K. Suzie Byun
D. L. Beveridge
Department of Chemistry, and Molecular Biophysics Program, Wesleyan University, Middletown, CT 06459

Received 1 April 2003; accepted 15 April 2003

Molecular Dynamics Simulations of Papilloma Virus E2 DNA Sequences: Dynamical Models for Oligonucleotide Structures in Solution

Abstract: *The specificity of papilloma virus E2 protein–DNA binding depends critically upon the sequence of a region of the DNA not in direct contact with the protein, and represents one of the simplest known examples of indirect readout. A detailed characterization of this system in solution is important to the further investigation hypothesis of a structural code for DNA recognition by regulatory proteins. In the crystalline state, the E2 DNA oligonucleotide sequence, d(ACCGAATTCGGT), exhibits three different structural forms. We report herein studies of the structure of E2 DNA in solution based on a series of molecular dynamics (MD) simulations including counterions and water, utilizing both the canonical and various crystallographic structures as initial points of departure. All MDs converged on a single dynamical structure of d(ACCGAATTCGGT) in solution. The predicted structure is in close accord with two of the three crystal structures, and indicates that a significant kink in the double helix at the central A_pT step in the other crystal molecule may be a packing effect. The dynamical fine structure was analyzed on the basis of helicoidal parameters. The calculated curvature in the sequence was found to originate primarily from Y_PR steps in the regions flanking the central AATT tract. In order to study the role of structural adaptation of the DNA in the binding process, a subsequent simulation on the 16-mer cognate sequence d(CAACCGAATTCGGTTG) was initiated from the crystallographic coordinates of the bound DNA in the crystal structure of the protein DNA complex. MD simulations starting with the protein-bound form relaxed rapidly back to the dynamical structure predicted from the previous simulations on the uncomplexed DNA. The MD results show that the bound form E2 DNA is a dynamically unstable structure in the absence of protein, and arises as a consequence of both structural changes intrinsic to the sequence and induced by the interaction with protein.* Biopolymers 73: 369–379, 2004

Keywords: *Papilloma virus E2; protein–DNA binding; DNA recognition; regulatory proteins; molecular dynamics*

Correspondence to: K. Suzie Byun (email: kbyun@wesleyan.edu); or D. L. Beveridge (email: dbeveridge@wesleyan.edu).

Contract grant sponsor: NIH; Contract grant number: GM-37909

Biopolymers, Vol. 73, 369–379 (2004)

INTRODUCTION

The structural biology of transcriptional regulation in papilloma viruses (PVs) has recently been described based on x-ray crystal structures of the human PV (HPV) E2 protein–DNA complex[1] and the corresponding uncomplexed DNA.[2] This system is of particular interest because the specificity of the E2 protein–DNA binding depends critically upon the sequence of a region of the DNA not in direct contact with the protein. This is indicative of a role for indirect readout in the molecular recognition process, and makes E2 an ideal system for detailed study of the hypothesis of a structural code for DNA recognition by regulatory proteins. However, the argument for indirect readout and a structural code for molecular recognition in this system depend on key features of the structures of the uncomplexed DNA in solution. Crystallography reveals three distinct conformations of the specific binding sequence d(ACCGAATTCGGT) in each asymmetric unit of the crystal. Given the observed polymorphism of the E2 DNA structure in the crystal and possible sensitivity of DNA structures to the crystalline environment,[3] the relevance of the crystal structures of the uncomplexed E2 DNA to the dynamical structure of the free oligonucleotide in solution is of interest. The E2 DNA in complexed form exhibits considerable deformation with respect to both canonical and crystallographic forms of uncomplexed E2 DNA, and the extent to which this is induced in the course of protein binding, or intrinsic to the solution structure, is of interest in understanding the nature of E2 protein DNA binding at the molecular level. Experimental determination of DNA structure in solution has been pursued particularly by NMR spectroscopy,[4,5] but NMR structures have not yet been reported for E2 DNA sequences. Recent studies of DNA structures in solution have been carried out theoretically using molecular dynamics (MD) simulations. Detailed comparisons between AMBER-based MD and experimentally determined DNA structures from both crystallography[6] and NMR spectroscopy[7] have shown close accord, with only minor discrepancies.[8,9] In order to investigate the E2 DNA and E2 protein DNA complexation, we present a set of MD studies of duplex d(ACCGAATTCGGT) in solution, using a set of simulation protocols essentially identical to that utilized in our recent MD and NMR comparisons of DNA solution structure.[7] The objective is to obtain an independent MD model of solution structure of the E2 DNA duplexes from different possible initial forms, including the canonical B form of the sequence, various crystal structures of the free DNA, and the protein bound form taken from the crystal structure of the complex. Based on these results, we examine in detail how the MD prediction of solution structure and corresponding crystallographic models of E2 DNA compare, and the extent to which the structural change of E2 DNA on complexation is intrinsic in the structure free in solution or induced by interactions with the E2 protein. The possible implications of the results with respect to indirect readout in E2 protein DNA recognition are considered.

BACKGROUND

The PVs are a family of small DNA viruses that cause the hyperplastic epithelial lesion[10,11] and have been linked to human conditions ranging for warts to cervical cancers.[12,13] The dominant transcriptional regulators of PV are the products of the E2 gene, which regulates transcription from all viral promoters and are essential to replication. The crystal structure of the binding domain of the bovine PV (BPV) E2 protein complexed as a dimer with a cognate DNA sequence has been reported by Hegde et al.[11] The E2 protein binds to a consensus DNA sequence ACCG NNNN CGGT (N=A, T, G, or C) that is found to occur through the viral genome. The binding affinity of the sequences for the E2 proteins is dependent upon the sequence of the central NNNN spacer, which in the crystal structure, as noted above, does not contact the DNA directly.

DNA sequences that bind with high affinity (NNNN=AATT) as well as low affinity (NNNN=ACGT) exhibit similar DNA conformations and protein–DNA contacts[1] in the complex. This suggests that the changes in E2 DNA binding affinity may be primarily dependent on the structural deformations that take place between free and protein-bound DNA targets. Since the difference involves regions not in direct contact with the E2 protein, this is presumably a case of "indirect readout" of a structure-based code.[1,11,14–16] Pursuing this line of thought further requires structures of high-affinity sequence in the free in solution and protein-bound form. Recently, Hizver et al.[2] reported crystallography at 2.2 Å resolution on the high-affinity HPV E2-DNA target d(ACCGAATTCGGT) (Figure 1a). Three independent DNA molecules per asymmetric unit of the P1 space group of the crystal structure were observed. Two of the DNA conformations (denoted X1 and X3) are very similar, while the other (X2) exhibits a distinct kink at the A_pT step. Hegde et al. reported the crystal structure at 2.4 Å resolution of the HPV E2 protein–DNA complex based on the DNA 16-mer sequence d(CAACCG AATTCGGTTG) (Figure 1b). In both

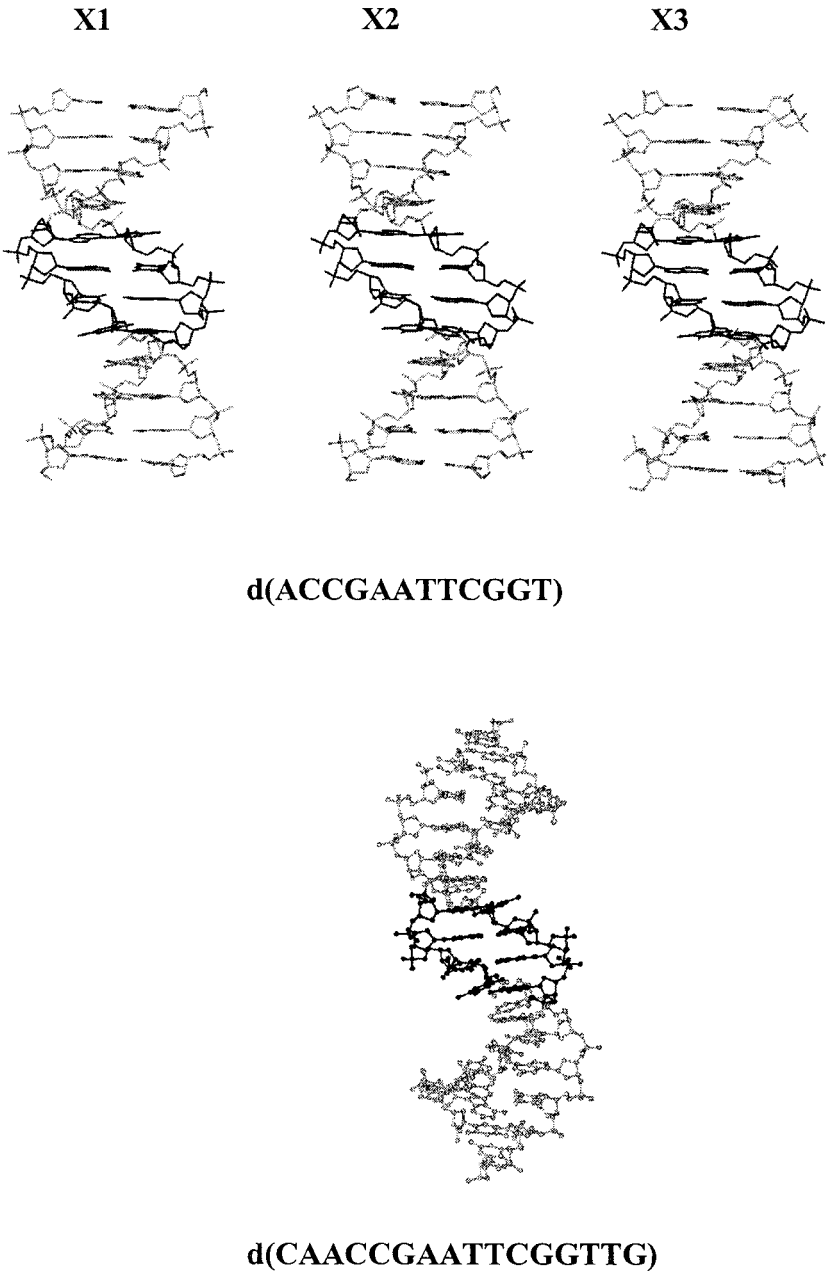


FIGURE 1 (a) Three DNA molecules (X1, X2, X3) from the x-ray structure for the native sequence d(ACCGAATTCGGT). Central AATT base pairs are shaded in black. (b) DNA cognate from the PDC crystal structure. Central AATT base pairs are shaded in black.

the free and bound crystal structures of E2 DNA, the DNA axis deformation is towards the minor groove with respect to the AATT region, which positions AATT on the concave side of the helix. In the case of the lower affinity ACGT sequence,[17] the uncomplexed DNA is bent toward the major groove of the structure in the ACGT region of the crystal, but towards the minor groove in the complex. The additional structural adaptation in the latter case was offered as a possible explanation of the differential binding affinity. However, the apparent flexibility of the DNA in the binding region of the molecules from crystal structure raises a question about the corresponding DNA structure in solution, and knowledge of the solution structure of E2 DNA is necessary to pursue this further.

The unique structures of sequences of adenines (A-tracts) are relevant in this system. DNA sequences

featuring multiple A-tracts situated in phase with a full turn of B-DNA double helix are noted for enhanced axis bending. Crothers et al.[18] determined that the bending is concave when viewed from opposite the minor groove of the A-tract regions. Whether this occurs by a deformation within the A-tracts (wedge model[19–21]), at or near the junction of A-tract and flanking sequence (junction model[22]), or in the non A-tract region, primarily at Y_PR steps (Goodsell Dickerson "nonA-tract" model[23–25]), has been discussed elsewhere.[26] Crystal structures of A-tract containing DNA sequences show that A-tracts are essentially straight, exhibit less dispersion across different instances,[26,27] and show no particular evidence for bending within A-tracts. Bolshoy et al.[19] showed that a wedge model of DNA could account for an extensive set of gel retardation results. However, measurements on diverse sequences in subsequent studies[28] show that result may not be unique, and that crystal structure data considering the uncertainty in stepwise structural deformations can also account for the data. An additional point of interest in the d(ACCGAATTCGGT) structures is the observation of a A_pT kink towards the minor groove of a duplex DNA. In previous cases of reported AATT structures, Hizver et al.[2] noted that the central decamer bends in this direction, but the overall sequence does not. They proceeded to utilize the crystal structure of this decamer to construct a model of a DNA 30-mer, which was found to bend toward the minor groove with a curvature of $\sim 10°$ per "A-tract." The bending in this model is partitioned evenly along the double helix and arises from negative *roll* angles in the AATT region, positive *roll* angles in the flanking regions, and buckle of the base pairs at junctions between A-tracts and flanking sequences. The magnitude of bending was noted to compare well with data obtained from solution studies on similar sequence motifs.[29]

MD modeling of DNA structure has recently been reviewed.[8,9,30] Several recent studies[31–33] have documented that AMBER MD[34] based on the Cornell et al (parm94) force field[35] with a particle mesh Ewald (PME) treatment of long-range interactions[36] produce stable MD trajectories for DNA on the nanosecond time frame, and provide a plausible description of the dynamical structure of B-form double helix. The basic protocol applied in this MD has been subsequently used to study the dynamical structure of a number of DNA and RNA oligonucleotides,[8] the dynamical structure of A-tracts and B′ form of DNA,[26,32] A_4T_4 and T_4A_4 motifs with periodic helix phasing,[37] A/B conformational preferences and transitions.[38,39] MD studies of a crystalline unit cell of four d(CGCGAATTCGCG)$_2$ under PME boundary conditions were carried out here[6] and compared directly with corresponding crystal structure data. The results show ~ 1 Å RMSD between the MD time averaged and crystal structure. Comparing the MD results on a B-DNA sequence in the crystal compared with those for the same sequence in solution provide a basis for a purely theoretical study of crystal packing effects.[6] Difference between the crystal MD and solution MD structures show packing effects at the 3′-end of the sequence, at which there is an interpenetration of helices and a G–G contact in the minor groove is prominent, but in this case the influence of crystal packing on structure seems to be local and not global.

As a more specific test of the ability of MD to predict solution structure for DNA, a detailed comparison between MD calculated and NMR observed indices of dynamical structure of DNA in solution has been carried out.[7] An MD trajectory of the structure of d(CGCGAATTCGCG) in solution based on the AMBER force field has been extended to 14 ns.[40] New measurements of the NMR spectrum for this sequence have been obtained in order to make the comparison between calculated and observed parameters correspond as closely as possible. Observable two-dimensional nuclear Overhauser effect spectroscopy (2D NOESY) intensities and scalar coupling constants were back calculated from the MD trajectory and compared with corresponding NMR measurements. The results indicate that the MD model is generally in good agreement with the NMR data, and shows closer accord with experiments than back calculations based on the d(CGCGAATTCGCG) crystal structure or canonical A and B forms of the sequence. The NMR parameters are not particularly sensitive to the deficiency in the AMBER MD model, a tendency toward undertwisting of the double helix.

The current state of the structural biology of the E2 DNA and protein DNA complexation indicates questions about the solution structure of the uncomplexed DNA as a basis for understanding the differential binding affinities, and the viability of the dodecamer crystal structures as an elementary unity in predicting and elucidating the nature of axis deformation in long DNA sequences. The specific questions we address are as follows: (a) To what extent do any of the various crystal structures of d(ACCGAATTCGGT) provide an accurate model for the corresponding DNA in solution? (b) Is the A_pT kinked structure an intrinsic feature of the DNA or a crystal packing effect? A model for a DNA 36-mer has been proposed based on one of the 12-mer structures. With the flanking sequence at or near the end of the 12-mer, this oligonucleotide may or may not be a good model of sequences flanking A-tracts in longer phased A-tract

sequences, and we investigate this assumption. To obtain an independent model of solution structure of the E2 DNA, we have performed a series of MD studies of duplex sequences with the E2 recognition motif in solution, and examine in detail how the MD and crystallographic models compare, and pursue the implications of the MD models of DNA solution structure and axis bending in E2 protein DNA recognition and binding.

METHODS

Four isothermal-isobaric (T, P, N) ensemble MD simulations were performed, two on the oligonucleotide duplex d(ACCGAATTCGGT) starting from X1 and X2 from the crystal structure (PDB 1ilc), respectively, and one on the canonical B-form structure. The fourth MD is on the DNA cognate sequence d(CAACCGAATTCGGTTG) taken from the crystal structure of the protein-DNA complex (PDB 1jj4). For each case, a net neutral charge on the system is maintained by using "minimal salt" ionic conditions, i.e., including 22 mobile Na^+ counterions for the dodecamer and 30 mobile Na^+ counterions for the DNA-cognate 16-mer. The calculations employed the all-atom AMBER parm94 force field developed by Cornell et al.[35] and the TIP3P model for water,[41] and were carried out using the AMBER 5.0 suite of programs. Details of the MD protocols are identical to those described elsewhere.[7,26] The DNA and ions were hydrated in a rectangular prism by TIP3P water molecules after placement of ions. The box dimensions were truncated such that a minimum distance of ~12 Å beyond all DNA atoms in all directions was achieved. The resulting box size is approximately 45 × 45 × 70 Å, solvating the dodecamer DNA with ~4000 water molecules, and 55 × 55 × 80 Å, solvating the DNA 16-mer with ~7000 water molecules. MD trajectories of 4 ns were obtained using the SGI Origin2000 array at NCSA. Conformational and helicoidal parameters were calculated using the program CURVES[42] with local parameters option, and the Molecular Dynamics Tool Chest MDTC.[43] Overall DNA bend angles were computed from global helicoidal axis obtained with Curves.

RESULTS

Root Mean Square (RMS) Displacements

The three MD simulations carried out on d(ACCGAATTCGGT) start from molecules 1 (X1) and 2 (X2) reported for the crystal structure of the native DNA[2] and from the canonical B-DNA form of the sequence, denoted MD X1, MD X2, and MD B, respectively. The fourth MD on the sequence d(CA ACCGAATTCGGT TG) taken from the HPV type 18 E2 protein–DNA complex[1] is denoted MD PDC (Protein DNA Complex). Snapshots of structures taken from the MD X1 trajectory superimposed onto the X1 starting structure are shown in Figure 2a. Here the crystal molecule X1 is represented as a white ball-and-stick model and the MD X1 snapshots are superimposed in framework representations. The RMS values of these snapshots in Figure 2a from X1 range from 2.2 to 3.7Å. The MD structure beginning with the X2 form (which features the A_pT kink) relaxed rapidly into essentially the same form as MD X1, indicating the A_pT kink is unstable structure in solution. The results of MD B are indistinguishable from that of MD X1 and MD X2. Thus all MDs on E2 DNA converge to essentially the same dynamical structure, henceforth referred to simply as MD.

RMS values comparing MD with the various crystal structures are reported in Table I. RMS values of each snapshot was computed every 2 ps over the 4-ns trajectory after the equilibration. The averaged RMS value is listed in Table I. The RMS difference between the molecules from the crystal range from 0.5 to 0.7 Å computed for heavy atoms only. The RMS differences between MD and X1, X2, and X3 are 2.9 ± 0.7, 2.7 ± 0.7, 3.0 ± 0.8 Å, respectively, and 3.3 ± 0.7 Å from B-DNA. The uncertainty arises due to the thermal motion in the MD trajectory and everywhere reported here as ± one standard deviation, in Å. The RMS differences between MD and X1, X2, and X3 computed for the central A_pT step for the heavy atoms are 0.7 ± 0.2, 0.8 ± 0.2, and 0.7 ± 0.2.

Snapshots of structures from MD PDC trajectory superimposed onto the E2 protein DNA crystal structure are shown in Figure 2b. The crystal structure conformation is represented as a white ball-and-stick model and the MD PDC snapshots are superimposed in framework representations. The bound form of the E2 DNA has a large bend angle of 40°, while the snapshots equilibrate to a less perturbed form. RMS values between the bound E2 DNA sequence and these snapshots from MD PDC trajectory range from 4.6 to 6.4 Å. The RMS difference between MD PDC the crystal protein–DNA cognate is 5.1 ± 0.9 Å. The RMS difference between MD PDC and B-DNA is 3.9 ± 0.5 Å. The RMS deviation of the averaged structure from MD PDC and MD for the central dodecamer is 0.9 Å (Figure 2c), i.e., not significantly different from the result obtained above for uncomplexed E2 DNA. Thus the MD predicted solution structure for the E2 recognition sequence is essentially the same independent of starting structure. The RMS difference is higher at the flanking ends, and smaller for the central region of the DNA. The RMS differences between the averaged structure beginning with the

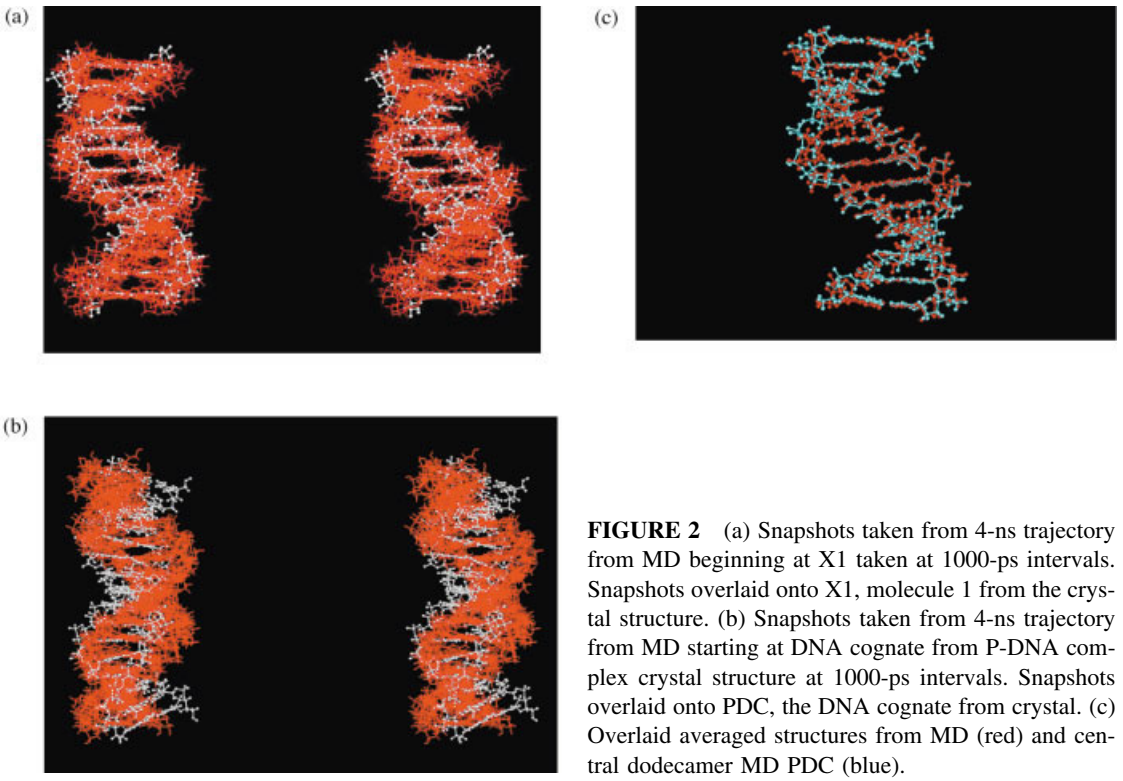


FIGURE 2 (a) Snapshots taken from 4-ns trajectory from MD beginning at X1 taken at 1000-ps intervals. Snapshots overlaid onto X1, molecule 1 from the crystal structure. (b) Snapshots taken from 4-ns trajectory from MD starting at DNA cognate from P-DNA complex crystal structure at 1000-ps intervals. Snapshots overlaid onto PDC, the DNA cognate from crystal. (c) Overlaid averaged structures from MD (red) and central dodecamer MD PDC (blue).

bound form and X1, X2, and X3 are 3.3, 3.2, and 3.5 Å, respectively.

Conformational and Helicoidal Analysis

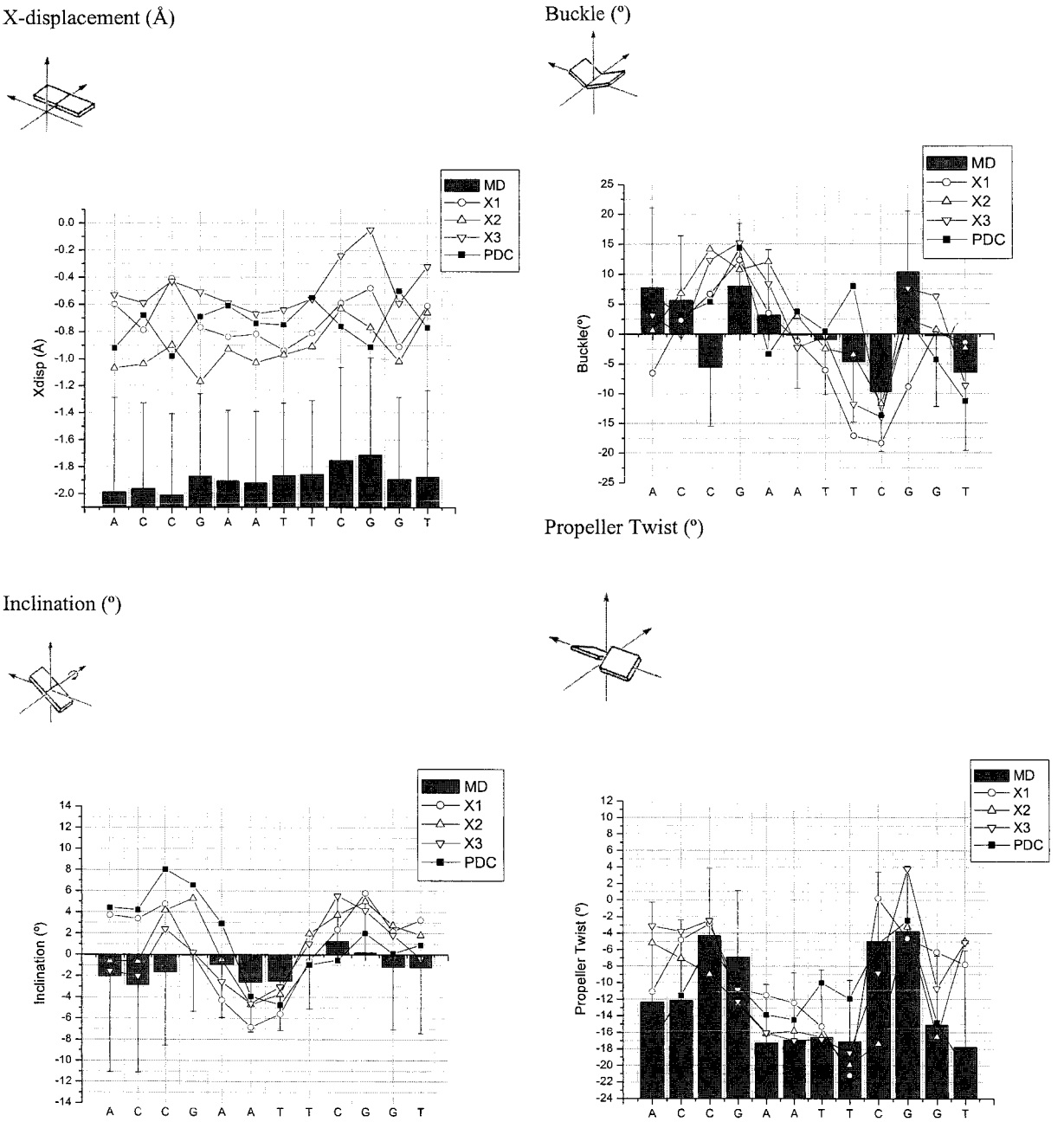
Analysis of the conformational dynamics of MD model of E2 DNA in solution showed that the sugar rings are mixed in the C2′–endo–O4′–endo–C3′–endo region. Notably, the sugar pucker associated with the 5′-end G of the C_pG step are significantly more localized to C2′-*endo* (B-form). There is typically increased but not systematic conformational fluctuations at or near the junctions of flanking sequences with the AATT region.

Analysis of the helicoidal parameters is given here in terms of the base-pair parameters *x-displacement*, *inclination*, *buckle,* and *propeller,* and the base-pair step parameters *roll*, *tilt*, *twist*, and *slide*, all defined with respect to the latest conventions. The values of *x-displacement* (Figure 3) for the MD structure are between −1.7 and −2.0 Å, and values of inclination (Figure 4) are $< 3°$. This places the MD structure well within the B family. MD solution structure shows systematically of 1 Å in *x-displacement* from crystal. The solution structure predicted by MD shows lower values of inclination and a less pronounced sequence dependence than is observed in the crystal conformation. There is good qualitative agreement between

Table I RMS Table for d(ACCGAATTCGGT) Dodecamer[a]

	MD	B-DNA	X1	X2	X3
MD	0	3.1	2.9 ± 0.7	2.7 ± 0.7	3.0 ± 0.8
B-DNA	3.3 ± 0.7	0	1.2	1.3	1.3
X1	2.9 ± 0.7	1.2	0	0.7	0.5
X2	2.7 ± 0.7	1.3	0.7	0	0.7
X3	3.0 ± 0.8	1.3	0.5	0.7	0

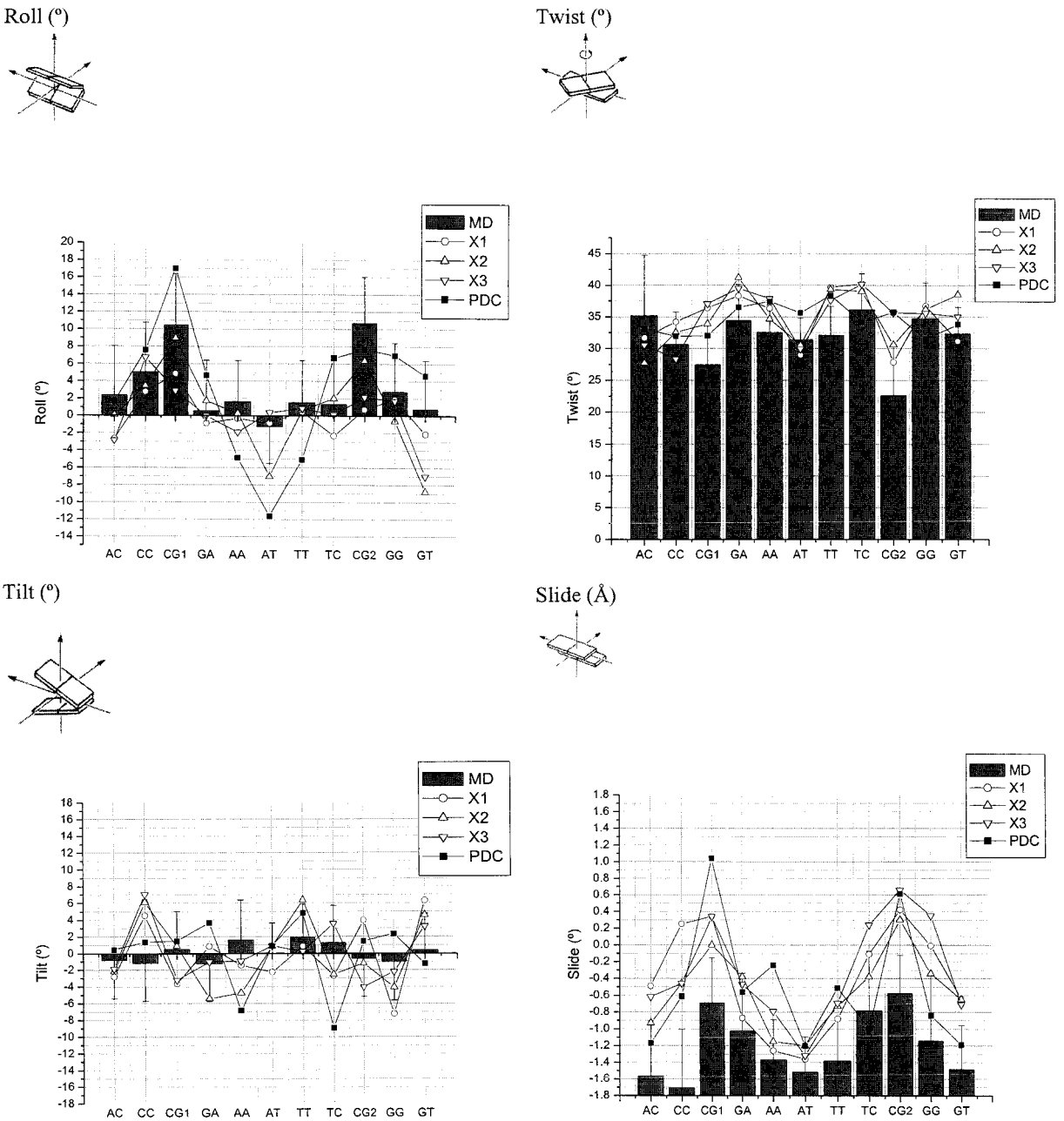
[a] RMS (Å) computed for MD solution structure and three molecules from x-ray (X1, X2, X3) of heavy atoms. RMS computed over 4-ns MD trajectory.



FIGURES 3–6 Base pair parameters *x-displacement* (Å), *inclination* (°), *buckle* (°), and *propeller* (°) computed for d(ACCGAATTCGGT). Values plotted in histogram are the ensemble average from MD. The vertical lines are one standard deviation of fluctuation from MD reported in Å and degrees, respectively. Values plotted in lines are the corresponding values for X1, X2, X3, and PDC.

MD and crystal values of *buckle* (Figure 5) shows between MD and x-ray at most steps in the sequence, especially considering the thermal range indicate for the MD values. There is a discrepancy at C3 for which the MD value of buckle is negative whereas crystal values are positive. MD values of *propeller* (Figure 6) show high propeller twisting in the AATT region.

The calculated base-pair step parameters for the MD and various molecules from crystal of E2 DNA are also shown in Figures 7–10. MD values of *roll* (Figure 7) are in good agreement with crystal, although here, as well in the base-pair parameters, smaller in magnitude. The discrepancy between the roll value of the MD form and large value of roll in



FIGURES 7–10 Base-pair step parameters *roll* (°), *tilt* (°), *twist* (°), and *slide* (°) computed for d(ACCGAATTCGGT). Values plotted in histogram are the ensemble average from MD, with vertical lines representing $\pm SD$ of fluctuation in (°). Values plotted in lines are the corresponding values for X1, X2, X3, and PDC.

crystal X2 is notable. However, the roll values of X2 are in better agreement with the MD at Y_PR steps, either than X1 or X3, indicating that the MD solution structure may best be described as a composite of the three molecules from crystal. A corresponding analysis of *tilt* values (Figure 8) shows little deviation from zero for MD. MD *tilt* values show most agreement with crystal molecules X1 and X3 in the AATT region. *Tilt* values of the crystal molecules in the flanking regions fall within the thermal fluctuation of the MD solution structure. Analysis of *slide* values (Figure 9, note the expanded scale on the ordinate) for MD in the AATT region is ~1 Å with respect to the x-ray structures, here again with the MD showing similar trends but reduced magnitudes. MD calculated *twist* angles (Figure 10) tend to run lower in MD than

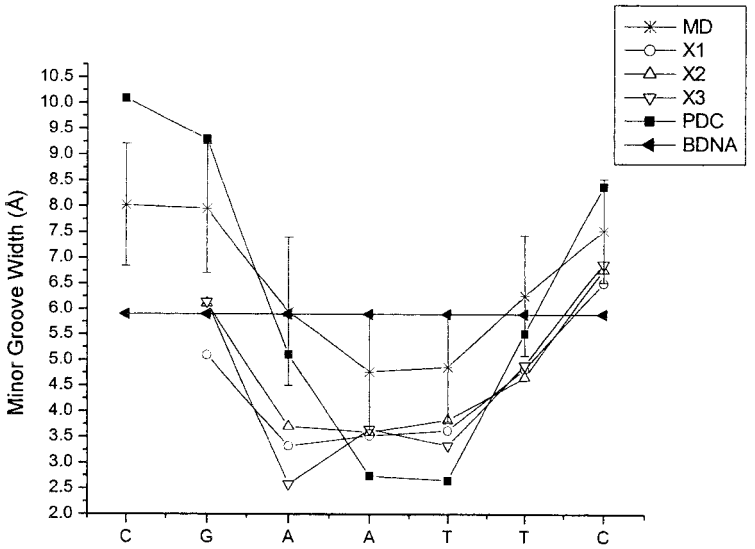


FIGURE 11 Minor groove widths for d(ACCGAATTCGGT) for MD, X1, X2, X3, PDC, and canonical B-DNA. For MD, ensemble average values are reported with one standard deviation in Å.

in the crystal molecules, discussed previously as one of the possible inaccuracies in the AMBER parm94 model of DNA in general.[9] Values obtained from MD[31,44] have been shown to be 2°–4° smaller than the value inferred from experimental data.[45,46] The discrepancy between MD and crystal values of *twist* for E2 DNA is most noticeable at the two C_PG steps in the sequence.

Analysis of the sequence dependence of the minor groove width for MD and the molecules from crystal structure of E2 DNA is shown in Figure 11. Both MD and crystal conformations show a narrowing of the minor groove as a function of sequence in the AATT region. The minor groove width of the MD structure is wider than canonical *B* value of 5.9 Å in the flanking sequences, and narrows to ~5 in the AATT region. Minor groove narrowing, accompanied by high propeller twist values and a small base-pair inclination are consistent with an incipient B´ structure[47] for the AATT region of d(ACCGAATTCGGT) for the MD solution structure. The minor groove width for X1, X2, and X3 fall within the range of 2.5–4.9 Å, with the average value being 3.8 Å. These values are smaller than MD solution models. Note that due to difference in convention, our calculated values for the crystal molecules from Curves are slightly smaller than those reported by Hizver et al. for these molecules (3.8–7.8 Å) based on P–P separations.[2]

The axis bending in the MD structure occurs as a consequence of pronounced *roll* values of ~+10° at the two C_PG steps, i.e., towards the major groove in this region. As a consequence of a convention introduced by Koo and Crothers,[22] it is customary to describe axis bending from a vantage point opposite to the minor groove of the central AATT tract. The helix axis in the MD structure bends towards the minor grove from this vantage point, but this happens as consequence of the two kinks at Y_PR steps in the flanking sequence, not because of bending within the AATT region. The bending angle for the DNA obtained from MD is 16° ± 8°, compared with values of 8°, 11°, and 8° for X1, X2, and X3, respectively (Figure 12); all values refer to a reference point opposite the minor groove of the AATT region.

The RMS values comparing crystal and MD obtained in this study fall within the range of results reported in the literature for MD studies of solution-phase DNA molecules of comparable length. McConnell and Beveridge report RMS values of sequences for the central 10-residue comparing MD with crystal structures of 2.5 Å.[26] Work presented on the 19-mer sequence for λ OL1 DNA operator showed RMS value of ~5 Å from starting canonical B-DNA form.[48] For a 30-mer containing phased A-tracts, the RMS value from the starting canonical form was reported to be ~7 Å.[32] Narrowing of the minor groove is observed for the MD solution structure, consistent with the work reported on phased A-tracts on *Bam* HI reported by Young and Beveridge,[32] studies of A_4T_4 sequences,[37] and seven MD solution studies of various DNA sequences containing A_3, A_4, A_5, and A_6 tracts by McConnell and Beveridge.[26]

DISCUSSION AND CONCLUSIONS

As described above, MD simulation predicts a solution structure for d(ACCGAATTCGGT) that is unique and independent of starting structure with structural features that are a composite of the three crystal molecules of the sequences. The A_pT kink in crystal for X2 disappears rapidly upon MD equilibration, indicating this to be an unstable deformation presumably induced by crystal packing. The overall axis bending is towards the minor groove with respect to the AATT-tract, but occurs as a consequence of kinks at Y_PR steps in the flanking sequences. The AATT element of the MD structure is essentially straight, showing values of roll between $-1°$ and $+1°$.

A subsequent MD on the 16-mer d(CAACCGAATTCGGTTG) beginning with the bent form observed in crystal structure of the complex was observed in the absence of protein to assume a solution form for the interior 12-mer essentially identical to that obtained in the preceding MDs. This indicates that the DNA structure in the complex is a protein induced structural adaptation of the solution structure. However, a comparison of the MD solution structure of E2 DNA with the canonical B form of the structure shows that the deformations in the intrinsic structure induced by sequence are in the direction of the protein-bound form. Interaction with protein induces further changes as indicated in the crystal structure of the complex. The collective results from this study and the crystal structures thus indicates that E2 protein DNA binding is a consequence of both intrinsic and induced structural changes with respect to the canonical B form of the sequence, and depend on both sequence-dependent deformation (intrinsic effects) and deformability (induced effects). Thus, if an indirect readout mechanism is operating, the MD results show it to be manifest via the flexibility of the sequence as well as the intrinsic structure.

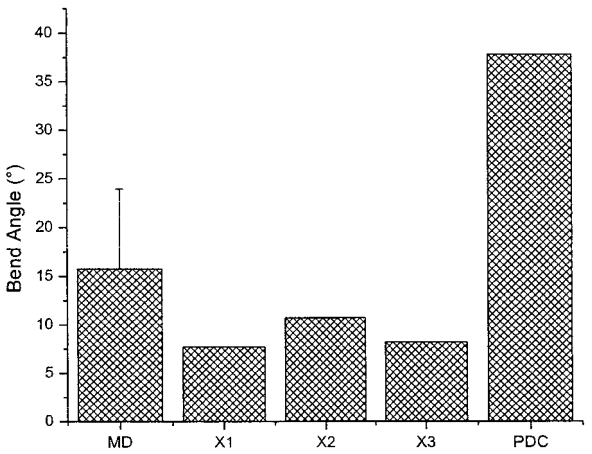


FIGURE 12 DNA bend angle computed for MD, X1, X2, X3, and PDC. Ensemble average value reported for MD with one standard deviation reported in degrees.

Table II RMS Table for (CA ACCGAATTCGGT TG)[a]

	MD PDC	B-DNA	PDC
MD PDC	0.0	3.9 ± 0.5	5.1 ± 0.9
B-DNA	3.9 ± 0.5	0.0	2.5
PDC	5.1 ± 0.9	2.5	0.0

RMS (Å) computed for MD PDC solution structure over the 4-ns trajectory, PDC, and B-DNA of heavy atoms. RMS computed for the central dodecamer.

The role of indirect readout in E2 protein DNA recognition is best viewed from a comparison of the crystal structures of d(ACCGAATTCGGT) and d(ACCGACGTCGGT). The latter sequence exhibits a distinctly altered structure in central four base pairs, coupled with the observation of reduced binding affinity.[17] Preliminary results from MD indicate that altered structure is also a feature of the solution structure of the ACGT sequence and are thus consistent with this interpretation. Further details of this calculation and MD simulations on E2 protein DNA complexes will be reported in a subsequent article.

The authors would like to thank Dr. Yongxing Liu and Dr. Surjit Dixit for helpful discussions. This research is supported by NIH grant GM-37909. Supercomputer facilities for these calculations were provided via the PACI program at the NCSA, University of Illinois.

REFERENCES

1. Kim, S. S.; Tam, J. K.; Wang, A.-F.; Hegde, R. S. J Biol Chem 2000, 275, 31245–31254.
2. Hizver, J.; Rozenberg, H.; Frolow, F.; Rabinovich, D.; Shakked, Z. Proc Natl Acad Sci USA 2001, 98, 8490–8495.
3. Berman, H. M. Biopolymers 1997, 44, 23–44.
4. Lane, A.; Jenkins, T. C.; Brown, T.; Neidle, S. Biochemistry 1991, 30, 1372–1385.
5. Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. J Am Chem Soc 2000, 122, 6190–6200.
6. McConnell, K. J. Ph.D. thesis, Wesleyan University, Middletown, CT, 2002.
7. Arthanari, H.; McConnell, K. J.; Beger, R.; A., Y. M.; Beveridge, D. L.; Bolton, P. H. Biopolymers 2003, 68, 3–15.

8. Cheatham, T. E., III; Young, M. A. Biopolymers 2001, 56, 232–256.
9. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182–196.
10. Sousa, R.; Dostatni, N.; Yaniv, M. Biochim Biophys Acta 1990, 1032, 19–37.
11. Hegde, R. S.; Grossman, S. R.; Laimins, L. A.; Sigler, P. B. Nature 1992, 359, 505–512.
12. zur Hausen, H. Biochim Biophys Acta 1996, 1288, F55–F78.
13. zur Hausen, H.; de Villiers, E. Annu Rev Microbiol 1994, 48, 427–447.
14. Hegde, R. S. Annu Rev Biophys Biomol Struct 2002, 343–360.
15. Hines, C. S.; Meghoo, C.; Shetty, S.; Biburger, M.; Brenowitz, M.; Hegde, R. S. J Mol Biol 1998, 276, 809–818.
16. Hegde, R. S. Adv Biophys Chem 1997,143–171.
17. Rozenberg, H.; Rabinovich, D.; Frolow, F.; Hegde, R. S.; Shakked, Z. Proc Natl Acad Sci USA 1998, 95, 15194–15199.
18. Zinkel, S. S.; Crothers, D. M. Nature 1987, 328, 178–181.
19. Bolshoy, A.; McNamara, P.; Harrington, R. E.; Trifonov, E. N. Proc Natl Acad Sci USA 1991, 88, 2312–2316.
20. DeSantis, P.; Pelleschi, A.; Savino, M.; Scipioni, A. Biochemistry 1990, 29, 9269–9273.
21. Ulanovsky, L. E.; Trifonov, E. N. Nature 1987, 326, 720–722.
22. Koo, H.-S.; Wu, H.-M.; Crothers, D. M. Nature 1986, 320, 501–506.
23. Calladine, C. R.; Drew, H. R.; McCall, M. J. J Mol Biol 1988, 201, 127–137.
24. Goodsell, D. S.; Kaczor Grzeskowiak, M.; Dickerson, R. E. J Mol Biol 1994, 239, 79–96.
25. Maroun, R. C.; Olson, W. K. Biopolymers 1988, 27, 585–603.
26. McConnell, K. J.; Beveridge, D. L. J Mol Biol 2001, 314, 23–40.
27. Young, M. A.; Ravishanker, G.; Beveridge, D. L.; Berman, H. M. Biophys J 1995, 68, 2454–2468.
28. Liu, Y.; Beveridge, D. L. J Biomol Struct Dyn 2001, 18, 505–526.
29. Shatzky-Schwatz, M.; Arbuckle, N.; Eisenstein, M.; Rabinovich, D.; Bareket-Samish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z. J Mol Biol 1997, 267, 595–623.
30. MacKerell, J., A. D.; Banavali, N. J Comp Chem 2000, 21, 105–120.
31. Cheatham, T. E., III; Kollman, P. A. J Am Chem Soc 1997, 119, 4805–4825.
32. Young, M. A.; Beveridge, D. L. J Mol Biol 1998, 281, 675–687.
33. Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. J Am Chem Soc 1999, 121, 5981–5991.
34. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. AMBER: Version 5, University of California, San Francisco, 1997.
35. 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. J Am Chem Soc 1995, 117, 5179–5197.
36. Darden, T. A.; York, D. M.; Pedersen, L. G. J Chem Phys 1993, 98, 10089–10092.
37. Sprous, D.; Young, M. A.; Beveridge, D. L. J Mol Biol 1999, 285, 1623–1632.
38. Cheatham, T. E., III; Kollman, P. A. J Mol Biol 1996, 259, 434–444.
39. Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. J Am Chem Soc 1998, 120, 10629–10633.
40. Young, M. A.; Jayaram, B.; Beveridge, D. L. J Phys Chem B 1998, 102, 7666–7669.
41. Jorgensen, W. L. J Am Chem Soc 1981, 103, 335–340.
42. Lavery, R.; Sklenar, H. Curves 5.1: Helical Analysis of Irregular Nucleic Acids; Institut de Biologie Physico-Chimique: Paris, France, 1996.
43. Ravishanker, G.; Wang, W.; Beveridge, D. L. Molecular Dynamics Tool Chest 2.0: Wesleyan University: Middletown, CT, 1998.
44. Young, M. A.; Ravishanker, G.; Beveridge, D. L. Biophys J 1997, 73, 2313–2336.
45. Kabsch, W.; Sander, C.; Trifonov, E. N. Nucleic Acids Res 1982, 10, 1097–1104.
46. Suzuki, M.; Amano, N.; Kakinuma, J.; Tateno, M. J Mol Biol 1997, 274, 421–435.
47. Saenger, W. Principles of Nucleic Acid Structure; Springer-Verlag: New York 1984.
48. Kombo, D. C.; McConnell, K. J.; Young, M. A.; Beveridge, D. L. Biopolymers 2001, 59, 205–225.

Reviewing Editor: Dr. David A. Case