The 5-Methyl Group in Thymine Dynamically Influences the Structure of A-Tracts in DNA at the Local and Global Level

Arvind Marathe and Manju Bansal*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore - 12, India

Received: November 20, 2009; Revised Manuscript Received: February 13, 2010

DNA sequences containing a stretch of several A:T basepairs without a 5′-TA-3′ step are known as A-tracts and have been the subject of extensive investigation because of their unique structural features such as a narrow minor groove and their crucial role in several biological processes. One of the aspects under investigation has been the influence of the 5-methyl group of thymine on the properties of A-tracts. Detailed molecular dynamics simulation studies of the sequences d(CGCAAAUUUGCG) and d(CGCAAATTTGCG) indicate that the presence of the 5-methyl group in thymine increases the frequency of a narrow minor groove conformation, which could facilitate its specific recognition by proteins, and reduce its susceptibility to cleavage by DNase I. The bias toward a wider minor groove in the absence of the thymine 5-methyl group is a static structural feature. Our results also indicate that the presence of the thymine 5-methyl group is necessary for calibrating the backbone conformation and the basepair and dinucleotide step geometry of the core A-tract as well as the flanking CA/TG and the neighboring GC/GC steps, as observed in free and protein-bound DNA. As a consequence, it also fine-tunes the curvature of the longer DNA fragment in which the A-tract is embedded.

1. Introduction

Structural and conformational properties of DNA have been observed to vary in a sequence dependent fashion. Of particular interest are DNA sequences containing a stretch of several A:T basepairs without a 5′-TA-3′ step, known as A-tracts, because of their unique structural features such as a narrow minor groove and their crucial role in several biological processes, such as formation of complexes with regulatory proteins,[1−3] as sites for minor groove binding drugs,[4−13] in DNA curvature,[14−17] and in DNA packaging and nucleosome positioning.[3,14,18−23] Hence, for the past two decades, the role played by A-tract sequences in modulating the structural properties of the DNA helix and its biological relevance has been under intense scrutiny. One of the aspects under investigation has been the influence of the 5-methyl group of thymine on the properties of A-tracts. Several experimental studies[24,25] indicated that the 5-methyl group in A-tract thymines did not play a significant role in influencing DNA curvature and also did not significantly influence the structure of the DNA helix, while Hagerman[26] had earlier shown that they did influence DNA curvature but in a complex, sequence dependent fashion. Later, it was shown that the 5-methyl groups were responsible for the damping of the basepair opening fluctuations in A-tracts.[27] The role of the methyl group in DNA cleavage by endonucleases has also been established.[28] However, there do not appear to be any studies comprehensively documenting the influence, if any, of the 5-methyl groups in A-tracts on DNA structure at the basepair and the dinucleotide step levels or its effect on the backbone conformation. The basepair parameters essentially quantify the inherent or induced basepair nonplanarity. Various studies have shown that A:T basepairs constituting the DNA A-tracts have unusually large, negative propeller twist values.[29,30] The dinucleotide step parameters are indicators of basepair stacking efficiency, and their sequence dependent variation modulates the groove width[31] and gives rise to a wide variety of global geometries,[32] including kinked and curved DNA. Variation in basepair and dinucleotide step parameters also influences the formation of cross-strand C−H···O and N−H···O hydrogen bonds, which have been frequently observed in A-tract containing sequences.[29,33−36] The DNA backbone which is involved in more than 50% of the contacts in regulatory protein−DNA complexes[37] could also play a conformation dependent role, apart from the (charge−charge) electrostatic interactions. The noncanonical backbone conformation BII, involving correlated motion of the ε and ζ torsion angles, is observed in several A-tract containing structures and is known to play a role in modulating DNA curvature and recognition by drugs and proteins.[38−47] Moreover, variation in backbone torsion angles can cause movement of the basepair positions, thus affecting the dinucleotide geometry. In the present study, we have carried out extended (12 ns) molecular dynamics simulations on the dodecamer sequence d(CGCAAAUUUGCG) and compared them with the results of simulation on the widely studied and well-characterized sequence d(CGCAAATTTGCG). We show that the thymine 5-methyl group has a significant influence on the dynamic evolution of the DNA structure at the basepair, dinucleotide step, and backbone conformational levels, and consequently may also influence the global structure and play an important role in protein−DNA interactions.

2. Methods

The molecular dynamics (MD) simulation was carried out using AMBER8 with the Duan et al.[48] force field. Coordinates in Arnott B-form fiber model structure[49] were generated using the nucgen module of AMBER8 for the 12-mer DNA sequences d(CGCAAAUUUGCG) (hereafter referred to as A_3U_3) and d(CGCAAATTTGCG) (hereafter referred to as A_3T_3). The system was then neutralized by placing 22 Na^+ counterions at the positions of minimum electrostatic potential. The system of DNA and counterions was immersed in a periodic TIP3P water bath which extended up to a distance of 10 Å from the

* Corresponding author. E-mail: mb@mbu.iisc.ernet.in.

10.1021/jp911055x
Published on Web 04/07/2010

DNA atoms. The final system contained 4110 water molecules and a total of 13094 atoms (including 480 heavy atoms of DNA molecule) for A_3U_3 and 4134 water molecules and a total of 13184 atoms (including 486 heavy atoms of DNA molecule) for A_3T_3. The ion positions were randomized so that no ion was within 5 Å of a DNA atom or any other ion.

Initial minimization of 1000 steps was carried out keeping a large restraint of 500 kcal/mol-Å^2 on the DNA. This was followed by 2500 steps of minimization with no restraints. Then, a constant volume 10 ps molecular dynamics run was done on the system, with a restraint of 100 kcal/mol-Å^2 on DNA, followed by a no restraints minimization for 500 steps. This was followed by successive 10 ps constant pressure molecular dynamics runs, with restraints of 50, 25, 20, 15, 10, and 5 kcal/mol-Å^2, respectively. Minimization of 1000 steps without restraints was then carried out. Finally, the system was heated from 0 to 300 K, in six 2 ps runs, with 50 K increments in temperature, followed by a production run for 12 ns at constant temperature and pressure. All of the analysis was carried out for the snapshots corresponding to the 1−12 ns simulation.

All of the analysis has been carried out excluding the terminal basepairs to eliminate end effects.

The rmsd and hydrogen bond values were calculated using the ptraj module of AMBER8. The distance cutoff between the donor and acceptor atoms was set at 3.6 Å, and the angle cutoff between the donor atom, hydrogen, and acceptor atom was set at 120° for the hydrogen bond calculation.

All of the plots were generated using MATLAB.

2.1. Clustering of Snapshots. Since A-tracts are often characterized by their minor groove width,[31,50] we have clustered the MD snapshots using this feature.

In the fiber model, the minimum interstrand phosphate−phosphate (P···P) separation for a dodecamer with residues numbered 1−12 in strand 1 and 13−24 in strand 2 on the minor groove side are between the phosphates P_5 and P_{24}, P_6 and P_{23}, P_7 and P_{22}, P_8 and P_{21}, and so on. We have used the minimum interstrand P···P separation on the minor groove side to quantify the minor groove width, and it spans two dinucleotide steps or three basepairs. Since our aim was to cluster snapshots based on the minor groove width in the central A-tract region of the molecules, encompassing the fragment d($A_4A_5A_6U(T)_7$-$U(T)_8U(T)_9$) and its complementary sequence d($A_{16}A_{17}A_{18}U$-$(T)_{19}U(T)_{20}U(T)_{21}$), we have classified a snapshot as belonging to a particular groove width range, if the minimum interstrand P···P separation values corresponding to both phosphates P_8 and P_9 lie in that range. The ranges were grouped at intervals of 1 Å, centered on groove width values of 7, 8, 9, ... , 15 Å. Thus, if the minimum interstrand P···P separation from both P_8 and P_9 is between 9.5 and 10.5 Å, the snapshot is assigned to the 9.5−10.5 Å range. The phosphates in strand 2 at minimum distances from P_8 and P_9 were most frequently P_{21} and P_{20}, respectively. Thus, our criteria effectively apply to the region comprising the central four basepairs of the two sequences, viz., $A_5A_6U_7U_8/A_{17}A_{18}U_{19}U_{20}$ and $A_5A_6T_7T_8/A_{17}A_{18}T_{19}T_{20}$.

With our stringent criteria, only 3544 and 3687 out of 11000 snapshots for A_3U_3 and A_3T_3, respectively, were observed to fall within all of the different groove width ranges. For each groove width range, the selected snapshots were observed to be evenly distributed across the 1−12 ns MD range. Of these, the groove width ranges of 8.5−9.5, 9.5−10.5, 10.5−11.5, and 11.5−12.5 Å were considered for detailed analysis, which together include 3116 (87.9% of total snapshots in all of the groove width ranges) and 3506 (95.1% of total snapshots in all of the groove width ranges) snapshots for A_3U_3 and A_3T_3, respectively. Hereafter, these ranges would be represented as 8.5−9.5, 9.5−10.5, 10.5−11.5, and 11.5−12.5.

In addition to minor groove width, the minimum interstrand P···P separation in the fiber model for a dodecamer with residues numbered 1−12 in strand 1 and 13−24 in strand 2 on the major groove side are between the phosphates P_2 and P_{20}, P_3 and P_{19}, P_4 and P_{18}, P_5 and P_{17}, and so on. Hence, these phosphates span three dinucleotide steps or four basepairs. We have used the minimum interstrand P···P separation on the major groove side to quantify the major groove width.

2.2. Calculation of Structural Parameters of the DNA. The backbone parameters as well as the basepair and dinucleotide step parameters were calculated using the NUPARM program, and details have been described in the Supporting Information, sections 1.1 and 1.2.

2.3. Principal Component Analysis (PCA). Principal component analysis was carried out to compare the flexibility of the core A-tract in the two molecules. The heavy atoms of the central six basepairs of all 11000 snapshots of the production run were superimposed with respect to the final minimized structure. A covariance matrix was then constructed for all of the backbone heavy atoms of the central 12 nucleotides and only the phosphate atoms of the central 12 nucleotides. The principal 10 eigenvectors and eigenvalues were calculated in both cases. All of these calculations were done using the ptraj module of AMBER8.

The overlap between sets of eigenvectors for two similar molecules is given by the square root of the cumulative mean square inner product of the eigenvectors in each set.[51] The top 10 eigenvectors of each molecule were observed to contribute predominantly to its motion. Hence, the root-mean-square inner product (RMSIP), quantifying the overlap, was calculated as

$$\mathrm{RMSIP} = \sqrt{\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}(X_i \cdot Y_j)^2}$$

where X and Y denote the eigenvector sets for molecules A and B. A value close to 0 indicates lack of overlap and a value close to 1 indicates nearly identical eigenvector sets, implying overall motion in the same direction for the two molecules.

2.4. Crystal Structure Data Set Selection. The free B-DNA data set described in ref 32 was used for calculating the rotational basepair parameters for the A:T basepair, and the dinucleotide step parameters for the AA/TT, AT/AT, and CA/TG steps. Hereafter, this data set would be referred to as the "thymine crystal data set". For calculating the rotational basepair parameters for the A:U basepair, and the dinucleotide step parameters for the AA/UU, AU/AU, and CA/UG steps, X-ray crystal structures of better than 2 Å resolution and containing deoxy-uracil residues were extracted from the Protein Data Bank.[52] This data set is comprised of structures with the following PDB id's: 1G75, 1G8N, 1G8U, 1G8V, 1SM5, 271D. Hereafter, this data set will be referred to as the "uracil crystal data set".

We have also compared the results of our simulations with the values obtained for two crystal structures of the dodecamer d(CGCAAATTTGCG) (referred to as A_3T_3, as in the case of the dodecamer used in MD) with PDB id's 1S2R[53] and 1D65[54] and with resolutions of 1.5 and 2.2 Å, respectively. No crystal structure of the dodecamer d(CGCAAAUUUGCG) is available for comparison with our results.

2.5. Pseudocrystallographic Refinement. To check for formation of the spine or the ribbon of hydration, and to test

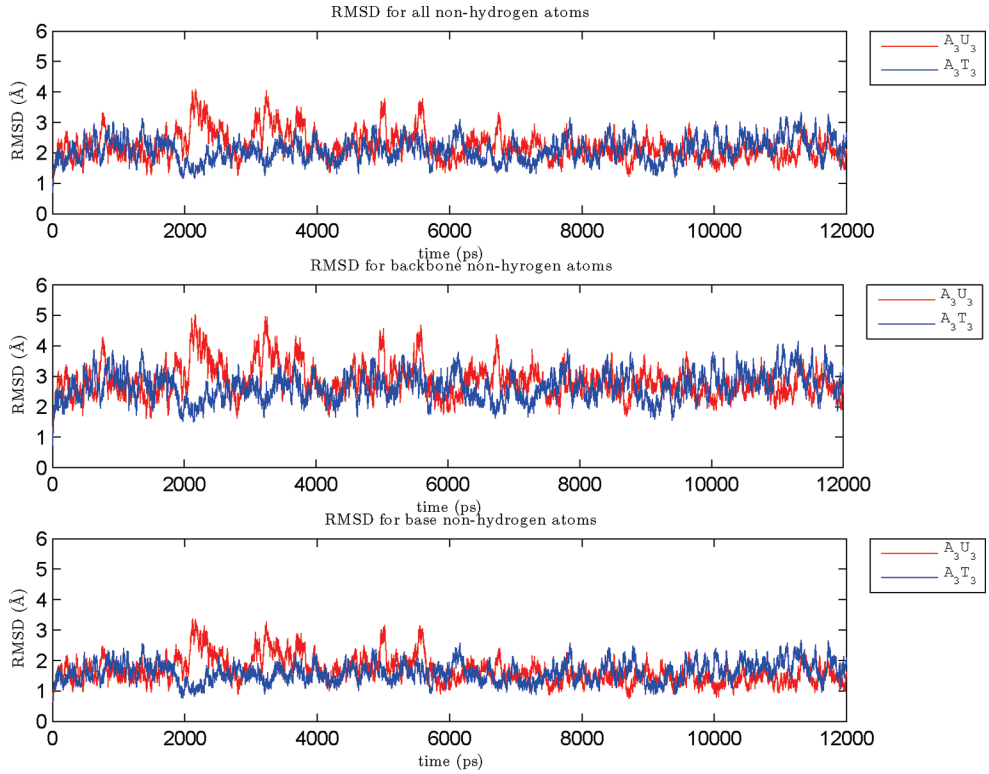

Figure 1. rmsd plot with respect to the final minimized structure for the entire 12 ns range, calculated for all of the heavy atoms, only the backbone heavy atoms, and only the base heavy atoms.

for the presence of ions, pseudocrystallographic refinement was carried out as described in ref 55 for several time windows of 500 ps length, and also for snapshots belonging to each groove width range, by treating them as pseudocontinuous in time.

3. Results

Figure 1 shows the rmsd plots with respect to the final minimized structure for the entire 12 ns range, calculated for all of the heavy atoms, only the backbone heavy atoms, and only the base heavy atoms. The nearly constant rmsd values (in the ranges 2.2 ± 0.5 Å for A_3U_3 and 2.1 ± 0.4 Å for A_3T_3 for the final 11 ns of the production run, for all heavy atoms) show that the molecules were well equilibrated.

The minor groove width values are observed to fluctuate within a small range over the entire trajectory. For the last 11 ns, the shortest interstrand P···P distances from P_8 and P_9 have mean values of 11.2 ± 1.3 and 11.1 ± 1.2 Å, respectively, for A_3U_3 and 10.2 ± 1.0 and 10.2 ± 1.1 Å, respectively, for A_3T_3, as compared to the value of 11.7 Å for the starting fiber model. We compared this with the values obtained for 1S2R and 1D65, the crystal structures of A_3T_3. For 1S2R, the shortest interstrand P···P distances from P_8 and P_9 are 9.8 and 9.6 Å, respectively, while, for 1D65, they are 10.4 and 10.3 Å. The minor groove width observed during the MD simulations is seen to be weakly correlated with the average of the roll, twist, and propeller twist values of the dinucleotide steps and the basepairs comprising the DNA region under consideraton. For example, the smallest interstrand P−P distance from P_8 has correlation coefficient values of 0.4, −0.5, and 0.3 with roll, twist, and propeller twist, respectively, for A_3U_3, while the corresponding numbers for A_3T_3 are 0.4, −0.4, and 0.3. Similarly, the smallest interstrand P−P distance from P_9 has correlation coefficient values of 0.4, −0.4, and 0.3 with roll, twist, and propeller twist, respectively, for A_3U_3, while the corresponding numbers for A_3T_3 are 0.4, −0.4, and 0.2.

TABLE 1: Number of Snapshots Selected for Each Minor Groove Width Range[a]

groove width range (Å)	A_3U_3	A_3T_3
7.5−8.5	3	156
	(0.1)	(4.2)
8.5−9.5	33	437
	(0.9)	(11.9)
9.5−10.5	1254	2066
	(35.4)	(56.0)
10.5−11.5	1412	883
	(39.8)	(23.9)
11.5−12.5	417	120
	(11.8)	(3.3)
12.5−13.5	292	22
	(8.2)	(0.6)
13.5−14.5	126	3
	(3.6)	(0.1)
14.5−15.5	7	0
	(0.2)	(0.0)
total	3544	3687

[a] The numbers in parentheses denote percentage proportion of the total number of selected snapshots, occurring in a particular minor groove width range. A snapshot has been counted "in" only when the minimum interstrand P···P separation from both P_8 and P_9 belong to the given range. For details, see the Methods section. Only snapshots in the 1−12 ns range have been considered.

As shown in Table 1, the A_3U_3 molecule has far lower populations in the narrow minor groove ranges as compared to the A_3T_3 molecule, while the proportion with a wider minor groove is much larger. There are 437 snapshots for A_3T_3 in the 8.5−9.5 range, and the corresponding number for A_3U_3 is only 33. On the other hand, there are only 120 snapshots for A_3T_3 in the 11.5−12.5 range, while the corresponding number for A_3U_3 is 417. The maximum number (2066) of snapshots for A_3T_3 occurs in the 9.5−10.5 range, and it is more than twice the number of snapshots in any other range. For A_3U_3, there is a

TABLE 2: Average and Standard Deviation Values for the Basepair Rotational Parameters for the A:U, A:T, and G:C Basepairs, for the Snapshots in the Four Minor Groove Width Ranges Selected for Detailed Analysis (Denoted by MD (SS)), the Entire Trajectory of Last 11 ns (Denoted by MD (11 ns)), and Also for the X-ray Crystal Structure Data Set (Denoted by X-tal), Described in the *Methods* Section[a]

	buckle (deg)			propeller (deg)			opening angle (deg)		
basepair	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal
A:U	4.6	3.5	3.1	−15.4	−14.9	−16.0	1.2	1.1	4.0
	(12.9)	(13.0)	(3.9)	(8.6)	(8.7)	(3.1)	(5.5)	(5.5)	(3.1)
	18696	66000	16	18696	66000	16	18696	66000	16
A:T	3.2	3.0	−1.7	−15.6	−15.6	−13.9	1.0	1.0	4.3
	(11.2)	(11.4)	(6.1)	(8.3)	(8.4)	(5.2)	(5.1)	(5.1)	(4.2)
	21036	66000	135	21036	66000	135	21036	66000	135
G:C[b]	−3.7	−3.4	2.8	−3.1	−3.7	−10.8	−1.3	−1.3	−0.6
	(13.0)	(13.3)	(7.7)	(10.2)	(10.2)	(5.7)	(3.6)	(3.5)	(2.7)
	12464	44000	48	18696	44000	48	18696	44000	48
G:C[c]	−2.1	−2.1	−0.5	−6.0	−5.8	−9.2	−1.3	−1.3	−0.7
	(12.2)	(12.1)	(9.1)	(9.2)	(9.2)	(7.3)	(3.3)	(3.3)	(3.4)
	14024	44000	179	14024	44000	179	14024	44000	179

[a] The numbers in parentheses denote standard deviation values. The numbers below the standard deviation values denote the number of data points. [b] G:C basepairs in the X-ray crystal structures with better than 2 Å resolution and containing deoxy-uracil residues. See the Methods section for details. [c] G:C basepairs in the free B-DNA crystal structure data set used in ref 32.

rather broad peak between 9.5 and 11.5, with 1254 snapshots in the 9.5−10.5 range and 1412 snapshots in the 10.5−11.5 range. Thus, the A_3U_3 structure has a clear bias toward a more open minor groove conformation.

We performed PCA for the snapshots of the last 11 ns for both molecules to test if the bias observed in A_3U_3 is a static bias or an indicator of flexibility. The RMSIP of the first 10 eigenvectors for the two molecules computed for the phosphate atoms of the central 12 nucleotides is 0.94. When we consider all of the backbone heavy atoms of the central 12 nucleotides, the RMSIP is 0.92. This indicates that the two sets of principal modes are nearly identical. Thus, the wider minor groove observed for a larger number of snapshots of A_3U_3 as compared to A_3T_3 does not indicate greater flexibility for A_3U_3 but is representative of a static structural feature.

Figure 1 in the Supporting Information shows representative snapshots belonging to the different groove width ranges, for both of the molecules. It can be seen that, despite similar groove width at the core, the two molecules differ considerably in several other features.

The distributions of different backbone conformations for all 11000 snapshots over the entire trajectory are observed to be very similar to the corresponding distributions for the subset of snapshots with defined minor groove. The average and standard deviation values for the basepair rotational parameters as well as the dinucleotide step parameters are shown in Tables 2−4, for the snapshots in the four groove width ranges. The values obtained for all 11000 snapshots are also listed and found to be very similar to those of the subset with defined minor groove.

For both molecules, the major groove width values are also observed to fluctuate within a small range over the entire trajectory. For all of the selected snapshots in the four groove width ranges, the minimum interstrand P···P distance from P_5, encompassing the region comprised of the four central basepairs, viz., $A_5A_6U(T)_7U(T)_8/A_{17}A_{18}U(T)_{19}U(T)_{20}$, is 18.3 ± 2.0 Å in A_3U_3 and 17.5 ± 1.8 Å in A_3T_3. For 11 ns of MD, the corresponding values are 18.3 ± 2.0 and 17.6 ± 1.8 Å, respectively. In the fiber model, the corresponding value is 17.2 Å, while in the crystal structures 1S2R and 1D65 the corresponding values are 17.6 and 16.5 Å, respectively. Thus, the major groove width value observed during the simulation of A_3T_3 matches the one observed in the high resolution crystal structure 1S2R.

No significant differences were observed between the spines and the ribbons of hydrations between A_3U_3 and A_3T_3, for any of the groove width ranges, or for any of the time ranges. Hence, these results have not been reported here.

3.1. Backbone Parameters. The backbone torsion angles assume characteristic values and variations in these angles influence the structure and function of DNA. The backbone torsion angles β, δ, and χ do not show any significant variation for different groove width ranges within or across the two molecules.

3.1.1. α/γ Transitions and Ion Motions. The α/γ torsion angles in DNA crystal structures most often take up the canonical g^-/g^+ state, but occasionally, they may take up the noncanonical t/t or g^+/g^- states.[32,56] We did a systematic study of the α/γ transitions for all of the steps over the entire trajectory. It is observed that some of the α/γ torsion angles in step $G_2C_3/G_{22}C_{23}$ in both molecules assume noncanonical conformations for a few nanoseconds. In the case of A_3U_3, α_3/γ_3 flips to g^+/t around 8 ns and continues in that region for the rest of the simulation, while, for A_3T_3, γ_3 assumes a t conformation at 8.5 ns and continues in that region for the rest of the simulation. Similarly, γ_{22} in step $C_3A_4/T_{21}G_{22}$ assumes a t conformation around 7.5−8.0 ns and stays there for the rest of the simulation. All other α/γ values in both of the molecules remain stable in the canonical g^-/g^+ region.

Long nucleic acid simulations of the order of 50 ns performed using the Duan et al. force field[48] indicate that α/γ transitions to noncanonical conformations persist for considerable time periods and are coupled to ion binding and transitions in twist and minor groove width.[57] We observe no changes in twist correlated to any of the persistent α or γ transitions. The width of the minor groove region encompassing basepairs G_2:C_{23}, C_3:G_{22}, and A_4:T_{21} is observed to reduce by ∼1 Å around 8.0 ns and remains in that region for the rest of the simulation. Transient ion density is observed on the minor groove side of step $C_3A_4/T_{21}G_{22}$, in the 8.0−8.5 ns range. As mentioned previously, the snapshots belonging to each of the selected groove width ranges are spread uniformly throughout the simulation trajectory, and hence, the number of snapshots contributed by the region of the trajectory, affected by ion binding at step $C_3A_4/T_{21}G_{22}$, the consequent α/γ transition, and possibly correlated change in minor groove width, is small.

TABLE 3: Average and Standard Deviation Values for the Rotational Parameters for the Dinucleotide Steps in the A_3U_3 and A_3T_3 Molecules, for the Snapshots in the Four Minor Groove Width Ranges Selected for Detailed Analysis (Denoted by MD (SS)), the Entire Trajectory of Last 11 ns (Denoted by MD (11 ns)), and Also for the X-ray Crystal Structure Data Set (Denoted by X-tal), Described in the *Methods* Section[a]

	tilt (deg)			roll (deg)			twist (deg)		
dinucleotide step	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal
AA/UU	−0.2	−0.2	1.2	4.2	4.5	1.0	33.3	33.1	36.0
	(3.2)	(3.2)	NA	(5.8)	(6.0)	NA	(5.1)	(5.2)	NA
	12464	44000	2	12464	44000	2	12464	44000	2
AA/TT	−1.9	−1.9	−0.7	2.0	2.2	0.2	34.5	34.4	36.1
	(2.9)	(2.9)	(1.6)	(5.4)	(5.5)	(4.0)	(6.0)	(6.1)	(3.9)
	14024	44000	45	14024	44000	45	14024	44000	45
AU/AU	0.1	−0.0	1.5	1.2	2.0	−0.3	34.0	33.5	32.2
	(3.0)	(3.0)	NA	(5.1)	(5.5)	NA	(3.8)	(4.0)	NA
	3116	11000	1	3116	11000	1	3116	11000	1
AT/AT	0.1	0.1	−0.4	−0.9	−0.8	−0.4	33.8	33.6	33.5
	(2.2)	(2.2)	(1.9)	(4.0)	(4.3)	(3.8)	(3.1)	(3.4)	(3.9)
	3506	11000	24	3506	11000	24	3506	11000	24
CA/UG	−0.0	−0.0	NA	7.4	7.9	NA	31.2	31.2	NA
	(3.5)	(3.5)	NA	(6.5)	(6.5)	NA	(7.5)	(7.2)	NA
	6232	22000	0	6232	22000	0	6232	22000	0
CA/TG	−0.0	0.0	−0.9	9.9	9.8	−2.1	26.0	26.1	45.9
	(3.5)	(3.6)	(1.3)	(6.3)	(6.3)	(6.4)	(7.8)	(7.9)	(6.4)
	7012	22000	16	7012	22000	16	7012	22000	16
GC/GC[b]	−0.3	−0.3	−0.5	0.6	1.0	−7.2	30.3	30.5	40.1
	(3.2)	(3.2)	(2.3)	(6.0)	(5.9)	(3.8)	(7.8)	(7.4)	(2.1)
	6232	22000	10	6232	22000	10	6232	22000	10
GC/GC[c]	−0.6	0.6	−0.6	−1.6	−1.5	−3.7	34.5	34.5	38.2
	(2.8)	(2.8)	(2.4)	(5.9)	(5.8)	(5.7)	(6.5)	(6.4)	(2.8)
	7012	22000	13	7012	22000	13	7012	22000	13

[a] The numbers in parentheses denote standard deviation values. The numbers below the standard deviation values denote the number of data points. [b] GC/GC dinucleotide steps in the X-ray crystal structures with better than 2 Å resolution and containing deoxy-uracil residues. See the Methods section for details. [c] GC/GC dinucleotide steps in the free B-DNA crystal structure data set used in ref 32.

TABLE 4: Average and Standard Deviation Values for the Translational Parameters for the Dinucleotide Steps in the A_3U_3 and A_3T_3 Molecules, for the Snapshots in the Four Minor Groove Width Ranges Selected for Detailed Analysis (Denoted by MD (SS)), the Entire Trajectory of Last 11 ns (Denoted by MD (11 ns)), and Also for the X-ray Crystal Structure Data Set (Denoted by X-tal), Described in the *Methods* Section[a]

	shift (Å)			slide (Å)			rise (Å)		
dinucleotide step	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal	MD (SS)	MD (11 ns)	X-tal
AA/UU	0.1	0.1	0.2	−0.7	−0.7	−0.5	3.3	3.3	3.1
	(0.5)	(0.5)	NA	(0.6)	(0.6)	NA	(0.2)	(0.2)	NA
	12464	44000	2	12464	44000	2	12464	44000	2
AA/TT	−0.2	−0.2	0.1	−0.4	−0.4	−0.1	3.3	3.3	3.2
	(0.6)	(0.6)	(0.3)	(0.6)	(0.6)	(0.4)	(0.2)	(0.2)	(0.1)
	14024	44000	45	14024	44000	45	14024	44000	45
AU/AU	0.0	−0.0	−0.0	−0.9	−0.9	−0.4	3.3	3.3	3.1
	(0.5)	(0.5)	NA	(0.4)	(0.4)	NA	(0.1)	(0.1)	NA
	3116	11000	1	3116	11000	1	3116	11000	1
AT/AT	0.0	0.0	−0.0	−0.9	−0.8	−0.3	3.3	3.3	3.2
	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.3)	(0.2)	(0.2)	(0.1)
	3506	11000	24	3506	11000	24	3506	11000	24
CA/UG	−0.2	−0.2	NA	−0.2	−0.2	NA	3.4	3.4	NA
	(0.6)	(0.6)	NA	(0.8)	(0.8)	NA	(0.2)	(0.2)	NA
	6232	22000	0	6232	22000	0	6232	22000	0
CA/TG	−0.4	−0.4	−0.1	−0.5	−0.4	1.7	3.4	3.4	3.3
	(0.7)	(0.7)	(0.2)	(0.6)	(0.6)	(1.2)	(0.2)	(0.2)	(0.1)
	7012	22000	16	7012	22000	16	7012	22000	16
GC/GC[b]	−0.2	−0.2	−0.3	−0.2	−0.2	0.6	3.4	3.4	3.4
	(0.7)	(0.7)	(0.8)	(0.7)	(0.8)	(0.1)	(0.2)	(0.2)	(0.1)
	6232	22000	10	6232	22000	10	6232	22000	10
GC/GC[c]	−0.2	−0.2	−0.3	−0.0	−0.0	0.4	3.4	3.4	3.3
	(0.8)	(0.8)	(0.7)	(0.5)	(0.5)	(0.5)	(0.2)	(0.2)	(0.2)
	7012	22000	13	7012	22000	13	7012	22000	13

[a] The numbers in parentheses denote standard deviation values. The numbers below the standard deviation values denote the number of data points. [b] GC/GC dinucleotide steps in the X-ray crystal structures with better than 2 Å resolution and containing deoxy-uracil residues. See the Methods section for details. [c] GC/GC dinucleotide steps in the free B-DNA crystal structure data set used in ref 32.

Hence, the effect of these changes on the overall results of the A_3T_3 simulation is negligible.

We observe transient ion densities around different regions during different time zones and also for the pseudocontinuous

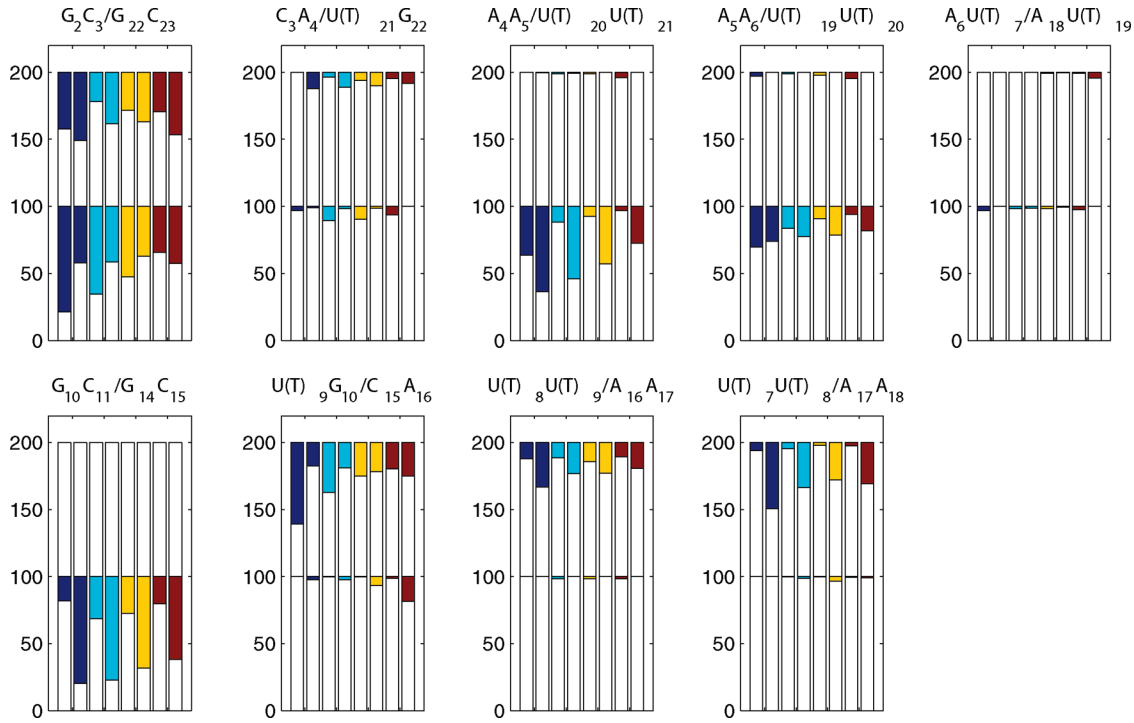

Figure 2. Strand-wise percentage distribution of BI and BII backbone conformations (for details, refer to the Methods section) for structures in each minor groove width range represented as bars for both of the molecules. The values between 0 and 100 represent the percentage distribution for strand 1, and the values between 100 and 200 represent the percentage distribution for strand 2. The white space in each bar represents the percentage proportion of BI conformation for the respective strand, and the colored space represents the proportion of BII conformation for the respective strand. The color code is as follows: dark blue, 8.5−9.5 Å; sky blue, 9.5−10.5 Å; yellow, 10.5−11.5 Å; maroon, 11.5−12.5 Å. For each minor groove width range, the bar on the left represents the values for A_3U_3 and the bar on the right represents the values for A_3T_3. For example, for the $G_2C_3/G_{22}C_{23}$ step in A_3U_3, strand 1 has a 21.2% proportion of BI conformation and 78.8% proportion of BII conformation, while strand 2 has a 57.6% proportion of BI conformation and 42.4% proportion of BII conformation.

trajectories encompassing snapshots with similar minor groove widths. In the case of A_3U_3, a striking observation with regards to ion motion is the trapping of an ion at ∼6 ns at the center of the A-tract on the minor groove side. The ion stays trapped in that region for the rest of the simulation. However, the ion densities do not correlate with changes in the backbone parameters, basepair parameters, or dinucleotide step parameters. The LEaP module in AMBER places the first counterion at the position where we observe the trapped ion, indicating that it is the position of minimum electrostatic potential. We carried out simulations of both A_3U_3 and A_3T_3 using the initial ion placement by LEaP. In these simulations, the ion placed at the center of the A-tract on the minor groove side stays there for the entire duration of 12 ns. However, the structural parameters for the simulation with or without randomizing initial ion positions are similar for both A_3U_3 and A_3T_3. This observation confirms reports by an earlier study[55] that ion binding is not a cause of groove narrowing of the A-tract. The lack of strong ion densities in the minor groove for the pseudocontinuous trajectories encompassing snapshots with groove width in the range 8.5−9.5 or 9.5−10.5 Å further corroborates this observation. Since no significant difference was seen in the results obtained with the initial ion positions being randomized as compared to those where the ion positions were not randomized, the latter results have not been discussed. Since the ion densities do not seem to be influencing the parameters of interest in this study, the ion motions have not been discussed.

A detailed description of the dinucleotide step parameters follows in a later section.

3.1.2. BI/BII Transitions. The major backbone flexibility is observed in the occurrence of BI and BII conformations. However, the variations in ε and ζ are observed to be uncorrelated to the α/γ transitions discussed in the previous section.

Figure 2 shows the percentage distribution of $\varepsilon-\zeta$ values for the two molecules in terms of BI and BII conformations, for the different groove width ranges. It is seen that, within the core A-tract, the AA/UU step has a far smaller proportion of BII conformation compared to AA/TT, while the AU/AU and AT/AT steps have negligible proportions of BII conformation. Among the selected snapshots, the overall proportion of AA/UU steps with one or both strands in the BII conformation is 10.1%, while the corresponding number for AA/TT is 32.5%. This correlates well with the larger roll and smaller twist values for most of the AA/UU steps, compared to AA/TT steps. For both structures, the BII conformation is seen only for the adenine nucleotides. The flanking nucleotide C_{15} in A_3U_3 has a larger proportion of BII conformation than the corresponding nucleotide in A_3T_3. Taken together, the $C_3A_4/U_{21}G_{22}$ and $U_9G_{10}/C_{15}A_{16}$ steps have 21.5% of snapshots in the defined groove width ranges with one or both strands in the BII conformation, while the corresponding number for the $C_3A_4/T_{21}G_{22}$ and $T_9G_{10}/C_{15}A_{16}$ steps is 17.2%. With the exception of G_{14}, all of the guanines assume the BII conformation in a large number of cases. With the exception of G_2, the proportion of guanines in the BII conformation in A_3U_3 is always smaller than the proportions for the corresponding guanines in A_3T_3. The $G_2C_3/G_{22}C_{23}$ and $G_{10}C_{11}/G_{14}C_{15}$ steps together have 31.2 and 65.3% data points in A_3U_3 and A_3T_3, respectively, with at least one strand in the BII conformation. Overall, the selected MD snapshots have 16.5 and 32.9% steps in A_3U_3 and A_3T_3, respectively, with one or both strands in the BII conformation. For an entire MD trajectory

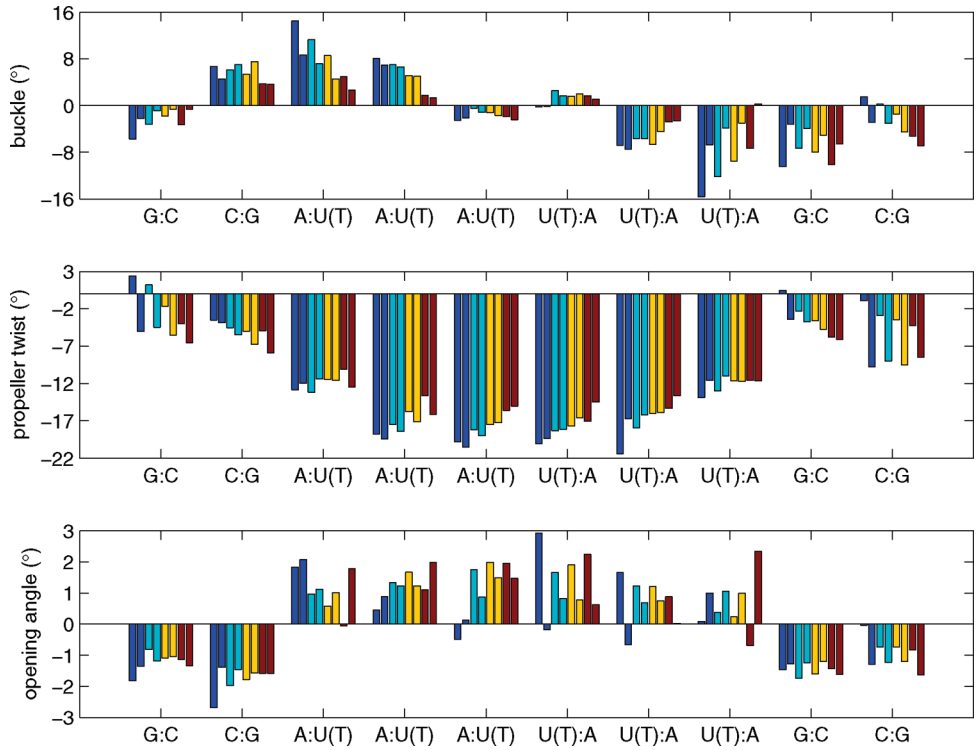

Figure 3. Mean values of the basepair parameters buckle, propeller, and opening angle for the structures occurring in each minor groove width range represented as bars for both of the molecules. The rest of the specifications are as detailed in the caption for Figure 2.

of 11 ns, the corresponding numbers are 15.7 and 33.6%. Thus, the presence of the thymine 5-methyl group leads to a much larger proportion of BII conformation, particularly in the GC/GC sequences, and in the adenines paired to thymines.

There are no CA/UG steps in the uracil crystal data set and only one AU/AU step with both strands in the BI conformation. Similarly, both of the AA/UU steps in the data set have both strands in the BI conformation. Of the 10 GC/GC steps, 8 (80%) have one or both strands in the BII conformation. In the thymine crystal data set, 2 out of 24 (8.3%) AT/AT steps, 7 out of 45 (15.6%) AA/TT steps, 12 out of 16 (75.0%) CA/TG steps, and 8 out of 13 (61.5%) GC/GC steps have one or both strands in the BII conformation.

3.2. Basepair Parameters. The basepair rotational parameters buckle, propeller twist, and opening angle are shown in Figure 3 for each groove width range. The average and standard deviation values for the basepair rotational parameters for the A:U, A:T, and G:C basepairs are shown in Table 2, for the snapshots in the four groove width ranges selected for detailed analysis, the entire trajectory of 11 ns, and also for the X-ray crystal structure data sets.[32] The basepair translational parameters have negligible variation within and across the two molecules, and hence have not been reported here.

As seen in Figure 3, within the A-tract, the mean value for buckle decreases with the widening of the groove for both molecules. A_3U_3 generally has a larger buckle value than A_3T_3 for all of the groove width ranges, particularly for the outer A:U/A:T basepairs, indicating that the presence of the 5-methyl group impedes buckling. As shown in Table 2, the buckle values for the A:U basepair are similar for the MD and the crystal structure data sets, while the corresponding value for the A:T basepair is in fact negative for the crystal structure data set, unlike its MD counterparts. This can perhaps be explained by the fact that only the two central A:U and A:T basepairs (i.e., basepair 6 and basepair 7) in the MD data sets have negative values of buckle while the outer basepairs have large positive buckle values, indicating the sequence dependence of this parameter. The buckle values for the outer G:C basepairs are more negative for A_3U_3 as compared to the corresponding basepairs in A_3T_3. On the other hand, the crystal structure buckle values are positive for the G:C basepairs in the uracil crystal data set but are small negative for the corresponding basepairs in the thymine crystal data set.

The propeller twist values also decrease as the groove widens, as reported for the crystal structures. For each groove width range, the A:U/A:T basepairs at the center of the molecule have a much larger magnitude of propeller twist as compared to the outer basepairs. There is no systematic trend for the small differences between the two molecules. The small difference between the crystal structure data set average propeller twist value for the A:T basepair and the corresponding MD data set values might owe its origin to wider sequence variation in the crystals. For the G:C basepairs, the propeller values for MD snapshots of A_3U_3 are significantly smaller (by 2.3−2.9°) than those for the corresponding basepairs in A_3T_3. The propeller values for the G:C basepairs in MD snapshots of both molecules are also significantly smaller than the crystal structure values (by 3.4−7.1°).

In the case of opening angle, it is hard to discern a trend with the change in groove width. The opening angle values are larger for the outer A:U(T) basepairs as compared to the inner ones. The four central A:U basepairs have significantly larger opening angle values than the corresponding A:T basepairs, while the outer A:U basepair has smaller opening angle values than the corresponding A:T basepair. The crystal structure data set values of opening angle for A:U are similar to those for A:T. Further, both A:U and A:T basepairs in the crystal structure data set have slightly larger positive average opening angle values compared to corresponding values for the MD data sets. It should however be noted that the crystal structure data set values have larger contributions from A:U(T) basepairs occurring in non-A-tract sequences, while the MD values correspond

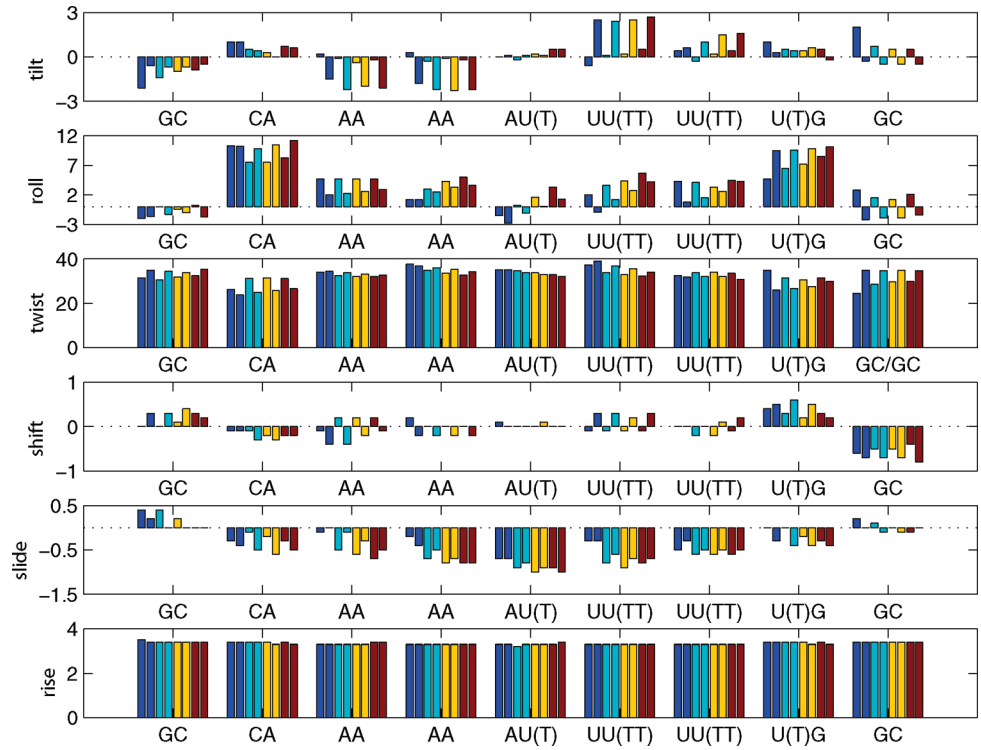

Figure 4. Mean values of the dinucleotide step parameters tilt, roll, twist (in deg), shift, slide, and rise (in Å) for the structures occurring in each minor groove width range represented as bars for both of the molecules. The rest of the specifications are as detailed in the caption for Figure 2.

only to A:U(T) basepairs constituting the A-tracts in this study, leading to smaller opening angles, as indicated by damping being correlated with the length of an A-tract.[27] The opening angle values for G:C basepairs in A_3U_3 are similar to those in A_3T_3. In both cases, the crystal structure values are slightly smaller.

3.3. Dinucleotide Step Parameters. The dinucleotide step parameters tilt, roll, and twist which define the rotation about the dimer *x*, *y*, and *z* axes, respectively, and the translational parameters shift, slide, and rise along the dimer *x*, *y*, and z axes, respectively, are shown in Figure 4. The average and standard deviation parameter values for the dinucleotide steps in the two molecules are shown in Tables 3 and 4, for the snapshots in the four groove width ranges selected for detailed analysis, the entire trajectory of 11 ns, and also the X-ray crystal structure data sets.

As seen in Figure 4, the tilt values are quite small (<3.0°) throughout for both of the molecules, but for A_3U_3, they are nearly 0°. A_3U_3 also seems to have little or no translational motion along the *x*-direction (shift) as compared to A_3T_3. There is very little variation in the rise parameter for either molecule.

Within the A-tract, the roll, twist, and slide values for the two molecules are observed to vary in a correlated fashion. As expected, an increase in groove width leads to larger roll and smaller twist and slide values. AA/UU steps have slightly larger roll, smaller twist, and more negative slide compared to AA/TT steps, while the AU/AU step has slightly more positive roll and a negligible difference in twist and slide as compared to the AT/AT step. The most significant difference between the two molecules is observed for the CA/U(T)G steps, where the roll values for CA/UG are significantly smaller (by 2−4°) than those for CA/TG, with correspondingly larger values of twist (by ∼5°) and less negative values of slide (by 0.2 Å). The large roll, small twist, and negative slide at the CA/U(T)G step gets compensated at the adjacent GC/GC steps on both ends. As a result, the averages of the roll, twist, and slide values over the entire molecule, for the different groove width ranges (data not shown), are similar for the two molecules.

As seen in Tables 3 and 4, the average values of basepair and dinucleotide step parameters for the crystal structure data sets differ in a few cases from the corresponding values obtained for the MD data sets. For most of the step parameters, the standard deviation values for the crystal structure data sets are smaller than those for the MD data set and might be due to the constraints imposed by packing forces in crystals. The most striking difference between the two data sets occurs in the case of the roll, twist, and slide values for the CA/TG step, and to a lesser extent for the neighboring GC/GC step in A_3U_3. While the CA/TG step in the MD data sets has large positive roll, very small twist, and negative slide, the same step in the thymine crystal data set has negative roll, very large twist, and positive slide. Similarly, the GC/GC step in the uracil crystal data set has negative roll, large twist, and small positive slide. This is owing to the unusually large proportion of steps in the crystal data sets with one or both nucleotides at the 5′-end being in the BII conformation, which skews the average geometry toward negative roll, larger twist, and positive slide. However, in the MD data sets, the proportion of BII conformation for the CA/U(T)G steps is very small for both of the molecules. The presence of two distinct conformations for the CA/TG step in the thymine crystal data set is also evident from the standard deviation value for the slide parameter for that step, which is twice the corresponding values for the MD data sets. As mentioned earlier, the proportion of MD snapshots with the BII conformation in one or both strands, for the GC/GC step, is 31.2% for A_3U_3. This is far lower than the corresponding proportion for the crystal structure data set and is reflected in the parameter values for this step.

The crystal structure data set average values for all of the dinucleotide sequences match the range of values obtained for

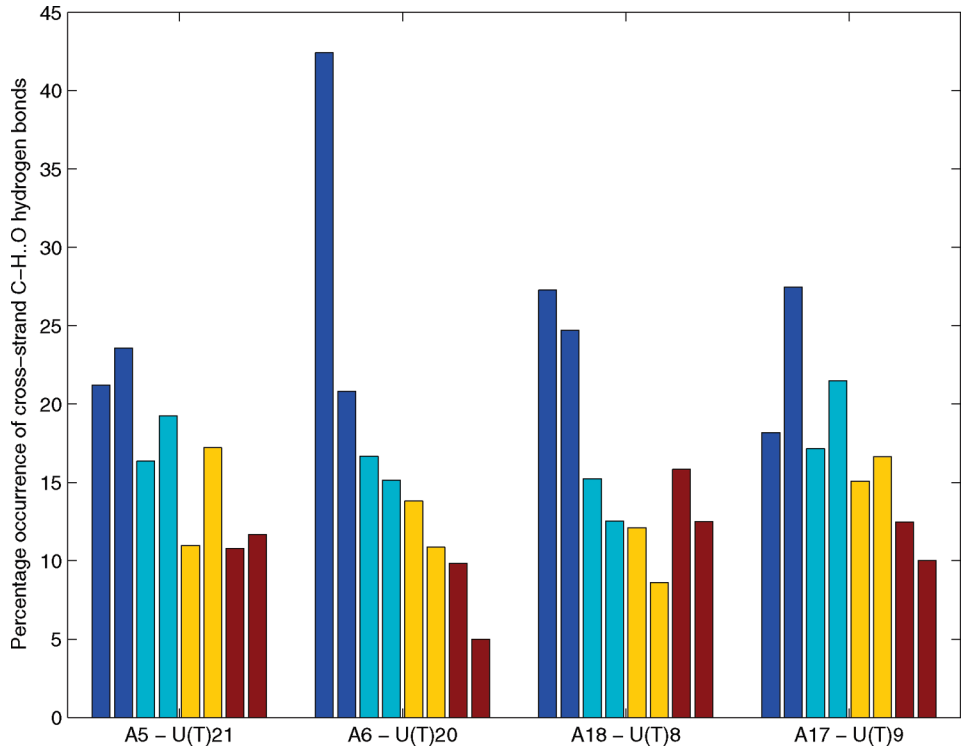

Figure 5. Occurrence of various $C_2-H_2\cdots O_2$ cross-strand hydrogen bonds on the minor groove side as a fraction (in percentage) of the total number of structures in each minor groove width range. In all of the cases, the donor C_2 belongs to the adenine on the 3′-side of an AA step and the acceptor O_2 belongs to the uracil/thymine bound to the adenine on the 5′-side of the same step. For assigning a hydrogen bond, the distance cutoff between the donor and acceptor atoms was set at 3.6 Å and the angle cutoff between the donor atom, hydrogen, and acceptor atom was set at 120°. The rest of the specifications are as detailed in the caption for Figure 2.

the MD structures with narrow minor groove width, ranging from 8.5 to 10.5 Å.

3.4. Cross-Strand Hydrogen Bonds. Cross-strand N−H···O and C−H···O hydrogen bonds across the major and minor grooves have been observed in several A-tract containing crystal structures[29,33−36] and molecular dynamics snapshots[55,58,59] and expected to contribute to their structural properties. We examined the MD snapshots for the cross-strand hydrogen bonds in the major as well as minor grooves, for the different groove width ranges for both molecules. The distance cutoff between the donor and acceptor atoms was set at 3.6 Å, and the angle cutoff between the donor atom, hydrogen, and acceptor atom was set at 120° for assigning a hydrogen bond. The percentage occurrence of N−H···O cross-strand hydrogen bonds in the major grooves was observed to be negligibly small (data not shown), in agreement with previous studies.[55,58,59] However, a significantly larger number of potential $C_2-H_2\cdots O_2$ type of hydrogen bond interactions was observed in the minor groove, where the donor C_2 belonged to the adenine on the 3′-side of an AA/UU(TT) step and the acceptor O_2 belonged to the uracil/thymine paired to the adenine on the 5′-side of the same step. Figure 5 shows the percentage distribution of these bonds for the four groove width ranges, for both of the molecules.

We observe that the proportion of potential $C_2-H_2\cdots O_2$ hydrogen bonds is lower for the outer AA/UU steps, viz., $A_4A_5/U_{20}U_{21}$ and $U_8U_9/A_{16}A_{17}$ steps compared to their counterparts in A_3T_3, while, for the inner AA/UU steps, viz., $A_5A_6/U_{19}U_{20}$ and $U_7U_8/A_{17}A_{18}$, it is larger than the corresponding steps in A_3T_3. The unusual geometry of the CA/TG step compared to the CA/UG step might be a factor leading to these differences.

In the high resolution A_3T_3 crystal structure 1S2R, the distances between C_2 of adenine and O_2 of thymine are 3.5, 3.3, 3.5, and 3.4 Å for A_5 and T_{21}, A_6 and T_{20}, A_{18} and T_8, and A_{17} and T_9, respectively, while the distances between N_6 of adenine and O_4 of thymine on the major groove side are 3.6, 3.6, 3.5, and 3.7 Å for A_4 and T_{20}, A_5 and T_{19}, A_{16} and T_8, and A_{17} and T_7, respectively, indicating slightly more favorable $C_2-H_2\cdots O_2$ interactions. An exhaustive analysis of all B-DNA crystal structures[36] (including the structure 1D65), carried out after fixing hydrogen atoms at ideal geometric locations, also showed that the H···O distances and N−H···O angles in N−H···O contacts often do not satisfy the criteria for hydrogen bond geometry, while the corresponding distance and angle values for C−H···O bonds are observed to satisfy these criteria.

3.5. Influence of A-Tract on DNA Bending. The mean values of the angles between the local helix axes corresponding to the penultimate, i.e., the 2nd and 10th dinucleotide steps, which describes the end-to-end bending angle when the terminal basepairs are excluded, are shown in Figure 6. The mean values of the angles between the local helix axes corresponding to the 3rd and 9th dinucleotide steps, which shows the bending of the A-tract including the CA/U(T)G step, as well as the mean values of the angles between the local helix axes corresponding to the 4th and 8th dinucleotide steps, which is a measure of the bending for the core $A_3U(T)_3/A_3U(T)_3$ tract, are also shown in the same figure. The bending angle between the 4th and 8th dinucleotide steps is slightly but uniformly larger for A_3U_3 compared to A_3T_3 for all of the groove width ranges. However, the values themselves are small (14.0−17.2°), compared to the two flanking angles, indicating that the A-tract is relatively straight for both molecules. The largest bending angle values are observed between the 3rd and 9th dinucleotide steps. The bending is much smaller for A_3U_3 compared to A_3T_3 for all of the groove width ranges and seems to be correlated with the difference in the values of bending angles between the 3rd and 4th dinucleotide steps and the 8th and 9th dinucleotide steps in the two molecules.

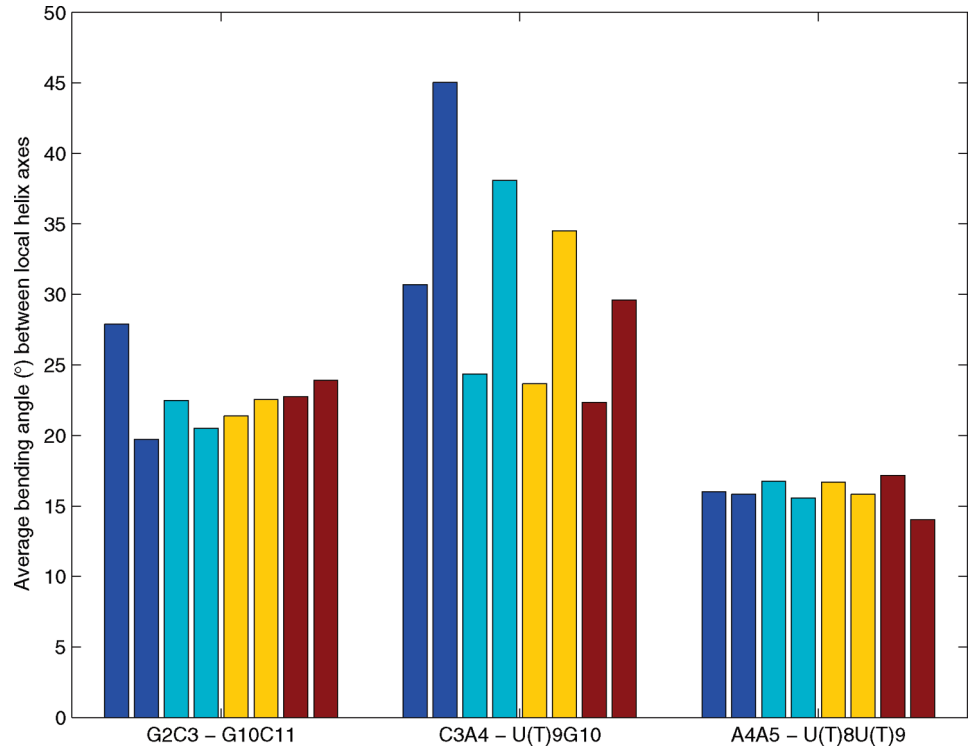

Figure 6. Histogram showing the mean values of the angles between the local helix axes vectors corresponding to dinucleotide steps $G_2C_3/G_{22}C_{23}$ and $G_{10}C_{11}/G_{14}C_{15}$, $C_3A_4/U(T)_{21}G_{22}$ and $U(T)_9G_{10}/C_{15}A_{16}$, and $A_4A_5/U(T)_{20}U(T)_{21}$ and $U(T)_8U(T)_9/A_{16}A_{17}$, for the four minor groove width ranges. The rest of the specifications are as detailed in the caption for Figure 2.

For all of the snapshots in the four defined minor groove width ranges, the angle between the 3rd and 4th dinucleotide steps in A_3U_3 is 17.6 ± 11.1°, while the corresponding value for A_3T_3 is considerably larger at 25.3 ± 14.0°. Similarly, the angle between the 8th and 9th dinucleotide steps in A_3U_3 is 17.7 ± 11.3°, while the corresponding value for A_3T_3 is 22.7 ± 12.9°.

The end-to-end bending angles between the 2nd and 10th dinucleotide steps are slightly larger for A_3U_3 compared to A_3T_3 for the two narrow minor groove width ranges but smaller for the next two wider groove width ranges. The GC/GC steps at both ends compensate to some extent for the large bending at the 3rd and 9th steps, by neutralizing the effect of the high roll angle at the CA/U(T)G steps, particularly for the A_3T_3 structure. This is evident from the proportions of snapshots with a significant difference in roll values between the two steps. The difference in roll values of the $C_3A_4/U_{21}G_{22}$ and $G_2C_3/G_{22}C_{23}$ steps and the $U_9G_{10}/C_{15}A_{16}$ and $G_{10}C_{11}/G_{14}C_{15}$ steps is >10° for 39.6 and 32.3%, respectively, of the total snapshots in the four selected groove width ranges. The corresponding numbers for A_3T_3 are larger at 56.5 and 58.2%, respectively.

We also compared the bending angle values obtained for the MD snapshots with the corresponding values obtained for the crystal structures 1S2R and 1D65. For 1S2R, the bending angles between the 4th and 8th, 3rd and 9th, and 2nd and 10th dinucleotide steps are 4.7, 23.7, and 9.6°, respectively, while the corresponding values for 1D65 are 12.8, 6.9, and 20.8°, respectively. Thus, the trend observed during MD for the bending angle values matches well with that observed in the high resolution structure 1S2R, while the values themselves are ∼10° larger. The constraints imposed by packing forces in crystals may be responsible for the smaller bending observed in the crystal structures.

4. Discussion

We observe that, when the MD snapshots of A_3U_3 and A_3T_3 are clustered on the basis of the minor groove width at the core of the A-tract region, they have different populations with varying minor groove width. The presence of the bulky 5-methyl group on the major groove side increases the frequency of a narrow minor groove conformation, implying that the minor groove is more likely to be narrow in the A-tract region for A_3T_3 as compared to A_3U_3. This observation, coupled with the recent report[3] that a narrow minor groove and the resultant strong negative electrostatic potential acts as a general readout signal for arginine residues, indicates that the presence of the 5-methyl group in thymines is critical for protein−DNA interactions. On the other hand, removal of the 5-methyl group has no effect on drug binding,[28] while it increases the susceptibility of the A-tract region in DNA to cleavage by DNase I and has been related to the increase in width of the A-tract minor groove, with a peak around a minor groove width of 12 Å.[60,61] Our observation that A_3U_3 structures take up a range of minor groove width values centered on 12 Å, with a frequency thrice that of A_3T_3, can explain why an A-tract region lacking the 5-methyl group is more susceptible to cleavage by DNase I. In addition, principal component analysis (PCA) indicates that the wider minor groove in A_3U_3 is a static structural feature and not a result of higher flexibility of that molecule. Finally, the average minor groove width of dodecamer crystal structures in A_3T_3 regions bound to a drug molecule is ∼10.2 ± 0.3 Å,[55] and both A_3U_3 and A_3T_3 have the maximum number of MD snapshots with minor groove width of the A-tract near 10.5 Å. This could be the reason why drug binding is insensitive to the presence of the thymine 5-methyl group.

It is also instructive to compare the change in proportion of BI/BII conformations caused by cytosine methylation with those observed in the A_3U_3 and A_3T_3 sequences. MD and NMR studies comparing a cAMP DNA responsive element containing a methylated cytosine with its unmethylated counterpart[42] showed that methylation did not change the proportion of BII conformation either for the step containing the methylated cytosine or

the guanine paired to it. In contrast, the adenines paired to thymines display a larger proportion of BII conformation compared to those paired to uracils. However, in the study mentioned above, a dinucleotide step positioned two basepairs upstream of the guanine paired to the methylated cytosine assumed a larger proportion of BII conformation compared to the unmethylated molecule, while its flanking step assumed a lower proportion of BII conformation. Our results for the A_3T_3 molecule containing the thymine 5-methyl group are very similar, where, with the exception of G_2, all other guanines belonging to the GC/GC steps, positioned at the 5′-end of the adenines paired to thymines, display larger proportions of BII conformation, compared to the corresponding numbers for A_3U_3. This is accompanied by a smaller proportion of BII for the neighboring CA/TG step as compared to that for the CA/UG step. This prevalence of the BII conformational state in the strand opposite to the one containing thymine may be significant, given the well-documented role of the BII conformational state in several biological processes.[40,42−47]

In the case of intra-basepair parameters, the presence of the thymine 5-methyl group is observed to impede buckling, while the propeller twist values show no significant trend for the differences between the A:U and A:T basepairs. This is in contrast to the prediction by Hunter[62] that A:T basepairs have a large, negative propeller twist to avoid clash between the thymine 5-methyl group and the neighboring sugar. Interestingly, the G:C basepairs in A_3T_3 structure have larger propeller twist as compared to those in A_3U_3. In the case of opening angle, the majority of the opening angle values for A:U and A:T basepairs are positive, indicating that opening into the major groove is preferred by A:T basepairs in A-tracts, in agreement with the prediction by Giudice and Lavery.[63] The outer A:U and A:T basepairs have larger opening angles as compared to the inner ones, as predicted by Moe et al.[64] The observation by Warmlander et al.,[27] using imino proton exchange measurements by NMR spectroscopy, that the basepair opening is damped due to the presence of the 5-methyl group, does not seem to always be true. The outer A:U/A:T basepairs in fact show an opposite trend, with the A:U basepair having smaller opening angle values than the A:T basepair. However, the four central A:T basepairs have significantly smaller opening angle values than the corresponding A:U basepairs, validating the observation[27] that the basepairs at the center of an A-tract display much larger damping of opening angle in the presence of the 5-methyl group. This can be explained by the minor groove being narrow in the core A_3T_3 region, which can hinder basepair opening.[63] In view of the coupling of damped opening in DNA A-tracts with a narrow minor groove and small bending,[63] our observations explain why the bending of the core A_3T_3 tract is smaller than the bending of the core A_3U_3 tract.

A comparison of the structural parameters of A_3T_3 molecule observed in our study with those observed for similar sequence motifs in DNA bound to proteins sheds further light on the biological relevance of the thymine 5-methyl group. A survey of the protein-bound DNA data sets used by Marathe et al.[32] reveals the presence of the hexanucleotide motif AAATTT/AAATTT, the pentanucleotide motif GCAAA/TTTGC, and the tetranucleotide motifs GCAA/TTGC and CAAA/TTTG in protein-bound DNA. The hexanucleotide fragment AAATTT/AAATTT occurs in DNA bound to the following proteins: MATα1/MATalpha2 homeodomain,[65] triple alanine mutant of MATalpha2,[66] and MATalpha2/MCM1 complex.[67] In all three cases, the A-tract motif assumes a narrow minor groove and most of the dinuceotide steps have small positive or negative roll, large twist, and nearly zero slide, as observed for the core A_3T_3 tract in this study. The bending angle values are 16.5, 9.3, and 19.0° for the fragments bound to the MATα1/MATalpha2 homeodomain, triple alanine mutant of MATalpha2, and MATalpha2/MCM1 complex, respectively, which are comparable to the bending observed in MD snapshots for A_3T_3.

The pentanucleotide motif GCAAA/TTTGC in the DNA bound to the transcription factor Oct-1[68] assumes a structure similar to the one observed for the same motif in this study. The CA/TG step assumes large roll, small twist, and negative slide, while the GG/CC step assumes small roll, large twist, and small positive slide. The GC/GC step and the AA/TT step at the 5′-end take up BI+BII backbone conformations. Narrowing of the minor groove is observed toward the 3′-end of the fragment.

The GCAA/TTGC and CAAA/TTTG tetranucleotide fragments occur in DNA bound to proteins from a variety of families such as transcription factors, histones, integration host factor, endonucleases, recombinases, homeodomains, etc. Certain common themes immediately emerge from a study of the structural parameters of these tetranucleotide motifs, as well as the pentanucleotide and hexanucleotide motifs described above. With the exception of the highly distorted DNA structures observed in a few protein−DNA complexes, the presence of two or more contiguous AA/TT steps ensures a narrow minor groove with small positive or small negative roll, large twist, and nearly zero slide for most of the AA/TT steps, very similar to the observation in this study. The A-tract regions display potential C−H···O hydrogen bond geometries, similar to those observed in this study. These observations, coupled with the comparative analysis of A_3U_3 and A_3T_3 in this study, highlight the importance of the thymine 5-methyl group in retaining the functionally important rigid and straight structure of the core A-tract, and a narrow minor groove. This may be critical in nucleosomes, where A-tracts are one of the most well-characterized positioning signals in regions where the minor groove faces the histone octamer, acting to bring the bendable DNA motifs in register with the histone octamer template, and allowing the histone residues to interact with the narrow minor groove region.[69]

Conclusions

Molecular dynamics simulation studies of an A-tract lacking the thymine 5-methyl group and a normal A-tract indicate that the thymine 5-methyl group is necessary for several characteristic features of an A-tract containing DNA structure. The clustering of MD snapshots based on the minor groove width of the core A-tract region demonstrates that the presence of a thymine 5-methyl group increases the frequency of a very narrow minor groove conformation, which in turn has been proposed to give rise to a strong negative electrostatic potential that can be recognized by arginine residues in proteins,[3] thus highlighting its essential biological role. Simultaneously, the presence of the thymine 5-methyl group decreases the frequency of a relatively wider minor groove with a width of ∼12 Å, which can explain why the thymine containing A-tract is less susceptible to cleavage by DNase I.[28,60,61] The bias toward a wider minor groove at the core of the A-tract lacking the thymine 5-methyl group is a static structural feature indicating that the absence of the 5-methyl group does not lead to an increase in flexibility. On the other hand, for both A_3U_3 and A_3T_3, the largest number of snapshots are observed to assume a minor groove with a width of ∼10.5 Å in the A-tract region, which can optimally accommodate minor groove binding drugs.

The thymine 5-methyl group is also observed to influence DNA structure at the level of backbone conformation as well as basepair and dinucleotide step parameters. The experimentally observed features of the core A-tract as well as that of the flanking CA/TG step and even the neighboring GC/GC step in free as well as protein-bound DNA are modulated by the presence of the thymine containing A-tract. This also influences the curvature of the longer DNA fragment in which the A-tract is embedded. Given that the noncoding regulatory regions in the genome are known to be AT-rich compared to their immediate neighbors in the downstream coding region,[70−72] the thymine 5-methyl groups may be influencing gene expression in a subtle and complex manner.

Acknowledgment. This work was partially supported by Department of Biotechnology, India.

Supporting Information Available: Description of the calculation and classification of backbone torsion angles and the calculation of basepair and dinucleotide step parameters and figure showing cartoon diagrams of representative snapshots of A3U3 and A3T3 for each minor groove width range. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) El Hassan, M. A.; Calladine, C. R. Two distinct modes of protein induced bending in DNA. *J. Mol. Biol.* **1998**, *282*, 331–343.

(2) Lane, W. J.; Darst, S. A. Computer modeling demonstrates that electrostatic attraction of nucleosomal DNA is mediated by histone tails. *PLoS Biol.* **2006**, *4*, e269.

(3) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248–1253.

(4) Brown, D. G.; Sanderson, M. R.; Skelly, J. V.; Jenkins, T. C.; Brown, T.; Garman, E.; Stuart, D. I.; Neidle, S. Crystal structure of a berenildodecamucleotide complex: the role of water in sequence specific ligand binding. *J. Mol. Biol.* **1990**, *11*, 281–287.

(5) Brown, D. G.; Sanderson, M. R.; Garman, E.; Neidle, S. Crystal structure of a berenil-d(CGCAAATTTGCG) complex as an example of drug-DNA recognition based on sequence-dependent structural features. *J. Mol. Biol.* **1992**, *226*, 481–490.

(6) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. Specific binding of Hoechst 33258 to the $d(CGCAAATTTGCG)_2$ duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.* **1997**, *271*, 244–257.

(7) Jenkins, T. C.; Lane, A. N.; Neidle, S.; Brown, D. G. NMR and molecular modelling studies of the interaction of berenil and pentamidine with $d(CGCAAATTTGCG)_2$. *Eur. J. Biochem.* **1993**, *213*, 1175–1184.

(8) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. Binding of an antitumor drug to DNA. Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. *Eur. J. Biochem.* **1985**, *213*, 1175–1184.

(9) Larsen, T.; Goodsell, D. S.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. The structure of DAPI bound to DNA. *J. Biomol. Struct. Dyn.* **1989**, *7*, 477–491.

(10) Laughton, C. A.; Tanious, F.; Nunn, C. M.; Boykin, D. W.; Wilson, W. D.; Neidle, S. A crystallographic and spectroscopic study of the complex between $d(CGCAAATTTGCG)_2$ and 2,5-bis(4-guanyl-phenyl)-furan, an analogue of berenil. Structural origins of enhanced DNA binding affinity. *Biochemistry* **1996**, *35*, 5655–5661.

(11) Quintana, J. R.; Lipanov, A. A.; Dickerson, R. E. Low temperature crystallographic analyses of the binding of Hoechst 33258 to the double-helical DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochemistry* **1991**, *30*, 10294–10306.

(12) Spink, N.; Brown, D. G.; Skelly, J. V.; Neidle, S. Sequence-dependent effects in drug-DNA interaction: the crystal structure of Hoechst 33258 bound to the $d(CGCAAATTTGCG)_2$ duplex. *Nucleic Acids Res.* **1994**, *22*, 1607–1612.

(13) Vega, C. M.; Garcia Saez, I.; Aymami, J.; Eritja, R.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Coll, M. Three-dimensional crystal structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor-groove binding drug Hoechst 33258. *Eur. J. Biochem.* **1994**, *222*, 721–726.

(14) Trifonov, E. N.; Sussman, J. L. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3816–3820.

(15) Levene, S. D.; Crothers, D. M. A computer graphics study of sequence directed bending in DNA. *J. Biomol. Struct. Dyn.* **1983**, *1*, 429–435.

(16) Wu, H. M.; Crothers, D. M. The locus of sequence directed and protein induced DNA bending. *Nature* **1984**, *308*, 509–513.

(17) Nagaich, A. K.; Bhattacharyya, D.; Brahmachari, S. K.; Bansal, M. CA/TG sequence at the 5′ end of oligo(A)-tracts strongly modulates DNA curvature. *J. Biol. Chem.* **1994**, *269*, 7824–7833.

(18) Satchwell, S. C.; Drew, H. R.; Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **1986**, *191*, 659–675.

(19) Lowman, H.; Bina, M. Correlation between dinucleotide periodicities and nucleosome positioning on mouse satellite DNA. *Biopolymers* **1990**, *30*, 861–876.

(20) Costanz, G.; Di Mauro, E.; Salina, G.; Negri, R. Attraction, phasing and neighbour effects of histone octamers on curved DNA. *J. Mol. Biol.* **1990**, *216*, 363–374.

(21) Widlund, H. R.; Cao, H.; Simonsson, S.; Magnusson, E.; Simonsson, T.; Nielsen, P. E.; Kahn, J. D.; Crothers, D. M.; Kubista, M. Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* **1997**, *267*, 807–817.

(22) Tolstorukov, M. Y.; Virnik, K. M.; Adhya, S.; Zhurkin, V. B. A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Res.* **2005**, *33*, 3907–3918.

(23) Segal, E.; Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **2009**, *19*, 65–71.

(24) Umemoto, K.; Sarma, M. H.; Gupta, G.; Sarma, R. H. Effect of the methyl group on DNA bending and curvature: structure of $d(GA_4U_4C)_2$ in solution. *Biochemistry* **1990**, *29*, 4714–4722.

(25) Diekmann, S.; Mazzarelli, J. M.; McLaughlin, L. W.; von Kitzing, E.; Travers, A. A. DNA curvature does not require bifurcated hydrogen bonds or pyrimidine methyl groups. *J. Mol. Biol.* **1992**, *225*, 729–738.

(26) Hagerman, P. J. Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. *Biochemistry* **1990**, *29*, 1980–1983.

(27) Warmlander, S.; Sponer, J. E.; Sponer, J.; Leijon, M. Three-dimensional crystal structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor-groove binding drug Hoechst 33258. *J. Biol. Chem.* **2002**, *277*, 28491–28497.

(28) Bailly, C.; Crowe, S.; Minnock, A.; Waring, M. J. Demethylation of thymine residues affects DNA cleavage by endonucleases but not sequence recognition by drugs. *J. Mol. Biol.* **1999**, *291*, 561–573.

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

(30) Coll, M.; Frederick, A.; Wang, A. H.; Rich, A. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385–8389.

(31) Bhattacharyya, D.; Bansal, M. Groove width and depth of B-DNA structures depend on local variation in slide. *J. Biomol. Struct. Dyn.* **1992**, *10*, 213–226.

(32) Marathe, A.; Karandur, D.; Bansal, M. Small local variations in B-form DNA lead to a large variety of global geometries which can accommodate most DNA-binding protein motifs. *BMC Struct. Biol.* **2009**, *9*, 24.

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

(34) Luisi, B.; Orozco, M.; Sponer, J.; Luque, F. J.; Shakked, Z. On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. *J. Mol. Biol.* **1998**, *279*, 1123–1136.

(35) Ghosh, A.; Bansal, M. Three-centre C-H—O hydrogen bonds in the DNA minor groove: analysis of oligonucleotide crystal structures. *Acta Crystallogr.* **1999**, *D55*, 2005–2012.

(36) Ghosh, A.; Bansal, M. C-H.O hydrogen bonds in minor groove of A-tracts in DNA double helices. *J. Mol. Biol.* **1999**, *294*, 1149–1158.

(37) Mandel-Gutfreund, Y.; Schueler, O.; Margalit, H. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J. Mol. Biol.* **1995**, *253*, 370–382.

(38) Lefebvre, A.; Mauffret, O.; el Antri, S.; Monnot, M.; Lescot, E.; Fermandjian, S. Sequence dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study. *Eur. J. Biochem.* **1995**, *229*, 445–454.

(39) Lefebvre, A.; Mauffret, O.; Lescot, E.; Hartmann, B.; Fermandjian, S. Solution structure of the CpG containing d(CTTCGAAG)2 oligonucleotide: NMR data and energy calculations are compatible with a BI/BII equilibrium at CpG. *Biochemistry* **1996**, *35*, 12560–12569.

(40) Shah, N.; Thomas, T.; Shirahata, A.; Sigal, L. H.; Thomas, T. J. Activation of nuclear factor kappaB by polyamines in breast cancer cells. *Biochemistry* **1999**, *38*, 14763–14764.

(41) Geahigan, K. B.; Meints, G. A.; Hatcher, M. E.; Orban, J.; Drobny, G. P. The dynamic impact of CpG methylation in DNA. *Biochemistry* **2000**, *39*, 4939–4946.

(42) Derreumaux, S.; Chaoui, M.; Tevanian, G.; Fermandjian, S. Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element. *Nucleic Acids Res.* **2001**, *29*, 2314–2326.

(43) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. Complex of B-DNA with polyamides freezes DNA backbone flexibility. *J. Am. Chem. Soc.* **2001**, *123*, 5044–5049.

(44) Banyay, M.; Graslund, A. Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR. *J. Mol. Biol.* **2002**, *324*, 667–676.

(45) van Dam, L.; Korolev, N.; Nordenskiold, L. Polyamine-nucleic acid interactions and the effects on structure in oriented DNA fibers. *Nucleic Acids Res.* **2002**, *30*, 419–428.

(46) Reddy, S. Y.; Obika, S.; Bruice, T. C. Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15475–15480.

(47) Djuranovic, D.; Oguey, C.; Hartmann, B. The role of DNA structure and dynamics in the recognition of bovine papillomavirus E2 protein target sequences. *J. Mol. Biol.* **2004**, *339*, 785–796.

(48) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W., Y. R.; Cieplak, P.; Luo, R.; Lee, T. A point charge force field for molecular mechanics simulations of proteins. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(49) Chandrasekaran, R.; Arnott, S. *Landolt-Bornstein: Numerical data and functional relationships in science and technology*; Springer-Verlag: Berlin, 1989; Vol. VII/1b.

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

(51) de Groot, B. L.; van Aalten, D. M.; Amadei, A.; Berendsen, H. J. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys. J.* **1996**, *71*, 1707–1713.

(52) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(53) Woods, K. K.; Maehigashi, T.; Howerton, S. B.; Sines, C. C.; Tannenbaum, S.; Williams, L. D. High resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]. *J. Am. Chem. Soc.* **2004**, *126*, 15330–15331.

(54) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S. Molecular structure of the B-DNA dodecamer d(CGCAAATTGCG)$_2$. An examination of propeller twist and minor-groove water structure at 2.2 Å resolution. *J. Mol. Biol.* **1992**, *226*, 1161–1173.

(55) Madhumalar, A.; Bansal, M. Structural insights into the effects of hydration and ions on A-tract DNA: a molecular dynamics study. *Biophys. J.* **2003**, *85*, 1805–1816.

(56) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. α-γA transitions in the B-DNA backbone. *Nucleic Acids Res.* **2002**, *30*, 5398–5406.

(57) Varnai, P.; Zakrzewska, K. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.

(58) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. Molecular dynamics studies of DNA A-tract structure and flexibility. *J. Mol. Biol.* **1999**, *267*, 595–623.

(59) McConnell, K. J.; Beveridge, D. Molecular dynamics simulations of B-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* **2001**, *314*, 23–40.

(60) Drew, H. R.; Travers, A. A. DNA structural variations in the E. coli tyrT promoter. *Cell* **1984**, *37*, 491–502.

(61) Brukner, I.; Jurukovski, V.; Savic, A. Sequence-dependent structural variations of DNA revealed by DNase I. *Nucleic Acids Res.* **1990**, *18*, 891–894.

(62) Hunter, C. A. Sequence-dependent DNA structure. the role of base stacking interactions. *J. Mol. Biol.* **1993**, *230*, 1025–1054.

(63) Giudice, E.; Lavery, R. Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc.* **2003**, *125*, 4998–4999.

(64) Moe, J. G.; Folta-Stogniew, E.; Russu, I. M. Energetics of base pair opening in a DNA dodecamer containing an A_3T_3 tract. *Nucleic Acids Res.* **1995**, *23*, 1984–1989.

(65) Li, T.; Jin, Y.; Vershon, A. K.; Wolberger, C. Crystal structure of the MATa1/MATalpha2 homeodomain heterodimer in complex with DNA containing an A-tract. *Nucleic Acids Res.* **1998**, *26*, 5707–5718.

(66) Ke, A.; Mathias, J. R.; Vershon, A. K.; Wolberger, C. Structural and thermodynamic characterization of the DNA binding properties of a triple alanine mutant of MATalpha2. *Structure* **2002**, *10*, 961–971.

(67) Tan, S.; Richmond, T. J. Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature* **1998**, *391*, 660–666.

(68) Remenyi, A.; Tomilin, A.; Pohl, E.; Lins, K.; Philippsen, A.; Reinbold, R.; Scholer, H. R.; Wilmanns, M. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol. Cell* **2001**, *8*, 569–580.

(69) Tolstorukov, M. Y.; Colasanti, A. V.; McCandish, M. M.; Olson, W. K.; Zhurkin, V. B. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J. Mol. Biol.* **2007**, *371*, 725–738.

(70) Kanhere, A.; Bansal, M. Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res.* **2005**, *33*, 3165–3175.

(71) Mitchison, G. The regional rule for bacterial base composition. *Trends Genet.* **2005**, *21*, 440–443.

(72) Abeel, T.; Saeys, Y.; Bonnet, E.; Rouze, P.; Van de Peer, Y. Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Res.* **2008**, *18*, 310–323.

JP911055X