

Molecular Dynamics Studies of DNA A-Tract Structure and Flexibility

Edward C. Sherer,[†] Sarah A. Harris,[†] Robert Soliva,[‡] Modesto Orozco,[‡] and Charles A. Laughton^{*,†}

Contribution from the Cancer Research Laboratories, School of Pharmaceutical Sciences, University of Nottingham, NG7 2RD, UK, and Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain

Received October 21, 1998

Abstract: Molecular dynamics simulations have been used to compare the structure and dynamics of three A-tract-containing DNA dodecamer sequences: d(CGCAAATTTGCG)₂, d(CGCAIATMTGCG)₂, and d(CGCI-IMMMGCG)₂, where M = 5-methylcytosine. The simulations shed light on experimental observations regarding DNA bending induced by these sequences. We find that replacing an A•T base pair by an I•M base pair does far more to the structure and particularly dynamics of the oligonucleotides than might be expected if the substitution were regarded as just exchanging a hydrogen bond donor and acceptor across the DNA major groove. The evaluation of the molecular dynamics data is greatly simplified by the application of the method of principal component analysis. This allows key differences in the structures and dynamics of the three systems to be readily discerned. Three major modes of deformation are observed, the amplitudes and/or average values of which can vary with sequence. The results allow a simple interpretation of the effects of A•T to I•M substitutions on DNA bending and point to the importance of DNA flexibility, as much as static structure, in determining macroscopic behavior.

Introduction

A-tracts, which may be defined as A_mT_n sequences in DNA where $m + n > 5$, have been shown to be capable of introducing a bend into duplex DNA, such that if a number of such tracts are present at intervals corresponding to the helical repeat, macroscopic curvature of the DNA can be observed (for a review, see Olson and Zhurkin¹). The most common way of examining this curvature has been gel electrophoresis,² but other, more direct methods have included electron microscopy³ and scanning force microscopy.⁴ Despite many studies at the molecular level, the reasons for this induced curvature remain unclear. Several models for A-tract-induced bending have been proposed; in essence, opinion is divided as to whether the bending takes place within the A-tract itself (wedge models, “A-tract bending”⁵), outside the A-tract in general sequence DNA (wedge models, “non-A-tract bending”⁶), or at the junctions between the A-tract and the preceding and following DNA sequences (junction models⁷). Structure determination of A-tract-containing oligonucleotides has not helped to completely

resolve this controversy (see, for example, recent commentaries by Dickerson et al.^{6a}), but has provided much interesting data. One of the key observations has been that A-tract sequences are characterized by high propeller twist, such that bifurcated H-bond interactions between neighboring bases can occur across the major groove (Figure 1a).^{8,9} It has been hypothesized that such interactions would tend to rigidify the helix in the region of the A-tract and also emphasize conformational transitions at the A-tract junctions. To test some of these ideas, the effect of replacing A•T base pairs in A-tracts with I•C or I•M base pairs has been investigated (Figure 2). The rationale behind this has been that the substitution should be equivalent to doing little other than exchanging a carbonyl and amino group across the major groove but that this should effectively inhibit the formation of bifurcated hydrogen bonds with neighboring A•T base pairs (Figure 1b). However, gel mobility studies have shown that such substitutions have a very minor effect on the curvature induced by the A-tracts.¹⁰

The results of recent crystallographic studies have indicated why this might be.¹¹ Somewhat surprisingly, it was found that the dodecamer d(CGCAIATMTGCG)₂ (M = 5-methylcytosine) showed a high propeller twist conformation very similar to that

[†] University of Nottingham.

[‡] Universitat de Barcelona.

(1) Olson, W. K.; Zhurkin, V. B. In *Biological Structure and Dynamics, Proceedings of the Ninth Conversation, State University of New York, Albany*; Sarma, R. H., Sarma, M. H., Eds.; Adenine: Schenectady, NY, 1996; pp 24–44.

(2) Diekmann, S. *Methods Enzymol.* **1992**, *212*, 30–46. Dlakic, M.; Harrington, R. E. *Nucleic Acids Res.* **1998**, *26*, 4274–4279.

(3) Griffith, J.; Bleyman, M.; Rauch, C. A.; Kitchin, P. A.; Englund, P. T. *Cell* **1986**, *46*, 717–724.

(4) Rivetti, C.; Walker, C.; Bustamante, C. *J. Mol. Biol.* **1998**, *280*, 41–59.

(5) Bolshoy, A.; McNamara, P.; Harrington, R. E.; Trifonov, E. N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2312–2316.

(6) (a) Dickerson, R. E.; Goodsell, D.; Kopka, M. L. *J. Mol. Biol.* **1996**, *256*, 108–125. (b) Olson, W. K.; Marky, N. L.; Jernigan, R. L.; Zhurkin, V. B. *J. Mol. Biol.* **1993**, *232*, 530–554.

(7) Crothers, D. M.; Haran, T. E.; Nadeau, J. G. *J. Biol. Chem.* **1990**, *265*, 7093–7096.

(8) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S. *J. Mol. Biol.* **1992**, *226*, 1161–1173.

(9) (a) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, *330*, 221–226. (b) DiGabrielle, A. D.; Steitz, T. A. *J. Mol. Biol.* **1993**, *321*, 1024–1039. (c) Aymami, J.; Coll, M.; Frederick, C. A.; Wang, H.-J.; Rich, A. *Nucleic Acids Res.* **1989**, *17*, 3229–3245.

(10) Diekmann, S.; Mazzarelli, J. M.; McLaughlin, L. W.; von Kitzing, E.; Travers, A. A. *J. Mol. Biol.* **1992**, *225*, 729–738.

(11) Shatzky-Schwartz, M.; Arbuckle, N. D.; Eisenstein, M.; Rabinovich, D.; Bareket-Samish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z. *J. Mol. Biol.* **1997**, *267*, 595–623.

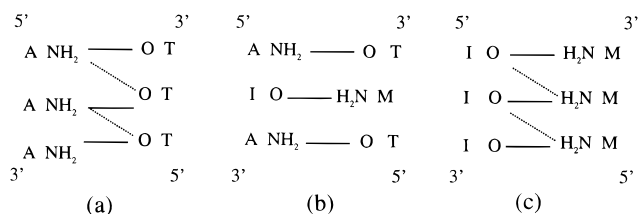
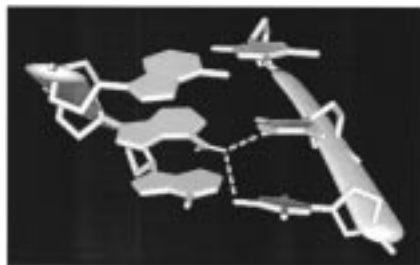


Figure 1. Relationship between sequence, propeller twist, and bifurcated cross-strand H bonding. The top panel illustrates the highly propeller-twisted conformation of the AT base pairs in the AAA crystal structure, and below (a) how this may be stabilized by nonstandard H-bonds between neighboring A and T bases (dashed lines). It might be expected that these interactions would not be possible in the AIA sequence (b), but would be possible in the III sequence (c).

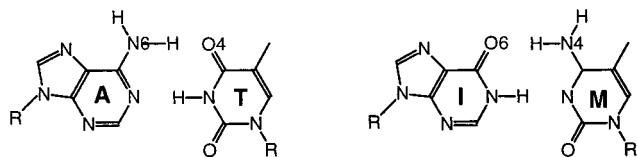


Figure 2. Structures of A•T and I•M base pairs, showing standard numbering of key atoms.

of the parent oligonucleotide $d(\text{CGCAAATTTGCG})_2$. It was proposed that this might be the result of the formation of unconventional amino–amino H-bond interactions across the major groove; theoretical and database analysis studies have provided some support for this idea.¹² In contrast, the structure of the decamer duplex $d(\text{CCIIICCCGG})_2$ was found not to show high propeller twist, although in theory it could form conventional cross-strand H-bonds (Figure 1c). This observation again supported gel mobility studies, which showed that sequences containing phased repeats of IIIICC (or IIIMMM) showed little curvature. As with previous structural studies,^{8,9} these results do not of themselves reveal the structural origins of the DNA bending but suggest possible explanations. It was concluded that hydration patterns may play an important role in stabilizing the highly propeller-twisted structures and that sequence-dependent variations in DNA flexibility might be as important as variations in DNA structure for producing macroscopic curvature.

Here we describe molecular dynamics simulations (MD) on three A-tract containing DNA dodecamers and their analysis in terms of the above hypotheses. The sequences studied were $d(\text{CGCAAATTTGCG})_2$ (“AAA”), $d(\text{CGCAIATMTGCG})_2$ (“AIA”), and $d(\text{CGCIIIMMMGCG})_2$ (“III”). We observe that all three simulations lead to time-averaged structures that are compatible with the crystallographic data. Analysis of hydration patterns does not reveal any clear-cut distinctions between the different sequences, but clear trends are evident in flexibility.

(12) Luisi, B. F.; Orozco, M.; Sponer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123–1136.

This analysis is greatly simplified by the application of the methods of principal component analysis. Three major modes of deformation are observed, the amplitudes and/or average values of which can vary with sequence. The results allow a simple interpretation of the effects of A•T to I•M substitutions on DNA bending and point to the importance of DNA flexibility, as much as static structure, in determining macroscopic behavior.

Methods

General. All simulations were performed using the AMBER 4.1 suite of programs,¹³ running on Silicon Graphics workstations. Partial charges for molecular fragments not present in the standard AMBER force field were obtained from HF/6-31G* calculations and the RESP procedure. Molecular visualization and manipulation were done using the programs MidasPlus,¹⁴ VMD,¹⁵ and Insight II from Biosym/MSI. Helix analysis was performed using Curves 4.1.¹⁶ Where this yields both global and local versions of parameters, the local parameters are quoted in all cases.

Construction of the Starting Structures. Crystal structure coordinates were obtained from the Nucleic Acid Database¹⁷ for $d(\text{CGCAAATTTGCG})_2$ ⁸ (“AAA”, NDB code BDL038, PDB code 1D65) and $d(\text{CGCAIATMTGCG})_2$ ¹¹ (“AIA”, NDB code BDLB76, PDB code 285D). The latter was then manipulated using Insight II to give an initial structure for $d(\text{CGCIIIMMMGCG})_2$ (“III”). Twenty-two sodium counterions were added along phosphate bisectors to each of the three structures which were then immersed in a periodic box of TIP3P water molecules to give a minimum 8-Å clearance between any solute atom and a box edge (approximately 90 002 800 water molecules in each case).

Molecular Dynamics Methods. Prior to unrestrained MD, each starting system was conditioned for a total of 90 ps as described previously.¹⁸ All MD simulations were performed at constant pressure and temperature (300 K) by means of periodic boundary conditions and the particle mesh Ewald (PME) method.^{19,20} SHAKE was used to constrain all bonds, allowing a 2-fs time step. Coordinates were saved every 0.5 ps of unrestrained dynamics for subsequent analysis. The production phase of each simulation was 1 ns.

Static Analysis Methods. Root-mean-square (rms) deviations of snapshots from initial structures were calculated over all DNA atoms, as were the autocorrelation functions. Each snapshot was also analyzed using Curves; where appropriate, helical parameters for III were compared with those from the crystal structure of $d(\text{CCIIICCCGG})_2$ ¹¹ (PDB code 286D). Time-averaged structures were computed from the last 800 ps of each trajectory and then minimized within AMBER to fix poor geometry from the averaging process as previously described.¹⁸ Bending dials were produced according to the method described by Young et al.²¹ Local helical parameters roll (ρ) and tilt (τ) were used to calculate the angle of axis deflection (θ) at each base step and its directionality (ϕ) measured from the direction of the major groove.

(13) Perlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1.

(14) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13–27.

(15) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(16) Lavery, R.; Sklenar, J. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.

(17) Abola, E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. Protein Data Bank. In *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107–132.

(18) (a) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463–7469. (b) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895–5904.

(19) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *99*, 8345–8348.

(20) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(21) Young, M. A.; Ravishanker, G.; Beveridge, D. L.; Berman, H. M. *Biophys. J.* **1995**, *68*, 2454–2468.

$$\theta = (\rho^2 + \tau^2)^{1/2} \quad (1)$$

$$\phi = \tan^{-1}(\tau/\rho) \quad (\text{for } \rho > 0) \quad (2a)$$

$$\phi = 180 + \tan^{-1}(\tau/\rho) \quad (\text{for } \rho < 0) \quad (2b)$$

Thus, ϕ values around zero correspond to bending into the major groove, while values around 180 degrees correspond to bending into the minor groove. Lu and Olson²² have recently shown that tilt is a parameter very sensitive to the choice of calculation method or, more precisely, the choice of reference frame. The reference frame used in Curves is not strictly consistent with the construction leading to θ and ϕ , but is used here for consistency with previous work. Water density calculations were performed as described previously,¹⁸ by using the time-averaged structures for reference.

PCA Methods. Principal component analysis was performed essentially as described by Wlodek et al.²³ From the equilibrated portion of each trajectory, or from the concatenated trajectories, the positional covariance matrix \mathbf{C} was calculated. For M snapshots of an N atom system, \mathbf{C} is a $3N \times 3N$ matrix whose elements are given by

$$c_{ij} = M^{-1} \sum_{k=1}^M \{x_i(k) - \langle x_i \rangle\} \{x_j(k) - \langle x_j \rangle\} \quad (3)$$

where x_i are the atomic coordinates. Diagonalization of \mathbf{C} to solve

$$\Lambda = \mathbf{V}^T \mathbf{C} \mathbf{V} \quad (4)$$

provides a set of $3N$ orthogonal eigenvectors, \mathbf{v}_n , as columns of matrix \mathbf{V} , in conjunction with their corresponding eigenvalues, λ_n , the diagonal elements of Λ . The eigenvectors provide a vectorial representation of each mode of structural deformation, and the eigenvalue for a mode indicates the relative contribution that this mode has made to motion within the trajectory; this may be expressed as a fraction of the total motion, $\lambda_i / \sum_i \lambda_i$. Projections of the trajectory ($\mathbf{r}(t)$) on the major eigenvectors (eq 5) were analyzed for their time-dependent behavior and probability distributions.

$$p_n(t) = \mathbf{v}_n \mathbf{r}(t) \quad (5)$$

The matrix \mathbf{V} may be regarded as a rotation matrix. We produced filtered versions $\mathbf{r}^m(t)$ of the original trajectory $\mathbf{r}(t)$ by rotating the coordinates into the new frame, producing the transformed coordinates $\mathbf{p}(t)$, (eq 6); replacing all elements of $\mathbf{p}(t)$, save that of the eigenvector of interest, m , with the corresponding time-averaged value, thus generating $\mathbf{p}^m(t)$ (eq 7), and then reversing the rotation (eq 8)

$$\mathbf{p}(t) = \mathbf{V} \mathbf{r}(t) \quad (6)$$

$$\mathbf{p}^m(t) = [\langle p_1 \rangle, \langle p_2 \rangle, \dots, \langle p_{m-1} \rangle, p_m(t), \langle p_{m+1} \rangle, \dots, \langle p_{3N} \rangle]^T \quad (7)$$

$$\mathbf{r}^m(t) = \mathbf{V}^{-1} \mathbf{p}^m(t) \quad (8)$$

To ease the interpretation of the deformations associated with each principal component, short MD trajectories were generated "artificially" by generating structures in which $p_m(t)$ (eq 7) was varied linearly between the minimum and maximum values observed. The resulting trajectories were inspected visually within VMD and analyzed by using Curves to identify helical parameters, or simple combinations of helical parameters, which captured, albeit imperfectly, the major motion involved.

Molecular Interaction Potential (MIP) Calculations. The potential associated with the interaction of an O^+ probe atom with the time-averaged duplex structures (the molecular interaction potential) was calculated using the method described previously.^{18a} Briefly, the interaction energy is the sum of an electrostatic contribution, calculated using the nonlinear Poisson–Boltzmann method, and a Lennard–Jones

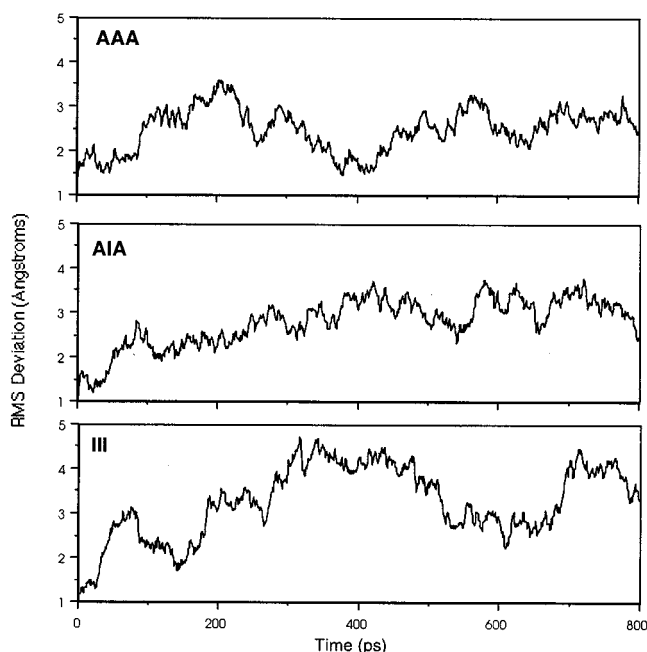


Figure 3. Root-mean-square deviation plots for the three trajectories, referenced to the corresponding starting structures.

contribution, calculated using standard OPLS parameters for the DNA and the TIP3P parameters for the probe. Regions of negative potential define areas where interaction with a positively charged ion or ligand would be favorable; regions of high potential, whether negative or positive, define areas where interaction with a polar ligand, e.g., a water molecule, would be favorable.

Calculation of Stacking Interaction Energies. To examine the effect of helical twist on the stacking energy of AAA and III sequences, Insight II was used to create 14 structures of both d(AAA)·d(TTT) and d(III)·d(MMM) in which the helical twist was varied between 27° and 40°. All sugar–phosphate backbone atoms were removed from each structure, and C1' was replaced by a hydrogen (H1). In the case of inosine, we are therefore effectively modeling the imino tautomer of hypoxanthine. Standard AMBER 5.0 partial charges were used on all base atoms, with the exception of H1 whose charge was adjusted to render the overall system neutral, and standard AMBER 5.0 van der Waals parameters were used for all atoms. The total stacking energy was computed as the sum of the intrastrand and interstrand contributions.

Results

General Features of the MD Trajectories. Animation of the trajectories revealed that in all cases the duplex structure of the DNA was maintained throughout. Figure 3 shows the rms deviations of snapshots from the trajectories from their respective starting structures. Although the rms deviations fluctuate widely, in all cases after 200–300 ps there is no evidence for long-term drift to higher values, suggesting the simulations are equilibrated. The time-correlation plots (Figure 4) indicate that the relaxation time for the DNA is of the order of 200 ps, providing further support for the view that these simulations are long enough to capture the essential flexibility of each DNA sequence. The rms deviation plots indicate that all simulations equilibrate to an envelope of conformations that differ to some extent from the starting structures. Measured over the last 800 ps of each simulation, the average rms from the starting structures are 2.57 Å for the AAA simulation, 2.88 Å for the AIA simulation, and 3.32 Å for the III simulation. As expected, the value for the III simulation is greatest; this simulation was begun from the "mutated" AIA crystal structure, rather than from a genuine crystal structure for this sequence.

(22) Lu, X.-J.; Olson, W. K. *J. Mol. Biol.* **1999**, *285*, 1563–1575.

(23) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513–9522.

Table 1. Selected Helical Parameters from the Three MD Simulations^a

base/step	AAA			AIA			III		
	roll (deg)	twist (deg)	PrTw (deg)	roll (deg)	twist (deg)	PrTw (deg)	roll (deg)	twist (deg)	PrTw (deg)
1 (C)			-13.2			-17.7			-21.7
2 (G)	10.7	34.6		11.9	34.5		10.9	35.9	
3 (C)	-6.1	33.7	-17.8	-3.2	33.3	-13.0	-10.1	35.4	-20.4
4 (A/I)	18.9	26.4	-2.5	14.7	29.3	-6.4	17.7	26.1	-1.1
5 (A/I)	1.0	32.3	-14.1	0.3	30.1	-12.8	0.5	29.8	-10.8
6 (A/I)	-0.5	33.4	-17.8	-1.3	34.2	-17.4	0.4	30.5	-13.4
7 (T/M)	-2.2	30.8	-18.5	-2.4	30.6	-14.7	-2.7	33.5	-9.4
8 (T/M)	0.0	36.2	-20.6	-0.7	34.4	-15.2	1.2	30.9	-11.2
9 (T/M)	-7.7	36.2	-25.3	1.5	29.8	-17.0	1.2	31.2	-15.0
10 (G)	21.7	22.3	-2.5	19.7	24.1	-9.7	15.7	27.2	-9.6
11 (C)	-8.1	33.9	0.1	-11.7	33.6	1.9	-1.9	30.9	-5.1
12 (C)	13.1	33.1	-14.8	20.7	27.6	-16.3	12.1	34.0	-13.7
12 (G)			-18.8			-8.1			-18.6
ave ^b	3.7	32.0	-13.8	4.5	31.0	-12.2	4.1	31.4	-12.5
ave (A) ^c	-1.8	33.8	-16.5	-0.5	31.8	-14.5	0.1	31.2	-11.6
crystal ^d	0.85	36.6	-22.6	-0.93	36.6	-19.2	-0.2	35.8	-12.6

^a Helical parameters calculated by means of Curves. Values are averages over the last 700 ps of each trajectory. ^b Average over all bases/base steps. ^c Average over A-tract only. ^d Average over A-tract in crystal structures (1D65,⁸ 285D,¹¹ and 286D¹¹).

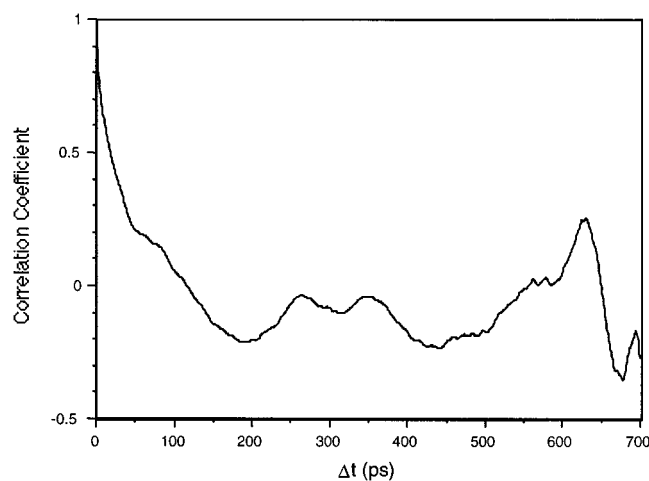


Figure 4. Time-correlation plot for the AAA simulation, calculated using data from the last 800 ps of the simulation. The AIA and III simulations produced similar plots.

Analysis of helical parameters indicates the nature of some of these conformational differences (Table 1). The structures from the simulations have, on average, 3–5° less helical twist than the crystal structures. They also show reduced propeller twist, 5° less in the case of AAA and AIA, just 1° less in the case of III. The effect of these differences on the possibility of forming nonstandard cross-strand H-bonding interactions is shown in Figure 5, in which the distributions in selected heavy atom–heavy atom distances through the last 200 ps of each simulation are presented. These show clearly that separations of 3.5 Å or less are increasingly less common as the sequence changes from AAA to AIA to III. To analyze bending, we used the bending dials approach of Young et al.²¹ The results are shown in Figure 6, from which it is clear that all three sequences show marked bending into the major groove at the A-tract junctions but that the degree of bending is very similar for all

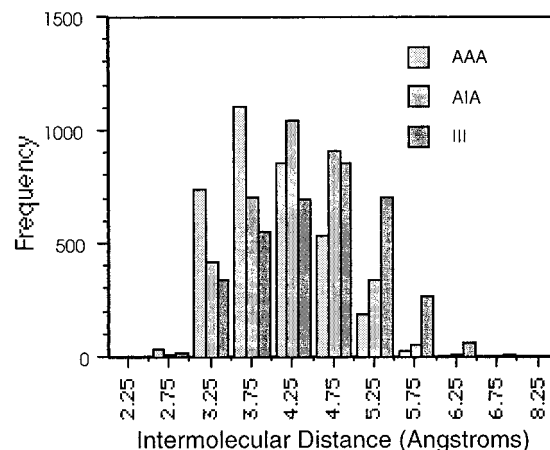


Figure 5. Distributions of selected atom pair separations associated with bifurcated H-bond formation. For AAA these are A–N6 to T–O4 distances (dashed lines in Figure 1a), for AIA these are A–N6 to M–N4 and I–O6 to T–O4 distances (equivalent diagonal relationships in Figure 1b), and for III I–O6 to M–N4 distances (dashed lines in Figure 1c). Data were collected from the last 800 ps of each simulation.

three sequences. Measuring the average roll at the two A-tract junction steps in each sequence also shows this. For AAA the average roll at these steps is 20.3°, for AIA it is 17.2°, and for III it is 16.7°. These figures may be contrasted with the average values within the A-tracts shown in Table 1 and with the crystal structure data, which in general do not show an unusual roll at the junction steps. The junction steps are also characterized by reduced helical twist and reduced propeller twist; the former is also evident in the crystal structures, but not the latter.

Analysis of Hydration Patterns. It has been hypothesized that differences in hydration patterns might play a role in the sequence dependence of A-tract bending. Shatzky-Schwartz et al.¹¹ concluded that highly propeller-twisted tracts appeared to

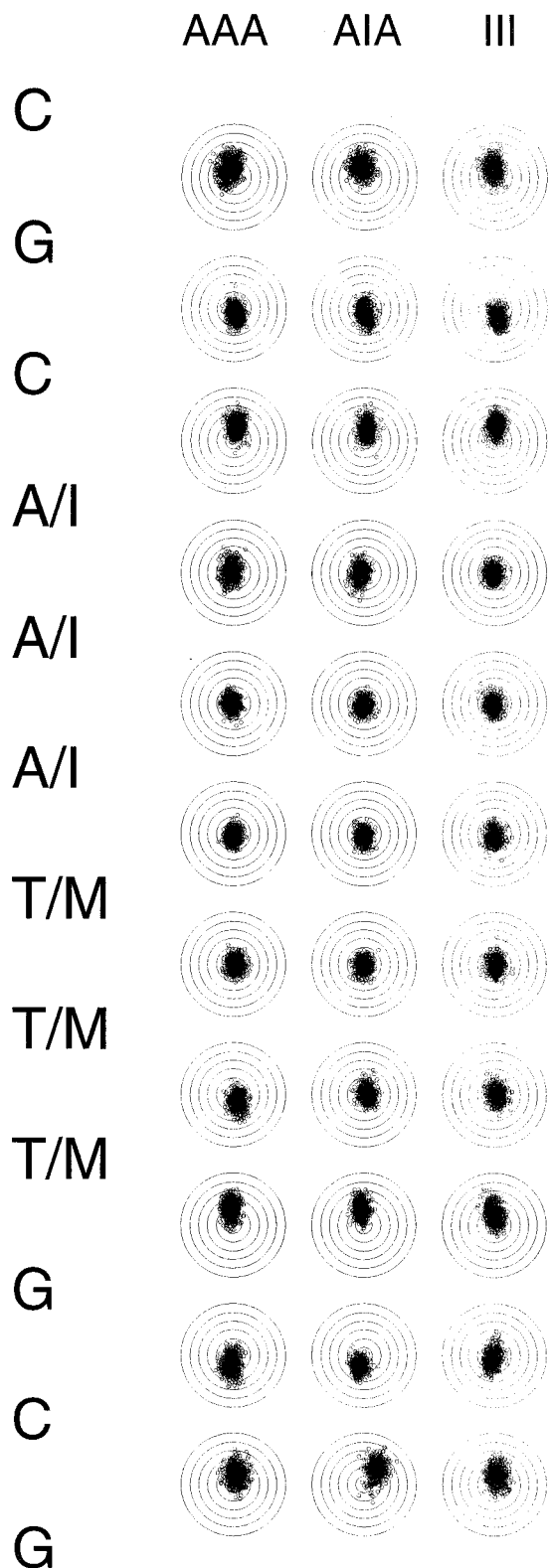


Figure 6. Bending dials analysis²¹ of the trajectories, calculated using local helix parameters. Bending through compression of the major groove is evident at the A-tract junction steps in all three simulations. The bending rings are drawn at 5° increments.

favor hydration patterns in the major groove where the water molecules lay between the base pair planes, whereas in less propeller-twisted I-tracts water molecules tended to lie within the base pair planes. We analyzed the hydration patterns within both the major and minor grooves of each sequence according to the methods previously described.¹⁸ The results are shown

Table 2. Principal Component Analysis of the Three MD Trajectories

trajectory	PC component	proportion of motion (%) ^a	interpretation
AAA	1	30	junction bending
	2	19	
	3	11	
AIA	1	33	helical twisting
	2	11	
	3	8	
III	1	39	helical twisting
	2	11	
	3	9	
combined	1	26	helical twisting
	2	21	
	3	15	

^a See methods section for definition.

in Figure 7a, where a clear spine of hydration in the minor groove is visible in all three structures. However, clear sequence-dependent hydration patterns in the A-tract region of the major groove, where A•T to I•M substitutions might be expected to have a major effect, are difficult to detect. Difference density maps reveal as many differences in hydration patterns outside the A-tract region, where the sequences are the same, as within this region, suggesting that any analysis may not be statistically significant.

Molecular Interaction Potentials. In contrast to the inconclusive results from the water density calculations, molecular interaction potential (MIP) analysis revealed a sequence-dependent feature. As expected, all structures showed a region within the minor groove where interaction with a positively charged ion, or charged or polar ligand, would be particularly favorable. However, the time-averaged III structure showed another such region within the A-tract major groove, whereas the AAA or AIA structures did not (Figure 7b). However, this feature does not appear to be localized within the base pair planes, as the crystallographic results suggest.

Principal Component Analysis. Analysis of time-averaged values of structural parameters did not reveal any clear-cut difference between the III A-tract and the other two sequences, in contrast to the experimental data on differences in induced bending. Therefore, we analyzed the flexibilities of the three sequences by using the method of principal component analysis (PCA). Initially, PCA was applied to each trajectory independently. Table 2 lists the top three eigenvalues obtained in each case and an interpretation of the motion of the systems when projected along each eigenvector.

It was clear that the principal components of the flexibility of all three simulations were very similar, although they varied in their relative importance. To quantify these differences more rigorously, the three trajectories were combined, and PCA was performed again. This yielded a single set of eigenvectors that apply to all three trajectories. The results of this analysis are also shown in Table 2. Figure 8 shows the variation in the coefficient of each of the three major principal components as a function of time. We see that in all cases the coefficients oscillate, that is, there is no evidence that in any of the simulations, one of the major components of the motion is an irreversible conformational transition (on the simulation time-scale). This is, therefore, further evidence that the equilibration procedure was adequate and that the sampling of the production phase was also sufficient. Figure 9 shows how deformation of the time-averaged structure along each of the major eigenvectors can be related to variation in selected conventional helical parameters; thus, principal component 1 results in major changes

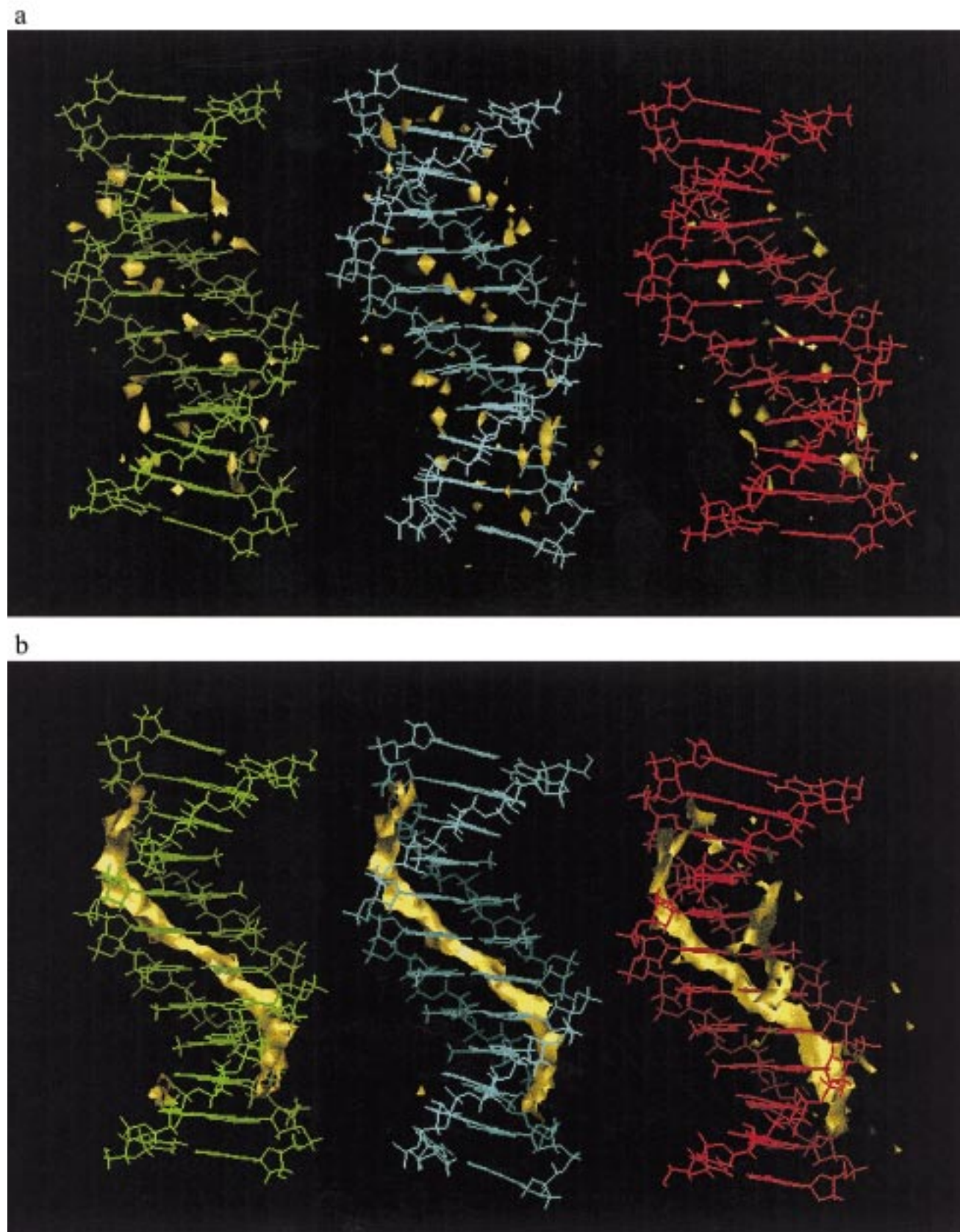


Figure 7. (a) Hydration density plots for the three simulations (left: AAA, center: AIA, right: III), referenced to the time-averaged structure in each case. The contour corresponds to a water density of approximately 4 times that of pure water. (b) Molecular interaction potential (MIP) plots for the three simulations (left: AAA, center: AIA, right: III), referenced to the corresponding time-averaged structures. The -4 kcal/mol contour is shown.

in helical twist, component 2 in junction bending, and component 3 in central (A-tract) bending. Sequence-dependent differences in these motions are evident from the histograms in

Figure 10. For the helical twisting motion (component 1), sequence III alone shows a bimodal distribution; one distribution is centered around coefficient (and thus twist) values also seen

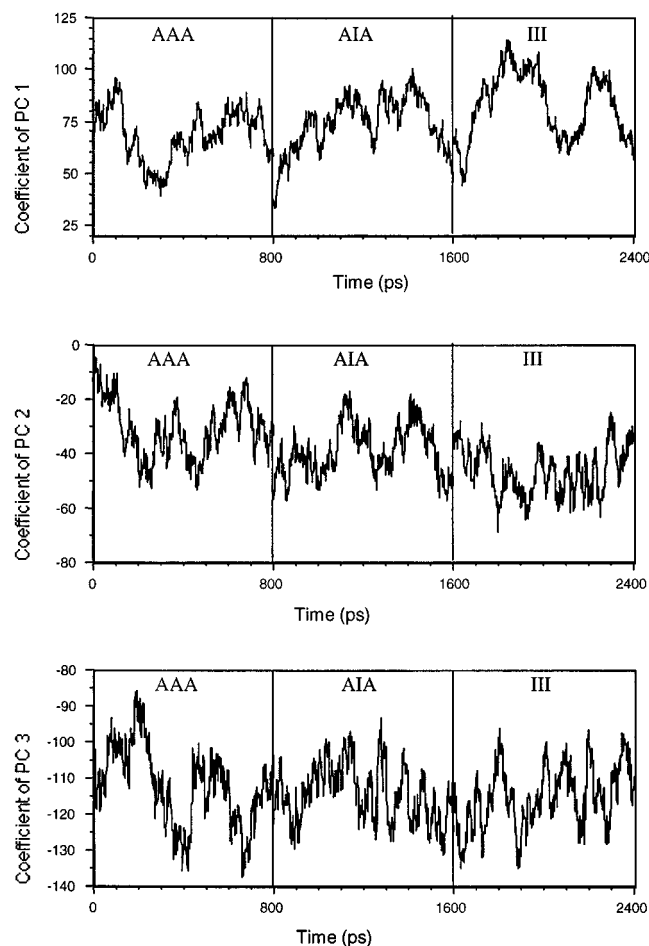


Figure 8. Projection of the three trajectories onto each of the first three principal components. The last 800 ps of each trajectory were joined together so that data for the AAA simulation runs from 0 to 800 ps, for AIA from 800 to 1600 ps, and for III from 1600 to 2400 ps.

for the other two simulations, but there is a second distribution centered around a coefficient of about 95, corresponding to structures with much reduced helical twist. For junction bending (component 2), there are also clear sequence-dependent variations, the average value of the coefficient decreasing from -31.2 (AAA) to -39.1 (AIA) to -46.6 (III). With reference to Figure 9, it can be seen that this corresponds to bend tending to decrease in the order $AAA > AIA > III$. In contrast, for central bending there is no evidence for sequence-dependent effects; all three simulations show very similar distributions. It must be borne in mind that the projection procedure involves the deformation of an arbitrary reference structure—in this case, the time-averaged structure—according to the selected eigenvector. For this reason, the absolute values of helical parameters measured for the projections can contain an offset relative to the original data; however, the shape of the distributions is not affected.

Calculations of Stacking Interactions. The PCA analysis suggested that alterations in stacking interactions accompanying the replacement of A•T base pairs by I•M base pairs might be significant. We calculated the molecular electrostatic potentials of A and I 3.4 \AA above the base plane (Figure 11). Gross differences are evident, which would be expected to contribute toward both the absolute stacking energy and its twist dependence. To examine this more closely, we calculated the base-stacking interactions in the trimers d(AAA)•d(TTT) and d(III)•d(MMM), built in standard B-form geometry but with varying helical twist. The total energy profiles are shown in Figure 12.

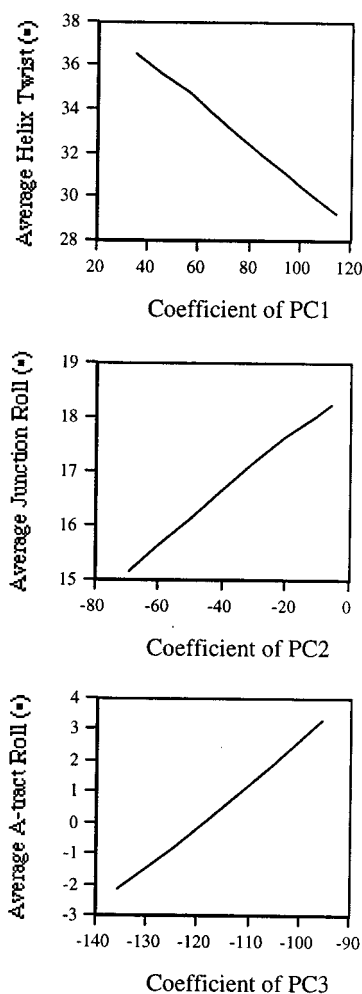


Figure 9. Relationship between the mode of deformation represented by the top three principal components and selected helical parameters. Component 1 (top): relationship to average helical twist; component 2 (center): relationship to the average of the roll at the two A-tract junction steps; component 3 (bottom): relationship to the average roll within the A-tract.

For both sequences, the base-stacking energy becomes more negative as helical twist increases. Stacking interactions are weaker for the I•M sequence, and also somewhat less sensitive to twist, than for the A•T sequence. Decreasing twist from 35° to 30° carries an energy penalty of 1.75 kcal/mol for the A•T sequence, but only 1.34 kcal/mol for the I•M sequence.

Calculations of Phosphate–Phosphate Interactions. One of the most important electrostatic interactions in DNA is that between the phosphate groups. Structural deformations that alter phosphate–phosphate distances are likely to have a major effect on the electrostatic interaction energy of the DNA, and conversely, factors that modulate electrostatic interactions—for example, changes in the ionic strength of the solvent—are likely to affect DNA structure and modes of deformation that involve the phosphate groups. Charge–charge interaction energies vary as r^{-1} ; therefore, to predict how the differing modes of DNA deformation detected by PCA might respond to external factors altering electrostatic interactions, we calculated the average value of the reciprocal of all phosphorus atom–phosphorus atom distances ($\langle r_{pp}^{-1} \rangle$) as the AAA structure was deformed according to each principal component (Table 3). We observe that $\langle r_{pp}^{-1} \rangle$ is particularly sensitive to deformation according to principal component 1 (helical twisting), to a lesser extent to deformation according to component 3 (central bending), and largely

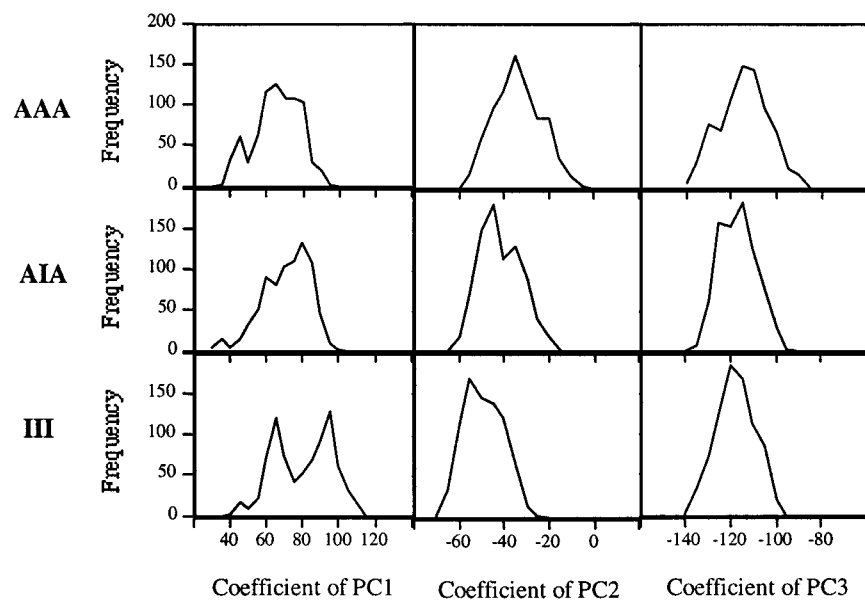


Figure 10. Distribution plots for the principal component projection data shown in Figure 8.

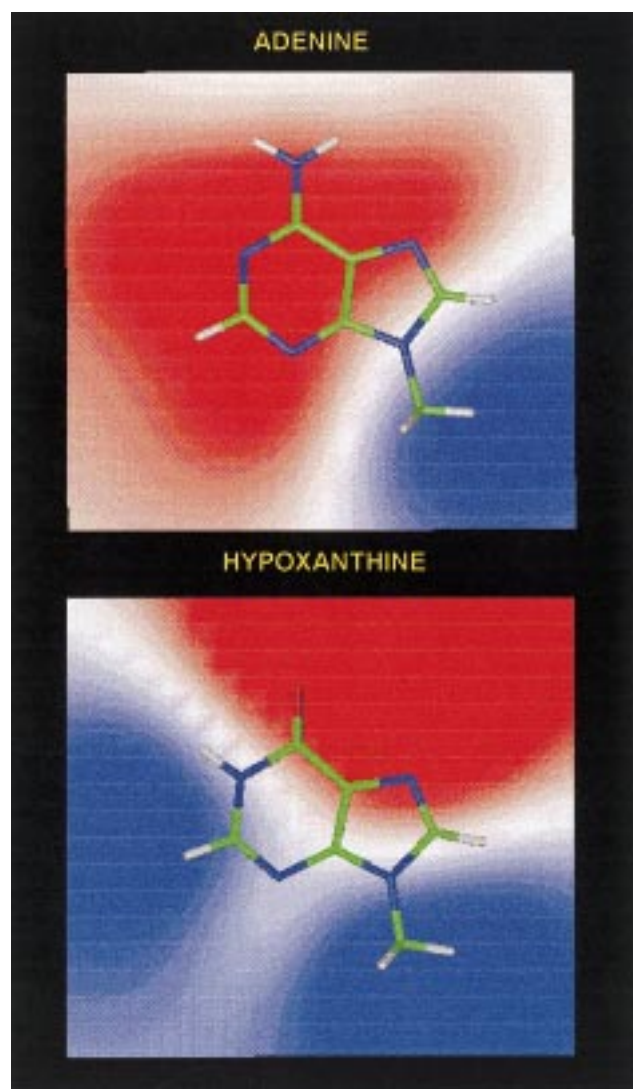


Figure 11. Molecular electrostatic potential maps for the bases A and X, calculated 3.4 Å above the base planes. The color coding covers the range -5 kcal/mol (red) to 5 kcal/mol (blue).

insensitive to DNA deformation according to component 2 (junction bending).

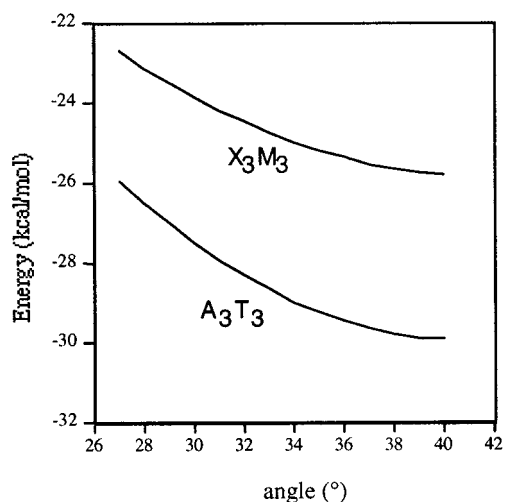


Figure 12. Stacking interaction energy for A_3T_3 and X_3M_3 model systems as a function of helical twist angle.

Table 3. Effect of Projection of the Structures along Each Principal Component on the Average Reciprocal Phosphorus–Phosphorus Atom Separation

principal component	coefficient	$\langle r_{pp}^{-1} \rangle$ ($\text{Å}^{-1} \times 10^2$)	geometry
1	35	6.15	high twist
1	55	6.12	.
1	85	6.01	.
1	114	5.84	low twist
2	-5	6.06	high junction bend
2	-20	6.09	.
2	-50	6.10	.
2	-69	6.06	low junction bend
3	-95	6.10	positive A-tract roll
3	-105	6.08	.
3	-125	6.02	.
3	-135	5.98	negative A-tract roll

Discussion

General Features of the Simulations. All three sequences proved stable over the time scale of the MD simulations and showed good evidence for both equilibration and adequate sampling of the conformational space accessible to the molecules

at room temperature in solution. The ensembles of equilibrated structures for AAA and AIA show rms deviations from the corresponding crystal structures of 1.5–3.5 and 2.0–3.6 Å, respectively, whereas for III the values from the mutated AAA crystal structure are 2.2–4.5 Å. A large part of this deviation arises from the difference in helical twist of the crystal structures and in the MD ensembles. While the crystal structures show a twist close to the canonical value for B-form DNA of 36°, the ensembles show average twist values in the range of 31° (III) to 34° (AAA). This however, is closer to the value of 34.8° estimated for A-tract-containing sequences in solution.²⁴ The increased roll and reduced twist suggest a somewhat “A-like” structure for the bases, but the backbone (and particularly the sugar pucker) remains quite “B-like”. This is in agreement with the simulations of Feig and Pettitt.²⁵ The reduced helical twist (compared with the crystal structures) correlates with a reduction in propeller twist, and the two together result in the cross-strand amino–amino distances, or equivalent depending on the sequence, being greater in the simulations than observed in the crystal structures. The histogram (Figure 5) shows, however, that in all cases there is a significant proportion of structures with N–N separations in the range 3.0–3.5 Å, where amino–amino interactions may be significant.¹²

Interactions with Water and Ions. Analysis of major groove hydration from the MD trajectories failed to show a statistically significant difference between the three sequences. However, the MIP analysis, based on time-averaged structures from the three simulations, did suggest that the major groove of the III structure has a conformation that is more favorable for interaction with polar and charged ligands than that of either other sequence. It is interesting to consider why this feature could not be detected in the MD-based analysis. Preferential hydration sites are created by the suitable alignment of H-bonding groups within the DNA grooves. As the conformation of the DNA varies through the MD trajectory, we may thus expect to see preferential hydration sites created and destroyed. Over time, a diffuse pattern of apparently low-affinity sites thus results. In contrast, the MIP analysis considers a single conformation of the DNA, albeit the time-averaged structure, and thus higher affinity hydration sites are better defined. Neither analysis could detect, as proposed from the crystallographic studies, preferential hydration between base pair planes for highly propeller-twisted structures, but within base pair planes where propeller twist was low. Rather, careful examination of the MIP analysis suggests that the improved hydration of the major groove is particularly associated with the inosine amino group. In the lower propeller-twisted state favored for the III sequence, this group is less involved in bifurcated H bonding, and thus may be more available to interact with the solvent.

Principal Component Analysis. Up to this point, analysis of the MD trajectories had failed to indicate clearly why the III sequence should induce less bending than the AAA or AIA sequences. The PCA, however, presented a much clearer picture of sequence-dependent differences in structure and flexibility. When PCA was applied to each trajectory independently, in each case three major modes of deformation were identified. Animation of these revealed that, by visual inspection, the same three modes of deformation occurred in all three trajectories, although their relative importance varied with sequence. It was clear that a more quantitative comparison of the three simula-

tions would be possible if the analysis produced identical eigenvectors for each, and thus the approach of combining the three trajectories into one was adopted. This was possible because, at the heavy atom level, the three sequences are topologically identical; in addition, it appeared to be a valid approach as we had already seen that, taken separately, the three trajectories yielded very similar eigenvectors anyway. Indeed, PCA of the combined trajectories yielded three major eigenvectors, as before, which represented modes of deformation very similar to those previously observed.

These modes could be related approximately to changes in selected conventional helical parameters. The first mode was strongly related to changes in overall helical twist and also to changes in propeller twist; animation of the mode showed that as the helical twist reduced, so did the propeller twist. All three simulations showed a distribution in the first principal component coefficient centered around a value of 65–70 (Figure 10), corresponding to a helical twist of around 34° (Figure 9). However, the III simulation also showed a second distribution in this coefficient, centered around 30–31° of helical twist. This suggests that the III structure can exist in two substates; in one it shows helical (and propeller) twist values very similar to that of AAA or AIA; in the other approximately equally populated substate it shows much lower helical twist and propeller twist. The second mode of deformation was dominated by in-phase bending at the two A-tract junctions, and could be approximated by roll motion at the two junction steps. The principal component coefficient distributions showed a move to more negative values as the sequence changed from AAA to AIA to III. This trend corresponds to a reduction in the average bend at the junction steps in the order AAA > AIA > III. Visualization of the third mode of deformation revealed that it was dominated by central bending and, in terms of helical parameters, could be approximated by a concerted roll motion at all base steps within the A-tract. This mode, although an important contributor to the overall flexibility of the structures, proved to be relatively insensitive to base sequence; all three simulations showing distributions in the coefficient of this mode that were very similar in both spread and average value, which corresponded to almost zero roll.

Analysis of Stacking Interactions. Calculation of the molecular electrostatic potentials above the base planes in I and A revealed that, despite their steric similarity, these bases are electrostatically very different (Figure 11). We find that this leads to a helical dependence in the stacking interactions of AAA and III sequences that is in accord with the analysis of the MD data. Stacking interactions between I•M base pairs are weaker than those between A•T base pairs, and the energy penalty associated with the adoption of a low-twist state is smaller for the III sequence than the AAA sequence. Partitioning the stacking interaction energy into van der Waals and electrostatic terms, we find that this difference stems almost entirely, as expected, from the electrostatic component of the interaction energy. It should be borne in mind that in this analysis we have considered only a variation in helical twist, all other helix parameters being fixed at values appropriate for B-form DNA. Thus, it does not take into account the correlation between helical twist and propeller twist found from the principal component analysis. This was done to simplify the analysis, since it would otherwise have to consider also the effect of changes in the H-bonding geometry of the base pairs.

Relationship to Other Modeling Studies. The only other recent modeling study on DNA A-tracts is that of Young and

(24) Drak, J.; Crothers, D. M. *Proc. Natl Acad. Sci. U.S.A.* **1991**, *88*, 3074–3078.

(25) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134–149.

Beveridge²⁶ on the 25-mer duplex d(ATAGGCAAAAATAGC-CAAAAATGG). This study also used the AMBER force field and PME method, but also considered the effect of varying salt concentrations. This study, too, showed that junction bending (roll into the major groove) provided a major contribution to the curvature of the sequence and that the A-tracts themselves were generally straight. This study was also able to find evidence for general sequence bending; our study does not include a suitable sequence for comparison. The authors do not comment on the dynamical and static characteristics of their simulations with respect to helical twist or bifurcated H-bonding patterns.

Relationship to Experimental Observations. Relating these results to the observed variations in macroscopic curvature must be done with caution, due to the complexities of the processes involved in determining gel retardation. Relating the results to competing theories of the origins of A-tract-induced bending is somewhat more straightforward. One of the modes of deformation detected by PCA is central bending, which would relate to the wedge model of A-tract-induced DNA bending. However, we observe that in all three sequences central bending occurs equally into the major and minor grooves, so that the time-averaged structure of the A-tract is essentially unbent and that the degree of flexibility in this mode is also sequence-independent. This suggests that the wedge model of A-tract-induced bending cannot be used to explain the sequence dependence in macroscopic curvature detected experimentally. Another mode of DNA deformation that proves to be important in the simulations involves bending at the junctions between the A-tract and the flanking sequences. The flexibility of pyrimidine-purine steps is well-known and is thought to be related to poor base stacking. Olson and others²⁷ have shown this in an analysis of protein-DNA structures, although the enhanced flexibility of CA steps appears more in twist and slide than in roll. Gel studies²⁸ have also shown that CA steps induce DNA curvature. We may relate this to the junction model of A-tract-induced DNA bending but point out that this usually considers the junction between long segments of DNA, whereas we have a short flanking sequence here. Junction bending in these simulations shows a sequence dependence that qualitatively matches the experimental results, which show that macroscopic curvature decreases in the order AAA > AIA > III. However, this result alone does not provide a possible explanation for the more quantitative aspects of this trend, that is, that whereas bending only decreases slightly from AAA to AIA, the reduction on going from AIA to III is much greater. Here, perhaps the simulation results concerning the differing torsional flexibilities of the three sequences may be relevant. The measurement of macroscopic DNA bending requires the accurate phasing of a number of A-tract sequences. Inaccuracies in phasing, resulting from incorrect estimation of the helical repeat, will tend to reduce the observed bend. We find that the III sequence, uniquely, shows a highly populated substate characterized by reduced helical twist and propeller twist, which would produce this effect of reduced apparent bend to some extent, although it is difficult to predict its magnitude. Because of the correlation we observe between helical twist and propeller twist, the low helical-twist substate we detect is also a low propeller-twist substate. The existence of substates with different degrees of propeller twist has been hypothesized for A-tract sequences on the basis of spectroscopic observation of premelt-

ing transitions.²⁹ Ultraviolet resonance Raman spectroscopy of A-tract duplexes suggests that the premelting transition corresponds to a change from a bent structure with high propeller twist and bifurcated hydrogen bonds to a "conventional" B-form structure.³⁰

Toth et al.³¹ have measured DNA curvature in solution by fluorescence resonance energy transfer (FRET). They found that the curvature of A-tract-containing sequences was strongly salt-dependent, increasing from 23° per helical turn at 30 mM NaCl to 41° at 500 mM NaCl. This contrasted somewhat with the results of gel electrophoresis measurements, where below 60 mM NaCl curvature has been reported to be constant or even decrease with increasing salt. One might expect that one of the major effects of increased salt concentration would be to effectively mask phosphate-phosphate electrostatic repulsions. We have analyzed how each of the modes of DNA deformation detected in the simulations, and which might contribute to A-tract-induced bending, might respond to factors modulating phosphate-phosphate interactions (Table 3). We observe that $\langle r_{pp}^{-1} \rangle$ is sensitive to the helical twisting and central-bending modes but insensitive to the junction-bending mode. The observation from the FRET studies that bending increases with salt concentration argues for the involvement of the central-bending mode, since structures with larger positive A-tract roll will be energetically less disadvantaged. In effect, therefore, this observation supports the wedge model of A-tract-induced curvature. We may hypothesize that the reason that the results of gel studies are at variance with this analysis relates to the fact that such studies involve much longer DNA sequences with a greater number of phased A-tracts and thus are much more sensitive to errors in A-tract phasing. In this case, raising the salt concentration, by facilitating helical twisting, may reduce the accuracy of A-tract phasing and consequently act to reduce the observed macroscopic curvature, thus canceling out the effect of increased wedge bending. Further FRET studies, using time-resolved methods,³² may be able to shed more light on this hypothesis.

Conclusions

The major conclusion that can be drawn from these studies is that the assumption that replacing A•T base pairs in an A-tract with I•M pairs does nothing other than alter H-bonding potential in the major groove is simplistic. The substitution also alters base-stacking interactions, and this may be at least as important as H bonding in explaining the observed trends in A-tract-induced DNA bending. These results support the view that bifurcated H bonding may well be a consequence of the formation of the characteristic highly propeller-twisted conformation of A-tracts, but it is not the driving force behind it. The three models of A-tract bending would predict three different outcomes for these simulations. The junction model would predict that the structures AAA to III would show decreasing persistent bend at the A-tract junctions. The A-tract wedge model would predict decreasing central bending through this sequence of structures. The non-A-tract model would, conversely, predict increasing central bending through the sequence. The modeling studies support the junction model of A-tract-induced DNA curvature, characterized by bending at the A-tract

(29) Herrera, J. E.; Chaires, J. B. *Biochemistry* **1989**, *28*, 1993–2000.

(30) Chan, S. S.; Austin, R. H.; Mukerji, I.; Spiro, T. G. *Biophys. J.* **1997**, *72*, 1512–1520.

(31) Toth, K.; Sauerbrey, V.; Langowski, J. *Biochemistry* **1998**, *37*, 8173–8179.

(32) Haran, G.; Haas, E.; Szpikowska, B. K.; Mas, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11764–11768.

(26) Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675–687.

(27) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.

(28) Bolshoy, A.; McNamara, P.; Harrington, R. E.; Trifonov, E. N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2312–2316.

junction steps toward the major groove through roll motions. In contrast, while wedge model A-tract flexibility is observed, bending into the major groove and minor groove occur to an equal extent, yielding a time-averaged structure with a straight A-tract, as is observed crystallographically. Since the I-tract also shows no signs of a persistent bend, these results do not support the alternative, non-A-tract bending wedge model either. It should be emphasized that our rejection of the A-tract wedge model only concerns its applicability to this particular situation—the effect of A•T to I•M substitutions on macroscopic curvature. Our analysis of the role of salt effects on DNA deformability suggests that, in contrast, wedge-type bending may play an

important role in other circumstances. A-tract-containing sequences can show several important modes of structural deformation, and no single model of A-tract-induced bending may provide a satisfactory explanation of all of the experimental observations.

Acknowledgment. We thank Dr. Ben Luisi for helpful discussions. This work has been supported by the Fulbright Commission (E.C.S.), the BBSRC (S.A.H.), the Wellcome Trust (C.A.L., M.O.), and the European Union (Project BI04-CT-96-0509; M.O.).

JA983715Z