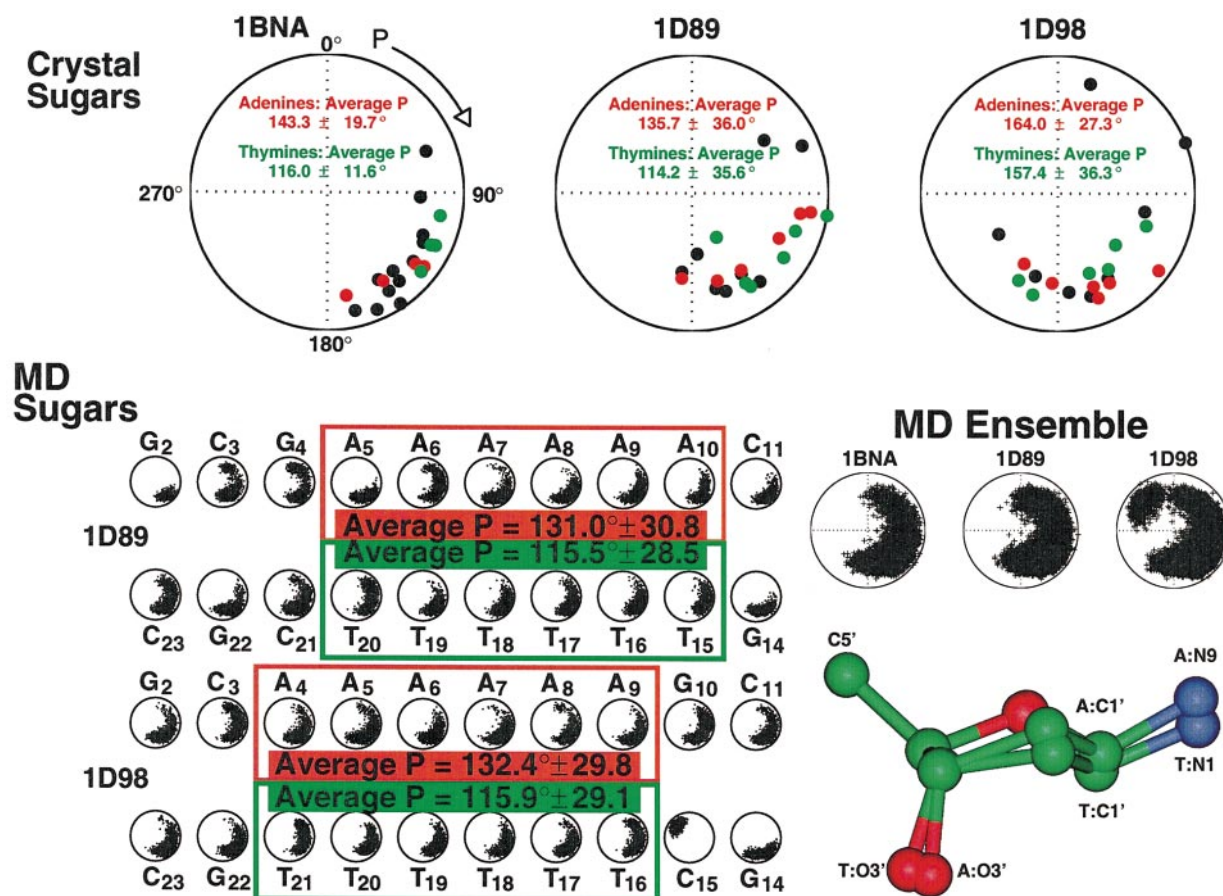
Figure 6. Phase/amplitude plots of crystallographic and MD ensembles of deoxyribose conformations. The pseudorotation phase angles P and amplitudes were calculated using the method described by Altona & Sundaralingam (1972); the endmost deoxyriboses have been disregarded in these calculations. The individual nucleotide values from the A-tract trajectories are plotted at bottom. Adenine deoxyriboses are indicated by red circles or boxes; thymine are indicated by green circles or boxes. A and T deoxyriboses with P values of 115° and 130°, respectively, are illustrated superimposed in the molecular model at lower right; the deoxyribose structural alignment minimizes the RMS deviation between the atoms C3′, C4′, and O4′ of the two sugar molecules.

We note that the minor groove widens slightly during the simulations relative to the crystallographic structure (Figure 7). Moreover, all three dodecamers have narrow minor grooves at regions of consecutive adenine residues. The minor groove of both A-tract dodecamers are asymmetric, whereas that of the control dodecamer is symmetric. The minimum width of the A-tract dodecamers is approximately 1 Å narrower than the control value (Figure 7). The A-tract pattern of narrowing is similar for the two dodecamers: the narrowing increases in the 5′ to 3′ direction from the first to third adenine and does not widen again until the end of the A-tract region. This pattern of minor-groove narrowing was predicted on the basis of NMR studies and hydroxyl radical footprinting of A-tract sequences (Katahira *et al.*, 1988; Burkhoff & Tullius, 1987; Nadeau & Crothers, 1989). Results of other MD simulations support similar minor-groove widths and narrowing patterns (Young & Beveridge, 1998; Sprous *et al.*, 1999; Young *et al.*, 1997).

A-tract stabilization by minor-groove water spine

Water spines were first observed in the minor groove of 1BNA, primarily associated with the central AT-rich region (Drew & Dickerson, 1981). The $\underline{S}$ crystal structure indicated the presence of an incomplete water spine within the A-tract (DiGabriele & Steitz, 1993), whereas analyses of the $\underline{C}$ oligomer did not resolve a water spine, presumably due to uncompensated lattice disorder (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993). Water spines are known to stabilize propeller twisting by forming bidentate hydrogen bonds crosslinking minor-groove acceptors (Crothers *et al.*, 1990; Shatzky-Schwartz *et al.*, 1997; Fratini *et al.*, 1982).

A water spine clearly emerges in our simulations of the A-tract dodecamers (Figure 8). The water spine state is well-occupied within the A-tract regions; specific spine sites are occupied in 50 % to 90 % of configurations analyzed. We find that both
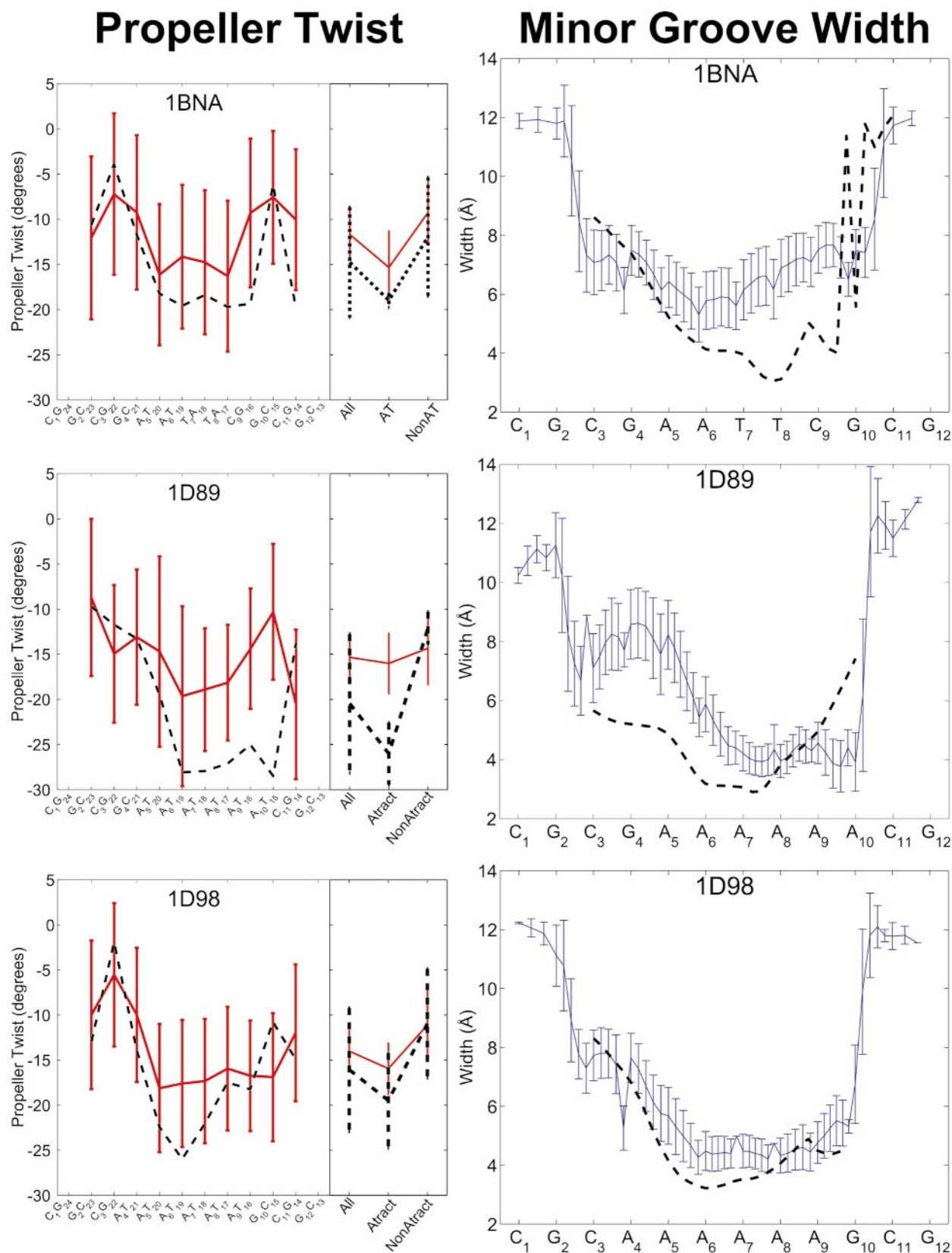
Figure 7. Average propeller twist (left) and minor-groove widths (right) of the three simulated dodecamers, bars indicating one standard deviation from the average, compared to the width of the crystallographic structure (heavy broken line). The propeller twist was calculated using Curves (Lavery & Sklenar, 1989, 1997). The central panel presents the average of all base-pairs (''All''), all AT base-pairs (''Atract''), and all non-AT base-pairs (''NonAtract''). The minor-groove width is calculated using the distances between cubic splines drawn tangential to the phosphate atoms (Stofer & Lavery, 1994).
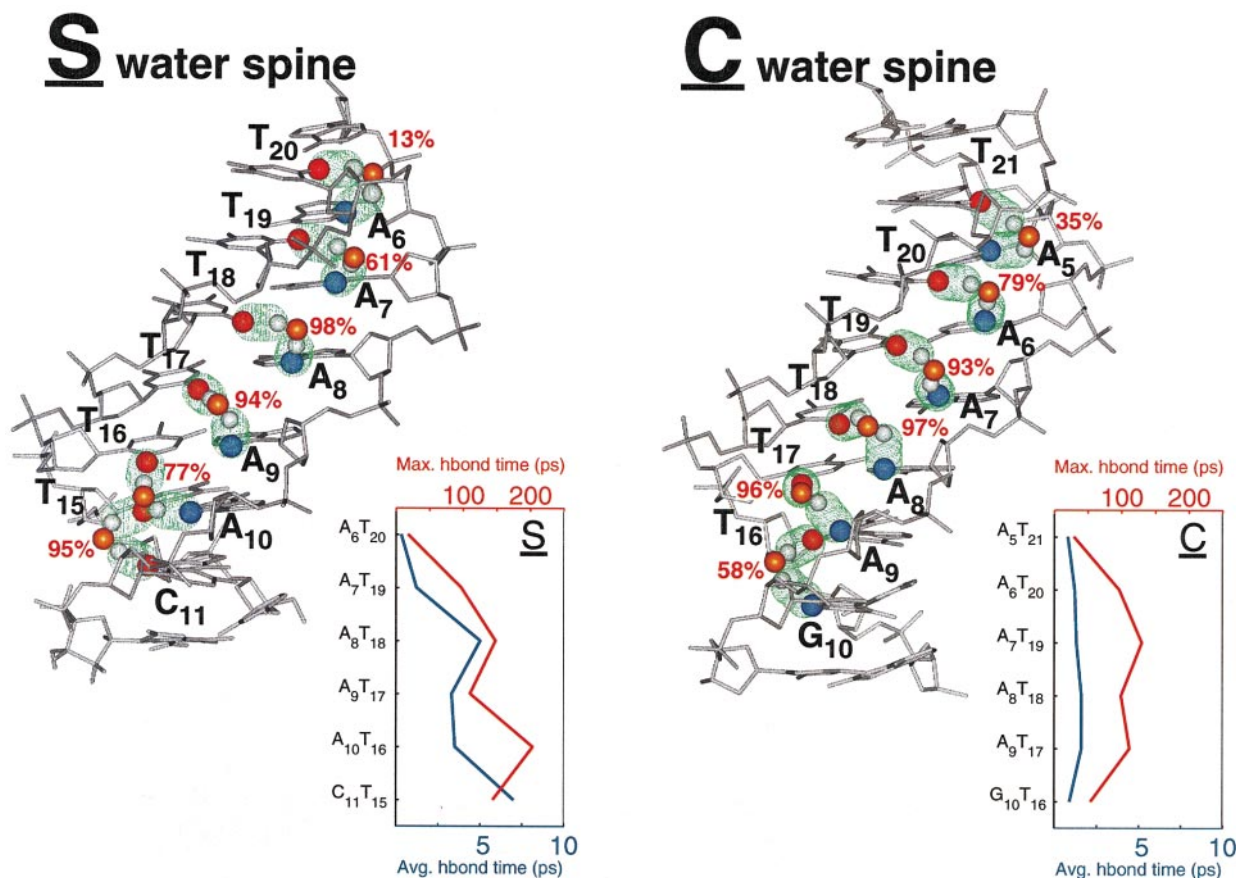
Figure 8. Minor groove water spine of the two A-tract dodecamers as viewed looking into the minor groove. Only water molecules in the primary hydration shell interacting within the A-tract and connecting the A-tract to the bp on the 3′ side are shown. Hydrogen bond interactions (green mesh surface) between the water molecule (oxygen, orange; hydrogen, white) and the adenine N3 (blue) and thymine O2 (red) atoms are represented. Next to each water position, the percentage of the 2 ns trajectory during which spine interactions were observed is indicated in red. The average (blue) and maximum (red) lifetimes of each spine position are plotted in the lower panels. The hydrogen bond lifetime was calculated by analyzing water molecules hydrogen bonded to both A:N3 and T:O2 atoms of adjacent bps at a frequency of 0.1 ps. Hydrogen bonds were scored using a distance cut-off between the acceptor/ hydrogen atoms of 3.1 Å and an angle criterion of 90° for the acceptor/hydrogen/oxygen atom sequence. Total hydrogen bond times were calculated as the sum of all interactions at a given water spine position.

the average and maximal water lifetimes increase as the 3′ end of the A-tract is approached (Figure 8), where the minor-groove width is narrowest (Figure 7). The geometrical considerations of water spines imply that base-pairs possess a negative propeller twist, although the precise level of propeller twist is variable (Shatzky-Schwartz *et al.*, 1997).

A novel feature that emerges is that both water spines extend in the 3′ direction beyond the A-tract to interact with one additional bp (Figure 8). In the S dodecamer, this extension forms an interaction between T_{15} and the O2 atom of C_{11} and, in the C dodecamer, between T_{16} and G_{10}. No extension is observed in the 5′ direction beyond the A-tract; the large positive roll at the major A-tract bend (Figure 5) and the widened minor groove (Figure 8) do not favor water spine interactions. These stabilizing interactions formed with the bp 3′ of the A-tract contribute to the structural features of this bp step.

Although a consistent trend, the water spine rapidly exchanges with bulk solvent. While occasional water molecules form spine interactions lasting as long as 0.2 ns, in agreement with recent estimates of analogous interactions of approximately 0.2 ns at 27° C (Phan *et al.*, 1999), average spine water molecules interact only briefly (less than 5 ps, Figure 8) with T:O2-A:N3 atom pairs. The average and maximal water oxygen residence times (defined by a 3.8 Å cut-off from the water molecule oxygen to A:N3/T:O2 atoms) are comparable to the calculated hydrogen bond times (data not shown). This indicates that water molecules are rapidly migrating to different interaction sites or to the secondary hydration shell after spine hydrogen bonds are broken (Figure 8). This fluidity of the water structure around DNA may be affected by dehydrating agents such as MPD (DiGabriele & Steitz, 1993; Sprous *et al.*, 1995; Dlakic *et al.*, 1996; Ganunis *et al.*, 1996).

Ion interactions with A-tracts

Although NMR (Hud & Feigon, 1997), crystallography (Tereshko *et al*., 1999), and simulation (Young & Beveridge, 1998; Young *et al*., 1997), have suggested a tendency of A-tract and related variants to bind cations within the minor groove, our simulations generally indicate only weak binding events for sodium ions which did not penetrate beyond the secondary hydration shell. An exception was one sodium ion coordinated between several water molecules bound to the N3 atom of A_5, located at the 5′ end of the $\underline{S}$ dodecamer A-tract. This interaction was captured four times during the 2 ns simulation, for a total time of 0.18 ns. These four events occurred during two longer periods (0.2 and 0.3 ns) with the ion shifting from the primary to the secondary hydration shell and back again. The multiple water molecules coordinated with this ion form a large complex; it is unlikely that this complex could bind in the narrower regions of the A-tract minor groove (Figure 7).

The position of this ion binding event at the 5′ end of the A-tract resembles manganese-binding events observed for A-tracts by NMR (Hud & Feigon, 1997). Localization around the largest bend site in A-tract oligonucleotides was also observed by Young & Beveridge (1998). Our ion interactions have a shorter duration than those observed for 1BNA ($\approx$0.2 *versus* 0.5 ns) (Young, *et al*., 1997), and a lower occupancy ($\approx$0.1 *versus* 0.5 fractional occupancy) (Tereshko *et al*., 1999); A-tracts may not provide the proper functional groups for long-term coordination of ions in the minor groove.

Discussion

Consensus structure and properties

The relationship between crystallographic and solution structures of A-tract DNA has been an area of intense interest. Since the crystallographic lattices of the $\underline{S}$ and $\underline{C}$ dodecamers are disordered, each lattice contains multiple conformations of the dodecamer (DiGabriele & Steitz, 1993). Disparities regarding bend directions are expected for different crystallographic structures (Figure 1). Despite the lattice disorder and the differences between bend directions, the A-tract moieties of the crystallographic structures are strikingly similar, with RMSD less than 1 Å (Dickerson *et al*., 1994; DiGabriele & Steitz, 1993). This agreement supports the notion that A-tract oligomers in solution possess a unique bend direction (Crothers *et al*., 1990); however, a unique bend direction clearly requires similarity extending beyond the A-tract region. The development of this similarity during our MD simulations merges these notions.

Our MD average structures for the two A-tract dodecamers superimpose well upon one another (RMS deviation less than 1 Å), indicating similarities extending beyond the A-tract (Figure 9). Each such average form is characterized by $\theta_R < 0$

(Figure 3), thus differing from the negative ($\underline{S}$) or positive ($\underline{C}$) global tilt in combination with small global roll associated with the initial structures. The new global bending framework introduced here helps clarify these bending trends; our program MaDBend is available *via* our web site (http://monod.biomath.nyu.edu).

Elements of these consensus structures include an adenine region with a gently curved helical axis, supported by a "wedge angle" of $\approx -1°$ tilt and $\approx 1°$ roll between each AA dinucleotide step. Thymine sugar molecules in the A-tract have an average pseudorotation of $P = 115°$, while the adenine sugar molecules have an average $P = 130°$ (Figure 6). This difference between thymine and adenine sugar molecules leads to a slight curvature intrinsic to the A-tract and the tilt/roll angles at each AA dinucleotide step. The narrow minor groove in the A-tract region progressively decreases in width in the 5′ to 3′ direction along the adenine strand (Figure 7) and is associated with approximately $16°$ of propeller twist at each A-tract bp, and a regular water spine (Figure 8). The bp step immediately on the 5′ side of the A-tract possesses a roll value of approximately $12°$ and a tilt of approximately $2°$ (Figure 5). In contrast, the roll and tilt values of the bp steps immediately 3′ of the A-tract regions are sequence dependent.

Our ensemble averages clearly indicate a bend direction of negative global roll (Figure 3), as predicted in solution (Crothers *et al*., 1990), but differing from the initial crystallographic bend directions (Nelson *et al*., 1987; DiGabriele & Steitz, 1993). Still, A-tract crystallographic structures currently emerging from the Shakked laboratory display negative global roll similar to our MD ensembles (Z. Shakked, personal communication). The changes leading to the bend conversion we observed are nonetheless subtle, as the average structures superimpose upon the crystallographic dodecamers with RMSD less than 3 Å (see Table 2). Each of the two MD average A-tract regions superimposes to the initial crystal structures with 1.5 Å RMSD, compared to a 0.9 Å difference between the two initial structures. Thus, the overall RMSD from the initial to the final structure is partitioned roughly equally between the A-tract and the non-A-tract dodecamer regions.

The detailed quantitative comparisons of our DNA and simulation results to experimental data, such as by hydroxyl radical cleavage (published on our web site http://monod.biomath.nyu.edu, see publications) and water lifetimes deduced from NMR (Figure 8), lend confidence in the interpretations made here regarding A-tract structure.

Interpretation of crystallographic A-tract structures

Our observed MD conversions suggest that small but subtle transformations have occurred which preserve the initial crystallographic models.
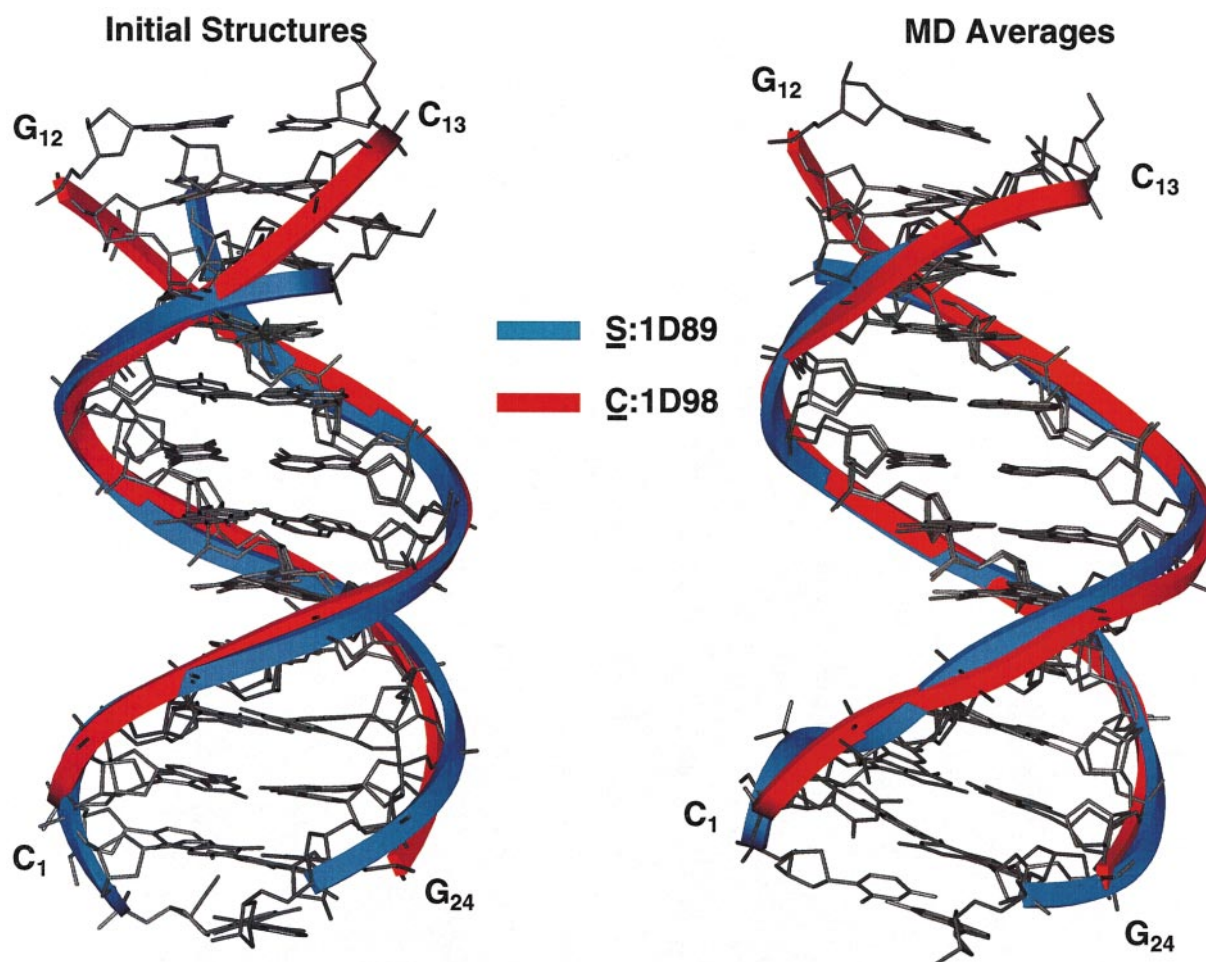
Figure 9. Superimposed crystallographic (left) and MD average (right) A-tract structures comparing the S̲ and C̲ dodecamers. The structures were superimposed minimizing the RMSD between A-tract regions. The blue and red backbone ribbons distinguish the S̲ and C̲ systems, respectively. Average structures are derived from the following trajectory segments: S̲: 309-2000 ps, C̲: 290-2000 ps.

These small differences include a widening of the minor grooves by ≈1 Å, some changes in average sugar conformations, and small angular changes for roll and tilt. Propeller twisting within the A-tracts exhibits a larger and consistent decrease of about 7° from the initial average value of 23°, while the twist decreases by about 2° per bp.

The overall similarity with the initial crystallographic models suggests that MPD effects are secondary and captured within the ensemble range of structures generated during the MD simulations (Dickerson *et al.*, 1994; DiGabriele & Steitz, 1993; Sprous *et al.*, 1995; Dlakic *et al.*, 1996; Ganunis *et al.*, 1996). We postulate that MPD exerts its effects through mechanisms affecting the DNA environment, such as the disruption of ordered water molecules (as observed in the minor-groove spine) or the chelation and removal of salt ions from solution (by formation of MPD/ion/water complexes, as observed in other diols). Ionic effects may also explain differences in bending magnitudes

observed for adenine-rich regions by crystallography (e.g., emerging from the Shakked laboratory) and simulation results (as reported here).

Relationship to models of A-tract bending

Our structural analyses possess elements that fit with many bending models available, such as "junction" (Crothers *et al.*, 1990; Koo & Crothers 1988; Haran *et al.*, 1994; Koo *et al.*, 1986), wedge (Trifonov, 1985; Ulanovsky *et al.*, 1986), and "curved general sequence" (Dickerson *et al.*, 1994, 1996).

Common features associated with the junction model are: (a) the A-tract dodecamers bend *via* negative global roll; (b) the A-tracts have a negative inclination to the global axis (Figure 10); and (c) the primary bend occurs at the 5′ end of the A-tract. This fit to the model is also manifested by the generally cooperative, nucleated structure of the A-tract region. The possible connection

Table 2. Average structural parameters from the dodecamer trajectories obtained from data sampled at a 1 ps frequency and deviations of MD-ensemble average structure from initial crystallographic structure

Property	1BNA	S	C
Avg. tilt (deg.)	0.1 ± 4.8	-0.1 ± 4.9	-0.4 ± 4.9
Avg. roll (deg.)	3.7 ± 7.1	2.7 ± 7.9	3.7 ± 7.0
Avg. twist (deg.)	31.7 ± 5.5	31.8 ± 8.7	32.0 ± 6.6
Avg. propeller twist (deg.)	-11.7 ± 8.8	-15.3 ± 10.9	-14.4 ± 9.3
Avg. x displacement (Å)	-1.9 ± 0.6	-1.4 ± 0.6	-2.0 ± 0.6
Avg. inclination (deg.)	2.0 ± 6.7	0.5 ± 8.3	2.7 ± 6.6
RMSD of average structure	2.68 Å	2.75 Å	2.63 Å

Nucleic-acid structural parameters were calculated using the Curves program (Lavery & Sklenar, 1989, 1997). In each dodecamer case, 9 bp steps were averaged, excluding the terminal bp steps. For the S dodecamer case, five trajectories (one Newtonian, four Langevin by LN) were averaged. Each trajectory produced 2000 snapshots. Average structures are derived from the following trajectory segments: 1BNA: 500-2000 ps, S: 309-2000 ps, C: 290-2000 ps.

between signs of inclination and the local A-tract bending angles might be further explored (Z. Shakked, personal communication).

Features that fit with the wedge-type bending models include: (a) the small but nearly uniform amounts of roll and tilt at the AA-dinucleotide steps within the A-tract; and (b) the similar bending geometry of the bp step at the 5′ end of the S and C dodecamer A-tracts.

Qualitative similarities between our results and characteristic of the curved general sequence model are noted from the smaller bends within each A-tract structure with respect to the bend on the 5′ side of the A-tract. Although not large, the A-tract unit curvature is an essential contributor to the negative global roll bend direction. We also observe flexing through roll at the junctions between the A-tracts and the surrounding sequence, as predicted by this model. The narrow minor grooves and large amounts of propeller

twist observed in our simulations and those of Young et al. (Young & Beveridge, 1998; Sprous et al., 1999) further concur with the curved general sequence model. However, observations that contradict the model include the similar stiffness between the A-tract bp steps and those outside the A-tract, and the negative inclination of the A-tract to the global helical axis (Dickerson et al., 1994).

In conclusion, our consensus model for A-tract structure incorporates various elements of several proposed models (Crothers et al., 1990; Dickerson et al., 1994, 1996; Koo & Crothers 1988; Haran et al., 1994; Koo et al., 1986; Trifonov, 1985; Ulanovsky et al., 1986) and extends and supports many experimental and theoretical studies.

Relationship to TATA-box elements

Current work in our laboratory is revealing additional contextual features of A-tract structure and their effect on biological processes. One interesting example under study is the selectivity of the TATA-box binding protein (TBP) on variants of the target DNA sequence (X. Qian, D. S. & T. S., unpublished results). TBP binds to many TATA-box variants; the resulting complexes have broadly varying levels of transcriptional efficiencies (Patikoglou et al., 1999). Our simulations reinforce the notion that the wild-type TATA-box element 5′-TATAAAAG-3′ (AdMLP) found in adenovirus particles has been selected by evolution because of its inherent flexibility.

Results from our simulations (to be detailed in a forthcoming paper) indicate that the AdMLP TATA-box element bends flexibly into the DNA major groove (Figure 11) and that a relatively disordered water/DNA interface further facilitates this motion and hence TBP binding. In contrast, the A-tract variant 5′-TAAAAAAG-3′ (A29), which could not be co-crystallized with TBP (Patikoglou et al., 1999), has been experimentally determined to be transcriptionally inactive. Our MD simulations have shown not only a marked difference in the intrinsic DNA flexibility of A29 relative to AdMLP, but also a bending tendency in the opposite direc-
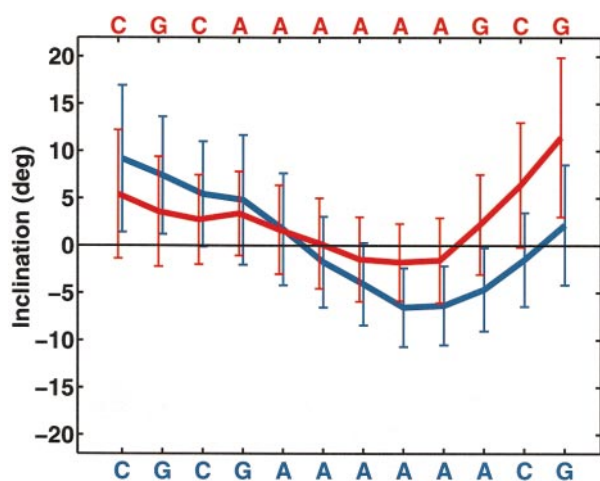

Figure 10. Average inclination to the S (blue) and C (red) global helical axes as computed at each base-pair from 2000 snapshots spanning 2 ns. The mean is plotted as a continuous line; one standard deviation is indicated by the bars.
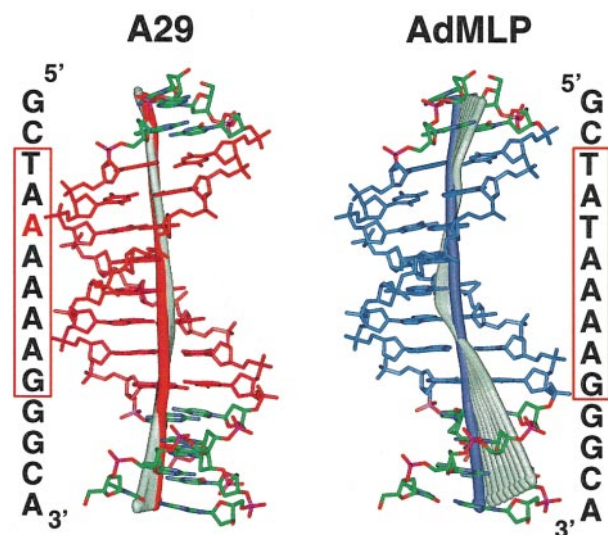
Figure 11. Average structures over 1.5 ns from simulations of the TATA-box variants A29 (left) and wild-type AdMLP (right) and principal component analyses of bending. The TATA-box elements (eight bps) are indicated in red (A29) and blue (AdMLP) DNA. The helical axes of the average structures, computed by Curves (Lavery & Sklenar, 1989, 1997), are shown in thick red (A29) and blue (AdMLP); the ten green curves in each case illustrate the observed range of motion along the first principal component. We note that the A-tract variant A29 bends towards the minor groove, while the AdMLP wild-type variant bends flexibly towards the major groove, as needed for TBP binding and activity.

tion (i.e., into the minor groove) with respect to that required for TBP binding.

Shown in Figure 11 is the ensemble average structures (over 1.5 ns) of the wild-type AdMLP sequence and the A29 variant, each with helical axes indicating the motion associated with the first principal component (groove bending). The different bending tendencies of the wild-type sequence (toward the major-groove) and the A-tract variant (toward the minor-groove) are apparent from the average helical axes. In a similar way, the large major-groove bending motion in the former *versus* the gentler minor-groove motion in the latter is evident from the principal component analysis. We also noted that A29, like other A-tract sequences, has a more ordered water spine than AdMLP and a higher concentration of counterions in regions that interfere with TBP binding and deformation.

Given the TATA-box consensus element octamer of TATA(a/t)A(a/t/g)R, where R indicates a purine base and the bases at positions 5 and 7 are variable, these striking sequence-dependent TATA-box variations suggest that the easily achieved single bp changes that convert a normal TATA-box element into an A-tract can act as a powerful biological regulator.

Methods

System preparation

Atomic coordinates for the DNA dodecamers 1BNA, S̲, and C̲ were obtained from the RCSB Protein Data Bank (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993; Wing *et al.*, 1980; Bernstein *et al.*, 1977). Crystallographically indicated water molecules were maintained with each dodecamer (80, 8, and 27 water molecules included with 1BNA, S̲, and C̲, respectively). Neutralization of charge requires one cation per phosphate ion; sodium ions were placed 5 Å from the phosphate groups along the O_1-P-O_2 bisector, for a total of 22 sodium ions per dodecamer. The S̲ and C̲ DNA/ion systems were overlaid with a pre-equilibrated box containing 11,500 TIP3P-modeled water molecules (Jorgensen *et al.*, 1983) (as implemented in the CHARMM molecular modeling package, version 2.5α3 (Brooks *et al.*, 1983)). Water molecules overlapping the DNA (defined as solvent with a non-hydrogen atom within 1.8 Å of DNA non-hydrogen atoms) were removed from the system. The systems were truncated to form regular hexagonal prisms with 10.0 Å between the DNA and the closest boundary of the periodic cell; this procedure generated regular hexagonal prisms of 65.0 Å length × 26.0 Å side for the S̲ system and 65.5 Å length × 25.1 Å side for the C̲ system. The 1BNA system was solvated using the Insight molecular graphics package, version 97.2. Excess water molecules were eliminated to generate a regular hexagonal prism of 65.5 Å length × 25.3 Å side. Final system sizes are 10,869, 12,062, and 11,474 atoms for the 1BNA, S̲ and C̲ dodecamers, respectively.

Minimization and simulations with the Leapfrog integrator

Energy minimization and MD simulations use the Cornell *et al.* (1995) force field as implemented in the CHARMM program (Brooks *et al.*, 1983). While the DNA and sodium ions in the cells were constrained in place, the water molecules in the system were subjected to minimization using the adopted-basis Newton-Raphson protocol and equilibrated at 300 °C using a Leapfrog integrator with a timestep of 2 fs. Non-bonded interactions were calculated using atom-based neighbor lists truncated at 12 Å; Lennard-Jones switch and electrostatic force shift functions were applied in the 10-12 Å range. SHAKE-type constraints were applied to all bonds containing hydrogen (Ryckaert *et al.*, 1977). After equilibration of the initial geometry, the entire system was minimized and reheated to 300 °C over 10 ps using protocols as described above. The three systems were equilibrated for 30 ps and data were collected for a period of 2 ns. All calculations were performed on a Silicon Graphics Power Challenge XL at New York University.

Simulations with the LN integrator

A recently developed multiple timestepping integrator for Langevin dynamics, termed LN, was also used to investigate the S̲ system. The LN integrator offers speed-ups of between 4 and 13-fold, allowing multiple ensembles or longer ensembles to be generated at a smaller cost; this speedup advantage allowed us to improve statistics and examine the S̲ system in greater detail *via* four additional independent simulations, in addition to the Verlet trajectory. The LN integrator uses stochastic

dynamics merged with extrapolation of slowly varying forces and a three-cycle integration (Barth & Schlick, 1998a,b; Sandu & Schlick, 1999). The shortest timestep represents the frequency used to evaluate the bonded interactions: bond length, bond angle, dihedral angles (both proper and improper), and Urey-Bradley bond lengths. The medium timestep is the frequency of evaluating non-bonded interactions within a specified distance (typically chosen in the 6-8 Å range). The outermost timestep is the frequency of calculating remaining non-bonded interactions and updating all non-bonded interactions, determined by the truncation limit. The LN simulations described here used inner, medium and outer timesteps of 1, 2 and 120 fs (1/2/120 fs), respectively. The medium force class was defined by a distance range of 7 Å associated with healing and buffer lengths of 4 Å each (7/4/4 Å). A relatively small Langevin collision parameter of $\gamma = 5$ ps^{-1} was chosen to couple the $\underline{S}$ system to a 300 °C Langevin heat bath. As shown by Barth & Schlick, a smaller γ generates trajectories closer to Newtonian trajectories; γ in the 5 ps^{-1} range also ensures stability by masking resonances (Barth & Schlick, 1998a,b; Sandu & Schlick, 1999). Four independent replicates (Caves et al., 1998) of the $\underline{S}$ system were initiated with different random number seeds and data were collected for a period of 2 ns.

Trajectory assessment

All dodecamer simulations equilibrated rapidly, reaching RMSD relative to the initial crystallographically-derived conformation of approximately 3 Å within the first 200 ps (Figure 12). The RMSD value for all dodecamers compare favorably to other DNA simulations using particle mesh Ewald techniques (de Souza & Ornstein, 1997; York et al., 1995; Cheatham et al., 1995; Young et al., 1997). After the initial 100 ps of simulation, all trajectories remain at a conformational distance of 2.5 Å with respect to the initial crystallographic conformation.

Average MD-ensemble structures generated from the trajectory are superimposable upon the original crystal structures with RMS deviation values less than 3 Å (see Table 2). As discussed below, roughly half this difference comes from the A-tract region. Average nucleic acid structural parameters (tilt, roll, etc.) derived from these trajectories verify that the structures and properties of the simulated dodecamers generated either by Newtonian or stochastic dynamics lie within the family of B-DNA structures (Table 2). It is noteworthy that our approach (a large domain in which the DNA is solvated by a 10 Å layer of water, non-bonded interactions are computed within a 12 Å cut-off range, and in which periodic atom lists are maintained within a 14 Å limit) yields stable trajectories that are comparable to a 5 ns trajectory with the particle mesh Ewald method (Young et al., 1997). Thus, such protocols are reliable for the properties examined here, as are particle mesh Ewald schemes (Darden et al., 1999).

The LN integrator used for some of the trajectories here demonstrates the improved sampling statistics as obtained with respect to single-timestep trajectories using the same computation time. This sampling advantage can be gleaned from Figure 3. The four pooled LN trajectories, each of length 2 ns, took a combined CPU time of 92 days, versus the single trajectory of length 65 days in comparison.

Although the LN simulations used the smaller inner timestep of 1 fs as opposed to the 2 fs step size used in the Leapfrog simulations, the LN simulations were overall faster than the Leapfrog simulations by a factor of three for simulating the same total time. That is, the LN simulations completed 1 ns in 23 days as opposed to the 65 days necessary for a Leapfrog integrator with a 2 fs timestep or the 98 days necessary for a Langevin (BBK) integrator using a single 1 fs timestep. An LN simulation without SHAKE requires a smaller inner timestep, such as 0.5 fs; a 0.5/2/120 fs multiple-timestep protocol for LN would have a speedup factor of approximately 9.4, completing a 1 ns simulation of the $\underline{S}$ system in 16 days as opposed to the 151 days for a Langevin (BBK) integrator using a single 0.5 fs timestep. This order of magnitude speedup correlates with values reported previously (Barth & Schlick, (1998a,b; Sandu & Schlick, 1999). Other macromolecular applications of LN to large systems can yield greater speedups (Schlick et al., 2000). The LN integrator is included in CHARMM version 2.7b1 (Brooks et al., 1983) and is being implemented in AMBER.

Structure analysis

Generalized nucleic acid structural parameters (such as roll, tilt and inclination) were derived using the Curves software package, version 5.2 (Lavery & Sklenar, 1989, 1997; Stofer & Lavery, 1994). Analytical tools within CHARMM were used to calculate various quantities such as distances, angles, deoxyribose puckers and RMS deviations of superimposed structures. Surface area was calculated using the Lee & Richards (1971) algorithm as implemented in CHARMM.

Global parameters for analysis of helical axis bend

Our procedure calculates DNA curvature by summing the projected components of local bp step tilt and roll angles, after adjusting for helical twist. Each term in the sum corresponds to a bp step j. First, we derive γ_j, the accumulated twist from the reference plane N_c to bp step j; $\gamma_j = \Sigma_{i = N_c}^{j} \Omega_i$ where Ω_j is the twist at bp step j (defining the rotation from bp j to bp $j + 1$ (Dickerson et al., 1989). Second, we rotate the local tilt (τ_j) and roll (ρ_j) angles (defining the angular deformations from bp j to $j + 1$ (Dickerson et al., 1989) through the angle γ_j; this can be compactly expressed as the matrix/vector product $R(\gamma_j)\theta_j$, where $R(\gamma_j)$ is the rotation matrix defined by γ_j:

$$R(\gamma_j) = \begin{pmatrix} \cos \gamma_j & \sin \gamma_j \\ -\sin \gamma_j & \cos \gamma_j \end{pmatrix},$$

and θ_j is the column vector $\theta_j = (\tau_j, \rho_j)^T$. Writing the two angles, global tilt (θ_T) and global roll (θ_R), as the vector $\Theta = (\theta_T, \theta_R)^T$, we have:

$$\Theta = \sum_j R(\gamma_j)\theta_j \tag{1}$$

or:

$$\theta_T = \sum_j (\tau_j \cos \gamma_j + \rho_j \sin \gamma_j) \tag{2a}$$

$$\theta_R = \sum_j (-\tau_j \sin \gamma_j + \rho_j \cos \gamma_j) \tag{2b}$$

In our right-handed coordinate system, the sign of twist above the reference plane is positive, and that below the reference plane is negative. If the reference plane lies
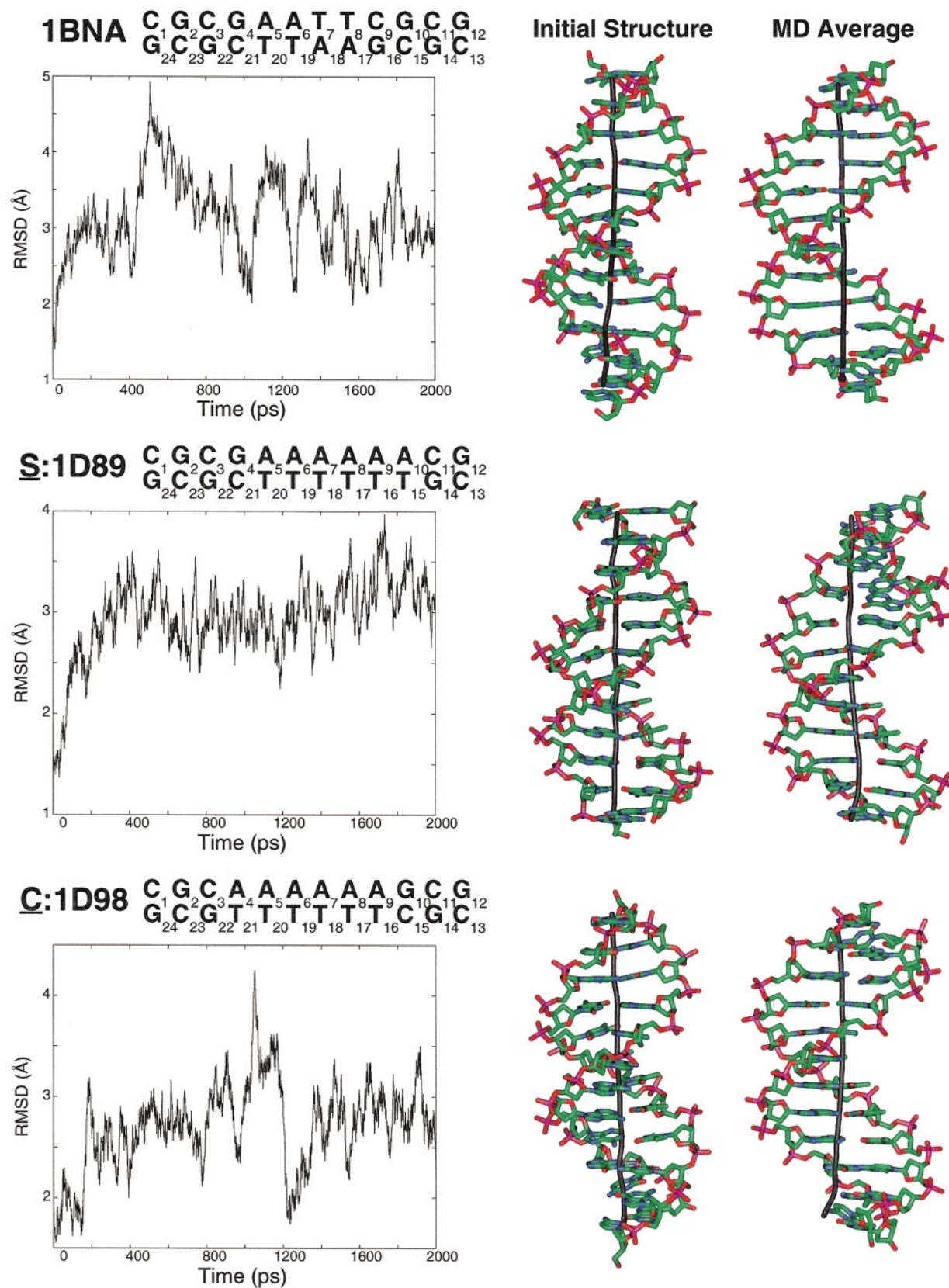
1BNA   C₁ G₂ C₃ G₄ A₅ A₆ T₇ T₈ C₉ G₁₀ C₁₁ G₁₂
            G₂₄ C₂₃ G₂₂ C₂₁ T₂₀ T₁₉ A₁₈ A₁₇ G₁₆ C₁₅ G₁₄ C₁₃

Initial Structure   **MD Average**

S:1D89   C₁ G₂ C₃ G₄ A₅ A₆ A₇ A₈ A₉ C₁₀ G₁₁ C₁₂
              G₂₄ C₂₃ G₂₂ C₂₁ T₂₀ T₁₉ T₁₈ T₁₇ T₁₆ T₁₅ G₁₄ C₁₃

C:1D98   C₁ G₂ C₃ A₄ A₅ A₆ A₇ A₈ A₉ G₁₀ C₁₁ G₁₂
              G₂₄ C₂₃ G₂₂ T₂₁ T₂₀ T₁₉ T₁₈ T₁₇ T₁₆ C₁₅ G₁₄ C₁₃

Figure 12. RMSD of superimposed structures from the crystallographic conformation (left) and selected structures (right) corresponding to the initial crystallographic structure and the MD average trajectory form. The crystallographic and average structures were superimposed using non-hydrogen atoms; both structures were rotated so that the minor groove at the A-tract (or dodecamer) center is oriented to the left. Global helical axes produced by the program Curves 5.2 are shown at the center of each structure in black (Lavery & Sklenar, 1989, 1997). A standard numbering scheme for nucleotide labeling is indicated for each system (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993; Wing *et al.*, 1980). Note the slight bend towards the minor groove in the helical axes of the S and C average structures at right.

midway between two bps, τ_j, ρ_j and Ω_j are halved for the two terms in step $j = N_c$. This procedure can be generalized to a reference plane located at any position between two bps.

To illustrate the use of the algorithm, consider a 6 bp oligonucleotide and a reference plane chosen at the oligomer center (i.e., between bp 3 and 4); θ_T and θ_R are obtained as follows:

$$\Theta = \frac{1}{2}R(\gamma_3/2)\theta_3 + R(\gamma_4)\theta_4 + R(\gamma_5)\theta_5$$
$$+ \frac{1}{2}R(-\gamma_3/2)\theta_3 + R(-\gamma_2)\theta_2 + R(-\gamma_1)\theta_1$$

If the reference plane coincides with the third base-pair, the formula instead becomes:

$$\Theta = R(\gamma_3)\theta_3 + R(\gamma_4)\theta_4 + R(\gamma_5)\theta_5$$
$$+ R(-\gamma_2)\theta_2 + R(-\gamma_1)\theta_1$$

Our analysis for the global angles is based on computed local tilt, roll and twist angles by the program Curves (Lavery & Sklenar, 1989). Other nucleic acid analysis programs may be used. It is well appreciated that different programs can produce different results (Babcock & Olson, 1994); hence, the global values can differ.

The utility for computing the global tilt and global roll is available from the authors and can be downloaded at http://monod.biomath.nyu.edu

Bends in the helical axis defined by $\theta_R < 0$ are equivalent to bending towards the minor groove in the local reference frame, as seen in Figure 1; bends defined by $\theta_R > 0$ correspond to bending towards the major groove. The total magnitude of the curvature is

$$\sqrt{\theta_T^2 + \theta_R^2}.$$

For our three oligonucleotides, we analyze θ_T and θ_R in the reference plane at the A-tract center (bp steps $\underline{S}:A_7A_8$ and $\underline{C}:A_6A_7$) or the central A_6T_7 bp step in the control 1BNA dodecamer. The A-tract reference planes are implied by the measurements reported by Crothers et al. (Zinkel & Crothers, 1987; Koo & Crothers 1988). The A-tract of the $\underline{S}$ dodecamer is located 1 bp askew from the oligonucleotide end; hence, the reference plane is offset by approximately 1 bp relative to the center of the dodecamer.

References

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

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.

Barth, E. & Schlick, T. (1998a). Overcoming stability limitations in biomolecular dynamics. I. Combining force splitting *via* extrapolation with Langevin dynamics in LN. *J. Chem. Phys.* **109**, 1617-1632.

Barth, E. & Schlick, T. (1998b). Extrapolation *versus* impulse in multiple-timestepping schemes. II. Linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.* **109**, 1633-1642.

Beamer, L. J. & Pabo, C. O. (1992). Refined 1.8 Å crystal structure of the lambda repressor-operator complex. *J. Mol. Biol.* **227**, 177-196.

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

Bossi, L. & Smith, D. M. (1984). Conformational change in the DNA associated with an unusual promoter mutation in a tRNA operon of Salmonella. *Cell*, **39**, 643-652.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.

Burkhoff, A. M. & Tullius, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell,* **48**, 935-943.

Caves, L. S. D., Evanseck, J. D. & Karplus, M. (1998). Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **7**, 649-666.

Cheatham, T. E., III, Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193-4194.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.

Crothers, D. M., Haran, T. E. & Nadeau, J. G. (1990). Intrinsically bent DNA. *J. Biol. Chem.* **265**, 7093-7096.

Darden, T., Perera, L., Li, L. & Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure,* **7**, R55-R60.

de Souza, O. N. & Ornstein, R. L. (1997). Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophys. J.* **72**, 2395-2397.

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

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). ''. . . the tyranny of the lattice. . . ''. *Proc. Natl Acad. Sci. USA,* **91**, 3579-3583.

Dickerson, R. E., Goodsell, D. S. & Kopka, M. L. (1996). MPD and DNA bending in crystals and in solution. *J. Mol. Biol.* **256**, 108-125.

DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. (1989). Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl Acad. Sci. USA,* **86**, 816-820.

DiGabriele, A. D. & Steitz, T. A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024-1039.

Dlakic, M., Park, K., Griffith, J. D., Harvey, S. C. & Harrington, R. E. (1996). The organic crystallizing agent 2-methyl-2,4-pentanediol reduces DNA curvature by means of structural changes in A-tracts. *J. Biol. Chem.* **271**, 17911-17919.

Drew, H. R. & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535-556.

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

Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 14686-14707.

Ganunis, R. M., Guo, H. & Tullius, T. D. (1996). Effect of the crystallizing agent 2-methyl-2,4-pentanediol on the structure of adenine tract DNA in solution. *Biochemistry,* **35**, 13729-13732.

Gartenberg, M. R. & Crothers, D. M. (1991). Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. *J. Mol. Biol.* **219**, 217-230.

Hagerman, P. J. (1992). Straightening out the bends in curved DNA. *Biochim. Biophys. Acta,* **1131**, 125-132.

Haran, T. E., Kahn, J. D. & Crothers, D. M. (1994). Sequence elements responsible for DNA curvature. *J. Mol. Biol.* **244**, 135-143.

Harvey, S. C., Dlakic, M., Griffith, J., Harrington, R., Park, K., Sprous, D. & Zacharias, W. (1995). What is the basis of sequence-directed curvature in DNAs containing A-tracts? *J. Biomol. Struct. Dynam.* **13**, 301-307.

Hud, N. V. & Feigon, J. (1997). Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Am. Chem. Soc.* **119**, 5756-5757.

Jorgensen, W. L., Chandrasekhar, J. & Madura, J. D. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.

Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R. & Tomita, K. (1988). One and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA) oligo(dT) tract. *Nucl. Acids Res.* **16**, 8619-8632.

Kim, J., Klooster, S. & Shapiro, D. J. (1995). Intrinsically bent DNA in a eukaryotic transcription factor recognition sequence potentiates transcription activation. *J. Biol. Chem.* **270**, 1282-1288.

Kintanar, A., Klevit, R. E. & Reid, B. R. (1987). Two-dimensional NMR investigation of a bent DNA fragment: assignment of the proton resonances and preliminary structure analysis. *Nucl. Acids Res.* **15**, 5845-5862.

Koo, H. S. & Crothers, D. M. (1988). Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl Acad. Sci. USA,* **85**, 1763-1767.

Koo, H. S., Wu, H. M. & Crothers, D. M. (1986). DNA bending at adenine-thymine tracts. *Nature,* **320**, 501-508.

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

Lavery, R. & Sklenar, H. (1997). *Curves 5.2: Helical Analysis of Irregular Nucleic Acids*, Laboratoire de Biochimie Theorique, CNRS URA 77, Institut de Biologie Physico-Chimique, Paris, France.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Nadeau, J. G. & Crothers, D. M. (1989). Structural basis for DNA bending. *Proc. Natl Acad. Sci. USA,* **86**, 2622-2626.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature,* **330**, 221-226.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA,* **93**, 4862-4867.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S.-H., Kodadek, T. & Burley, S. K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217-3230.

Phan, A. T., Leroy, J.-L. & Guéron, M. (1999). Determination of the residence time of water molecules hydrating *B'*-DNA and *B*-DNA, by one-dimensional zero-enhancement nuclear Overhauser effect spectroscopy. *J. Mol. Biol.* **286**, 505-519.

Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327-341.

Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. & Tegtmeyer, P. (1986). An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. *Cell,* **44**, 719-725.

Sandu, A. & Schlick, T. (1999). Masking resonance artifacts in force-splitting methods for biomolecular simulations by extrapolative Langevin dynamics. *J. Comp. Phys.* **151**, 74-113.

Schlick, T. (1998). Some failures and successes of long-timestep approaches for biomolecular simulations. In *Computational Molecular Dynamics: Challenges, Methods, Ideas - Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21-24, 1997* (Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A. E. & Skeel, R. D., eds), vol. 4, pp. 227-262, *Lecture Notes in*

Computational Science and Engineering, Springer-Verlag, Berlin and New York.

Schlick, T. (1999). Computational molecular biophysics today: a confluence of methodological advances and complex biomolecular applications. *J. Comp. Phys.* **151**, 1-8.

Schlick, T., Beard, D. A., Huang, J., Strahs, D. & Qian, X. (2000). Computational challenges in simulating large DNA over long times. *IEEE Comp. Sci. Eng.* In the press.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science,* **253**, 1001-1007.

Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F. & Shakked, Z. (1997). X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.* **267**, 595-623.

Snyder, M., Buchman, A. R. & Davis, R. W. (1986). Bent DNA at a yeast autonomously replicating sequence. *Nature,* **324**, 87-89.

Sprous, D., Zacharias, W., Wood, Z. A. & Harvey, S. C. (1995). Dehydrating agents sharply reduce curvature in DNAs containing A tracts. *Nucl. Acids Res.* **23**, 1816-1821.

Sprous, D., Young, M. A. & Beveridge, D. L. (1999). Molecular dynamics studies of axis bending in d(G_5-(GA_4T_4C)$_2$-C_5) and d(G_5-(GT_4A_4C)$_2$-C_5): effects of sequence polarity on DNA curvature. *J. Mol. Biol.* **285**, 1623-1632.

Stofer, E. & Lavery, R. (1994). Measuring the geometry of DNA grooves. *Biopolymers,* **34**, 337-346.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for con-stitutive transcription in yeast. *Proc. Natl Acad. Sci. USA,* **82**, 8419-8423.

Tereshko, V., Minasov, G. & Egli, M. (1999). A ''Hydrat-Ion'' spine in a B-DNA minor groove. *J. Am. Chem. Soc.* **121**, 3590-3595.

Trifonov, E. N. (1985). Curved DNA. *CRC Crit. Rev. Biochem.* **19**, 89-106.

Ulanovsky, L., Bodner, M., Trifonov, E. N. & Choder, M. (1986). Curved DNA: design, synthesis, and circularization. *Proc. Natl Acad. Sci. USA,* **83**, 862-866.

Ulanovsky, L. & Trifonov, E. N. (1987). Estimation of wedge components in curved DNA. *Nature,* **326**, 720-722.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of *B*-DNA. *Nature,* **287**, 755-758.

York, D. M., Yang, W., Lee, H., Darden, T. & Pedersen, L. G. (1995). Toward the accurate modeling of DNA; the importance of long-range electrostatics. *J. Am. Chem. Soc.* **117**, 5001-5002.

Young, M. A. & Beveridge, D. L. (1998). Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* **281**, 675-687.

Young, M. A., Ravishanker, G. & Beveridge, D. L. (1997). A 5-nanosecond molecular dynamics trajectory for *B*-DNA: analysis of structure, motions, and solvation. *Biophys. J.* **73**, 2313-2336.

Zinkel, S. S. & Crothers, D. M. (1987). DNA bend direction by phase sensitive detection. *Nature,* **328**, 178-181.

Edited by Dr I. Tinoco