Published in final edited form as:

J Phys Chem B. 2012 November 26; 116(46): 13672–13681. doi:10.1021/jp3084887.

Molecular Dynamics Study of the Role of the Spine of Hydration in DNA A-Tracts in Determining Nucleosome Occupancy

Xiao Zhu^{1,2} and George C. Schatz²

¹Texas Advanced Computing Center, Austin, TX 78758-4497

²Department of Chemistry, Northwestern University, Evanston IL 60208-3113

Abstract

A-tracts in DNA are generally associated with reduced nucleosome occupancy relative to other sequences, such that the longer the A-tract, the less likely that nucleosomes are found. In this paper we use molecular dynamics methods to study the structural properties of A-tracts, and in particular the role that the spine of hydration in A-tracts plays in allowing DNA to distort to the highly bent structure needed to form nucleosomes. This study includes a careful assessment of the ability of the Amber (parmbsc0), CHARMM27 and BMS force fields to describe these structural waters for the AAATTT sequence (here capped with CGC and GCG), including comparisons with X-ray results. All three force fields show a spine of hydration, but BMS and Amber show better correlation with measured properties, such as in narrowing of the minor groove width associated with the A-tract. We have used Amber to study the spine properties for several 6 and 14 base-pair A-tracts (all capped with CGC and GCG). These calculations show that the structural waters are tightly bound for “pure” A-tracts that allow for A-water-T links, and for AT steps that allow for a T-water-T link, but other sequences disfavor structural water, especially those that lead to A-water-A, G-water-G and C-water-A structures. In addition, we show that pure A-tracts favor *roll* values close to the Watson-Crick value for linear DNA, while A-tract sequences containing embedded T’s, C’s or G’s that are less favorable to structural water are more flexible. This implies the essential role of the spine of hydration in disfavoring nucleosome formation.

Keywords

DNA; nucleosome; A-tract; spine-of-hydration; nucleic acid; molecular dynamics

1. Introduction

An important advance in molecular biology in the past decade has been the determination of the dependence of nucleosome occupancy on DNA sequence.^{1,2} Nucleosomes are 4.1 nm radius cylindrical structures composed of a 147 base-pair (bp) sequence of DNA that is wrapped 1.67 turns about a protein complex known as the histone octamer. Nucleosomes are associated with 75–90% of the DNA in a cell nucleus, so understanding where along genomic DNA that nucleosomes are located, and where not, is important to understanding transcriptional regulation. Two trends that are gradually becoming clearer are (1) that nucleosomes prefer to occur for sequences where the dinucleotides AA,TT, AT or TA occur

Correspondence to: George C. Schatz.

Supporting Information Available: A detailed analysis of the sequence dependent structural propensities of the AAATTT sequence is provided, including Figures S1-S8. Figure S9 shows the minor groove width as a function of position for the 6 base-pair A-tracts. Figure S10 plots water diffusion coefficient for the same A-tracts. This material is available free of charge via the Internet at <http://pubs.acs.org>.

every 10 base pairs,³ and (2) that A-tracts or poly-dA/dT tracts (i.e., long strings of A's) tend to disfavor nucleosomes as shown in various studies,^{2,4–11} with the degree of disfavoring being correlated with the length of the A-tract. The former issue has been examined using simple models^{12,13} and molecular dynamics,¹⁴ demonstrating that the mechanical properties of DNA, and in particular its ability to bend to the 4.1 nm radius of curvature of the nucleosome, is strongly favored by the dinucleotide repeats. These studies also have demonstrated that the DNA minor groove is narrowed where these repeats are located, corresponding to the minor groove pointing towards the histone complex every 10 bp. Many analyses of sequences enriched in nucleosome-occupied and nucleosome-depleted regions in genome-wide data sets have demonstrated the importance of nucleosome-excluding sequences, in particular A-tracts.^{6,15–17} Recently, Tillo and Hughes were able to fit the same data sets as in Ref 12 with a similar but much simpler model that was heavily dependent on G+C and poly-dA/dT content.¹⁸ In another interesting study, Locke et al. found that the nucleosome occupancies can be rationalized by the differences in mono- and dinucleotide content between nucleosomal and linker sequences, with A:T dinucleotides being depleted in nucleosomes and enriched in linkers.¹⁹

In this paper we address the effect of A-tracts on nucleosome occupancy using molecular dynamics calculations. A-tracts are well known to have unique structural properties, including the formation of bent duplex structures.^{20–22} A property that has long been noted is their tendency to have structural water molecules in the minor-groove, leading to what is often called the spine of hydration in DNA.²³ This was characterized in some depth using X-ray measurements for the sequence GCGCAATTCGCG (i.e., an AATT A-tract that is capped by GCGC and its complement)²³, and more recently other A-tracts have been studied.²⁴ In the spine of hydration, water molecules are able to form hydrogen bonds that link a nucleotide X on the 3'–5' strand to the nucleotide Y of the next bp on the 5'–3' strand. We will denote such a structure using the notation X-water-Y, and based on the schematic in Fig. 1 one might expect that for the sequence d(A)_n there can be n-1 T-water-A links and therefore n-1 structural waters. A-tracts are often associated with groove narrowing and slight bending of short DNA duplexes, although the details depend on precise details of the A-tract. However it is intuitive to assume, although sometime debated,²⁵ that the additional stability associated with cross-strand hydrogen bonds is responsible for the unusual properties of A-tracts.⁴ Thus if it is harder to bend an A-tract, then the formation of the bent structures of DNA that are needed for nucleosome binding will not occur. However the work to date on this topic has not provided quantitative modeling, so the role of the spine of hydration in nucleosome occupancy is still an open question.

A-tracts have long been of interest to the modeling community, as the unusual properties of A-tracts together with the availability of high quality crystal structures for short duplexes, make these an easily accessible target. Indeed some of the early MD studies of DNA were concerned with A-tracts, as there was uncertainty over the importance of counterions (such as Na⁺) in the minor groove.^{26–30} An important issue in both this early DNA work, and even today, is the reliability of the force fields being used.³¹ These force fields have evolved significantly over the past decade, but their ability to describe the spine of hydration (where the water force field is also important) has not been carefully studied. Thus in the first section of this paper, we present a detailed comparison of the water spine for three force fields (all of which have been optimized for DNA studies): Amber (parm99³² with parmbsc0 nucleotide modifications³³), CHARMM27³⁴ and Bristol, Myers, Squib (BMS)³⁵. The water model in all cases is chosen to be TIP3P³⁶ and all calculations include explicit solvent and ions. We do this comparison for the sequence CGCAAATTTGCG, which has been of significant interest as a model system for studies of protein-DNA and drug-DNA interactions.^{37–42} This sequence has also been the subject of prior MD and other theory studies,^{38,43,44} and a high resolution structure (1S2R) is available from Woods et al,²⁴

making it a good system for calibration studies. A conclusion of the calibration studies which we present is that although structural waters are found in all force fields, Amber (parm99) and BMS are to be preferred compared to CHARMM27 for describing the water spine, as the number of structural waters and their location in the minor groove is reasonably consistent with experiment.

With this conclusion, we provide a systematic study of many A-tracts using the Amber (parmbc0) force field, in particular characterizing the variation of the spine of hydration with base pair sequence. These studies include four A-tracts with 6 bps: A₃T₃, A₆, A₂T₂A₂, (AT)₃ (plus the CGC and GCG caps), and four with 14 bps: A₁₄, A₄TA₄GA₄, A₃T₂A₃CGA₄, A₃T₂AGACGATA₂ (plus caps) that were the subject of nucleosome binding affinity experiments.^{1,5,6} The goal of this part of the paper is to determine the rules which govern the presence of structural waters, and to determine how these waters influence the structures and stability of the A-tracts. From this, the factors which influence the ability of different A-tracts to form nucleosomes will become apparent.

2. Calculations

The high resolution dodecamer structure 1S2R having the CGCAAATTTGCG sequence was picked for validation of the force field and simulation protocol, here focusing on the central A-tract. Scheme 1 shows the structure and the numbering convention.

We compared molecular dynamics results for the A₃T₃ sequence using three different force fields: (1) Amber parm99³² with parmbc0 nucleotide modifications (parmbc0)³³, (2) CHARMM27³⁴ and (3) Bristol-Myers Squibb (BMS)³⁵. To build the molecular system, the high-resolution crystal structure (Protein Data Bank code: 1S2R) including crystallographic water was solvated in a cubic box of TIP3P water.¹⁶ Each side of the box extended 10 Å away from any solute atom and water molecules within a distance of 2.6 Å of any solute atom were deleted. Sodium ions were randomly added to the water box in each case to neutralize the net charge. Calculations for the other 6 bp sequences A₆, A₂T₂A₂, (AT)₃ followed the same procedures except using ideal B-DNA starting structures with parmbc0 only. All trajectories were extended to 150 ns and carried out with the CHARMM program.^{45,46} The SHAKE⁴⁷ algorithm was used for all bonds involving hydrogen to allow a time step of 2 fs. A shift scheme⁷ was employed for van der Waals interactions with a cut off distance of 9 Å and long range electrostatics were treated with the Particle-Mesh Ewald (PME) method.^{8,9} A convergence parameter of 0.33 Å⁻¹ and a sixth degree B-spline interpolation was employed with the PME method. In the constant temperature and pressure simulations, the temperature was constrained using the Nose-Hoover algorithm^{10,11} with a mass of 250 amu for the thermostat and a Hoover reference temperature equal to the desired system temperature (298K). The pressure was controlled with the Langevin piston method¹⁸, with a mass of 500 amu for the pressure piston, a reference pressure of 1 atm, a Langevin piston collision frequency of 10 ps⁻¹ and a Langevin piston bath temperature equal to the desired system temperature. In the case of the AMBER force field, explicit scaling of the 1–4 electrostatic function was applied, as recommended for nucleic acid simulations. In addition, calculations for the 14 bp sequences A₁₄, A₄TA₄GA₄, A₃T₂A₃CGA₄, and A₃T₂AGACGATA₂ were done using 15 ns simulations at 298K, including explicit solvent and ions and parmbc0. Water molecules more than 15 Å from any solute atom were removed and a 14 Å cut-off distance was employed in real space; other simulation details were kept from the dodecamer simulations.

3. Global structure: Force Field Validation

In all DNA simulations, end-fraying is a well-known issue, so in the following comparisons we will focus on the part of the sequence in Scheme 1 that excludes the two outermost GC

base pairs. We find that the inner base pairs keep a Watson-Crick (WC) arrangement during the simulation time; with a stretch parameter that is well inside regular fluctuations (usually less than 0.6 Å) for each base pair. Distributions of RMSDs of the non-hydrogen atoms of these base pairs with respect to crystal structure are shown in Figure 2. This shows that the BMS force field has the best agreement with the X-ray structure, with an average RMSD of ~1.6 Å. CHARMM27 and AMBER give MD trajectories that deviate more or less equally from the crystal structure; the average deviation is 2.4 Å for AMBER and 2.5 Å for CHARMM27. In addition, the RMSD using the BMS force field is about half of that of the AMBER and CHARMM27 simulations.

Conformational changes in the DNA backbone involve a complex interplay of several torsional degrees of freedom. A simple way to investigate the average effect of these motions is to measure the interstrand phosphate distance, which defines both minor and major groove width.⁴⁸ In the high resolution A₃T₃ structure, the narrow minor groove is associated with the spine of hydration, which could further correlate with the deformability of DNA.⁴⁹ Figure 3 plots the minor and major groove widths at positions 4–9 including comparisons with X-ray structure results. None of the force fields can reproduce the detailed trends of the interstrand phosphate distances in the crystal structure quantitatively. Both AMBER and BMS show a symmetrical Ushape minor groove profile and close to 3 Å narrowing in the center. The CHARMM27 force field leads to larger minor groove widths and less narrowing of the A-T base pair region; similar behavior was observed for the Dickerson dodecamer.⁵⁰

Concerning the major groove widths, CHARMM27 systematically underestimates the major groove width by 2–3 Å. The anti-correlation of minor and major groove widths suggests that the CHARMM27 conformations tend to bend toward the major groove and widen the minor groove. Similar to what is found for the minor groove widths, AMBER produces greater interphosphate distances than BMS. The crystal structure distances in general lie between the BMS and AMBER results. Concerted widening/narrowing of the major and minor groove could be realized by unwinding/winding the double helix⁵¹ as is further related to the twist angle.

3.1. Sequence dependence of structural propensities

We have performed detailed studies of the sequence dependence of the structural propensities. The results are described in the Supporting Information (SI), including base pair geometries in Figs. S1, S2, S3, bending dials in Figs. S4, and backbone parameters in Figs. S5, S6 and S7. Also, Table S1 presents the sugar pucker parameters. Where comparisons with experiment are possible, these results underline the conclusions that BMS and AMBER tend to give more quantitative agreement than CHARMM27 while non-canonical backbone conformations are more frequently sampled with BMS.

3.2. Flexibility

In order to probe the influence of force field on nucleic acid backbone flexibility, we calculated the root mean squared fluctuations (RMSF) values for non-hydrogen backbone atoms (including sugar and phosphate groups) of each nucleotide. Because the sequence is palindromic, RMSFs of both strands are almost symmetrical; figure 4 shows the RMSFs for the main strand for clarity. The qualitative features from all force fields are similar, in which the central TT dimer is the least flexible. Overall, the backbone in BMS and CHARMM27 is more rigid than in parmbsc0.

Helical stiffness parameters at the base pair level allows us to understand sequence dependent flexibility in helical space^{52–54} and can be used for mesoscopic simulations of

very long pieces of DNA. The CHARMM27 and BMS force constants are in general higher than those in parmbsc0 in average by 10–15%. The relative rigidity order of translational movement can be captured by all three force fields while the relation among rotational stiffness is not clearly defined. BMS and parmbsc0 tend to present the order of *tilt twist roll*. Figure S8 (in Supporting Information) shows a comparison of rotational rigidity. It is clear that the dependence of stiffness parameters on sequence is beyond the dinucleotide level.⁵⁵ The AT step is the least flexible of all rotational deformations in parmbsc0 while it is less rigid by modifying the tilt angle in CHARMM27 and BMS.

3.3 Hydration and water dynamics

Solvent plays a crucial role in DNA structure, stability and dynamics. We calculated the hydration numbers in the first solvation shell of the duplex grooves and the phosphate group of the central duplex and A₃T₃ regions. Please noted that the first solvation shell water molecules are identified as those within 3.0 Å of DNA atoms belonging in each region (backbone, minor and major groove). The results are summarized in Table 1. Parmbsc0 solvates the phosphodiester backbone oxygens (O3', O5', O1P, O2P) best followed by BMS and CHARMM27. On average, somewhat less than two fewer water molecules are found in CHARMM27 than parmbsc0. Groove solvation is more complicated for the different force fields as both groove width and depth could contribute to the specific hydration number. Not surprisingly, there is always the highest hydration number around the minor groove and the fewest water molecules around the major groove in CHARMM27. DNA in BMS has a narrow and deep minor groove while the one in parmbsc0 is shallower (by ~ 0.7 Å) and slightly wider. This leads to somewhat less than one water molecule located in the minor groove in BMS than in parmbsc0. Major groove hydration is higher with parmbsc0 in the A₃T₃ region and lower for the whole sequence than BMS.

In order to further probe the water structure in the duplex minor groove at atomic detail, we analyzed the hydrogen bond pattern in the simulations. In the crystal structure, a water molecule in the first layer of the spine of hydration bridges the N3/O2 atoms of a particular nucleotide with the nucleotide one base pair upstream on the complementary strand. This sequence dependence of the one-water bridge percentage is believed to correlate with DNA deformability.⁴⁹ The fraction and lifetime of the one-water bridge are calculated and plotted in Figure 5. In these results, a one-water bridge is considered to be present when the water forms hydrogen bonds with both N3/O2 from each strand of the duplex. The criterion for hydrogen bond formation is defined such that the maximum distance between hydrogen bond acceptor (A) and hydrogen bond donor (D-H) is 2.4 Å and the minimum D-H...A angle is 125°. The one-water bridge fraction is defined as the ratio of the number of frames where the interstrand bridge is present to the total number of frames. The bridging water molecules undergo exchange with other water molecules from the bulk solvent so the lifetime reflects the average residence time of a specific water molecule in such a bridge. Both parmbsc0 and BMS present similar profiles in Fig. 5, with high occupancy (~50%) and long lifetime (~20 ps) in the central A-tracts. A small difference can be identified between them: the parmbsc0 occupancy appears to be higher at all positions except T7. On the contrary, the percentage of the one-water bridge in CHARMM27 is generally quite low along the sequence, less than 8% even in the A₃T₃ region. The distinctions can be again explained by the minor groove widths in the three force fields; one water molecule cannot form hydrogen bonds simultaneously with the N3/O2 atoms on both strands for the wider minor groove in CHARMM27. In Figure 6, we compare the water hydration pattern in the minor groove with the spine of hydration seen in the X-ray results. Not surprisingly, parmbsc0 and BMS are able to capture the location of first layer crystal water very well. Two water molecules seem to be necessary in CHARMM27 to form an effective bridge. While 20 ps lifetimes for water in the minor groove cannot be considered long-lived for

most measurements, the high occupancy suggests that the bridging sites are only briefly vacant. Also, these MD results for DNA in solution are not directly comparable to the conditions of the X-ray measurements so lifetimes are not expected to be comparable.

3.4. Summary of validation results

Both parmbsc0 and BMS are able to capture the spine of hydration very well while CHARMM27 tends to present a different solvation pattern than is determined by the X-ray results. In general BMS predicts the best overall structure compared to crystallographic data while parmbsc0 does better in modeling the backbone torsions; BMS tends to favor non-canonical backbone configurations only seen in protein-DNA complexes (please see SI for details). Concerning flexibility, all three force fields qualitatively agree with each other and CHARMM27 and BMS in general produce more rigid structures. In the following we only consider parmbsc0 given its good performance in describing all important degrees of freedom.

4. Studies of 6 bp A-tracts

We now extend the parmbsc0 MD studies to consider the A-tracts A_3T_3 (same as above), A_6 , $A_2T_2A_2$, and $(AT)_3$. Scheme 2 shows these A-tracts as well as possible water bridges that would form X-water-Y links. Of course, the results for A_3T_3 presented above have already indicated that there is only an imperfect correlation between calculated water occupancy and the idealized picture in Scheme 2. Figure 7 (left panel) shows how this result generalizes (for parmbsc0) to other sequences. What this shows is that for A_6 there is a relatively flat distribution of waters with a probability of 0.4–0.5 for the waters labeled 6–10 in Scheme 2. This contrasts with the results for A_3T_3 , which show a higher probability, 0.8, for the T-water-T link numbered 7 in Scheme 2. These results are further amplified in the $A_2T_2A_2$ results, where Fig. 7 shows a peak for the T-water-T link labeled 5 in Scheme 2, and then a minimum for the A-water-A link labeled 8. For the $(AT)_3$ structure in Fig. 7, we see peaks for each T-water-T link, and minima for each A-water-A link. We conclude from this that water bridges between two thymines are especially stable, those between two adenines are not very stable, and water between thymine and adenine are stable but not as stable as T-water-T links. The symmetry of the A_3T_3 results about T7 in Scheme 1 suggests that A-water-T is comparable in stability to T-water-A. Fig. 8 shows the water densities for A_3T_3 , A_6 , and $A_2T_2A_2$. Here we see that A_6 shows the expected density maximum for water linking each A with a T, while there is extra probability for a water between two T's and a very much reduced probability for water between two A's. Unfortunately we do not have X-ray data to confirm this prediction other than for A_3T_3 , and in this case it is not clear that there is extra probability for water to link the two T's. The third panel of Fig. 7 shows the parameter V_{step} which we have calculated using the MD results to provide a measure of bp flexibility. Here we have followed the work of Yonetani and Kono,⁴⁹ who defined V_{step} as the square root of the product of covariance matrix eigenvalues, with the 6×6 covariance matrix determined using the variables shift, slide, rise, roll, tilt, twist associated with adjacent base-pairs. As expected Fig. 7 shows low values for A-water-T and T-water-T links, indicating that these structures are the least flexible, while V_{step} is high for A-water-A steps, showing how unstable are these structures.

In Fig. S9, we plot the minor groove width (mgw) as a function of bp position, similar to the top panel of Fig. 3 for the A_3T_3 sequence. This shows behavior that is consistent with the water occupancy results, with the most narrowed mgw being associated with A_3T_3 and A_6 , with the minimum for A_3T_3 being roughly in the middle of the A-tract while that for A_6 is shifted to the right. The least amount of narrowing is associated with $(AT)_3$, while $A_2T_2A_2$ shows a minimum to the left and a maximum to the right of position 6.

Fig. 7 (middle panel) presents lifetime results for the same sequences, and we see that the trends are largely consistent with the left panel, with the longest lifetimes correlated with the highest one-bridge fractions. Note that even the most stable waters have average lifetimes of about 20 ps, which means that there is still significant water exchange between minor groove and bulk even though the motion of minor groove water is greatly quenched. Water diffusion coefficients, generated from mean-square displacements according to the Einstein relation (see below), and plotted in Fig. S10, show values in the range $0.6\text{--}0.8 \times 10^{-5} \text{ cm}^2/\text{s}$ with slightly larger values for $(\text{AT})_3$ and $\text{A}_2\text{T}_2\text{A}_2$ than for the other two sequences. The one-order of magnitude retardation in water dynamics compared to bulk water ($D_{\text{water}} \sim 5.0 \times 10^{-5} \text{ cm}^2/\text{s}$) is likely caused by the water bridge between DNA strands and by nanoconfinement of non-bridging water in the minor groove.⁵⁶ Note that the translational diffusion coefficients were calculated for water molecules in the first solvation shell of the minor groove, which was determined every 10 ps. The distance that each water molecule traveled was monitored for the subsequent 10 ps, and then the diffusion coefficient was obtained from the slope of the linear range of mean-square displacement curve.

5. Studies of 14 bp A-tracts

Now we consider MD results for the 14 bp sequences A_{14} , $\text{A}_4\text{TA}_4\text{GA}_4$, $\text{A}_3\text{T}_2\text{A}_3\text{CGA}_4$, and $\text{A}_3\text{T}_2\text{AGACGATA}_2$. These sequences are presented in Scheme 3, along with the abbreviations PAT_n , $n=0,2,4,6$ that we use in our discussion of the results, where n indicates the number of bases on the 5'–3' strand that are different from A (choices that match work by Segal and Widom^{1,5,6}). To simplify Scheme 3 we have omitted the linking waters and only the 5'–3' strand is presented.

Figure 9 presents the water occupancies for the 14 bp A-tracts. This shows that the A_{14} sequence labeled PAT_0 has a nearly constant water occupancy with a value close to 0.5 except for the first two A's on the 5' end. This is very much the same behavior as we found for the A_6 results in Fig. 7 (left panel). PAT_2 shows two distinct dips in water occupancy, with the more notable dip corresponding to water bridging between A (on the complementary strand in position 7) and A in positions 8–9, and the second dip to a C-water-A water bridge at positions 13–14. PAT_4 shows more pronounced dips, especially in positions 12–13 and 13–14, where the water bridges are G-water-G and C-water-A. PAT_6 shows even more serious dips, so if overall water occupancy correlates to DNA bending flexibility (as we considered in the 6-bp study and V_{step} plotted in right panel of Fig. 9), and then to nucleosome binding, then one would expect that nucleosome binding would increase as one goes from PAT_0 to PAT_6 , which is what was observed in the experimental studies of these sequences.^{1,5,6}

To further study the 14 bp A-tracts, in Fig. 10 we present results for the *mgw* and for the *roll* and *slide* coordinate as a function of bp position for $\text{PAT}_0\text{--PAT}_6$. We found in our previous study that *roll* was significantly deviated from the WC value in the nucleosome structure.¹⁴ Also a novel roll-and-slide mechanism was proposed by Tolstorukov and his co-workers for DNA folding in nucleosome.⁵⁷ The *mgw* results are largely what would be expected based on our earlier results for the 6 bp structures, with *mgw* being almost constant for PAT_0 in its A-tract region, with a value of around 11 Å. This is smaller than the WC *mgw*. The other A-tracts show oscillations in *mgw*, with a distinct peak in the $\text{PAT}_2\text{--PAT}_6$ results at positions 8–9 where the water occupancy is low due to the weak water binding at the A-water-A link. Also note that the *mgw* is generally larger for $\text{PAT}_2\text{--PAT}_6$ than for PAT_0 , an indication of the correlation between stability of the water spine and *mgw*.

The *roll* and *slide* values in Fig. 10 provide new insights concerning the effect of the water spine on DNA structure. This figure shows that for PAT_0 , *roll* and *slide* are nearly flat

functions of position, with both values close to the WC value (0°). The other A-tracts show peaks at about the same positions as the peaks in the mgw (Fig. 10), and the dips in water occupancy (Fig. 9). To provide further insight, we have plotted the stiffness parameter for roll and slide deformations in Fig. 11. Both curves have similar sequence-dependence, with TA, CG, GA steps more flexible against DNA bending, and are anti-correlated with the magnitude of water binding at the interstrand water link. More base steps appear as hinge points when more non-deoxyadenosine nucleotides are present in the sequences. These results for the *roll* and *slide* coordinates reveal the ability of A-tracts with impurities to distort from the WC structure, which is essential for DNA bending in nucleosomes. The unique role of lateral displacements of the adjacent base pairs in nucleosome formation was recently emphasized.¹⁹ Large positive *slide* and negative *roll* occur at locations where the DNA bends into the minor groove. Fig. 10 shows that PAT0 greatly favors the WC structure, and therefore is unlikely to bend in either direction. In other words, inter base-pair motion is limited by the water bridge in a chain of homo adenines. Combining the 6 and 14 basepair results, we can infer that nucleosome depletion in A-tracts is likely length dependent since it is mediated by the length of the spine of hydration. The other sequences show more flexibility, but typically giving positive *roll* and negative *slide* values. Only positions 6–7 in PAT2–PAT6 show the negative *roll* values that might allow DNA to bend into a nucleosome configuration. However the negative values we find are much smaller in magnitude (-3° vs -10°) than are found for nucleosome structures.¹⁴ Of course the results in Fig. 10 are for an average of 3000 DNA structures, so fluctuations about the average would be needed to make the nucleosome structure accessible. Please also note that we only simulate free DNA and do not consider any important protein-DNA interactions. Studies involving DNA-protein complex are beyond the scope of this paper.

6. Conclusion

We have performed a molecular dynamics study of A-tract structures with emphasis on characterizing the spine of hydration, and its dependence on base-pair sequence, and on the connection of this behavior with the bending of DNA that leads to nucleosome formation. The base-pair sequences studied included the 6 bp structures A_3T_3 , A_6 , $A_2T_2A_2$, $(AT)_3$ (plus CGC and GCG caps), and the analogous 14 bp structures A_{14} , $A_4TA_4GA_4$, $A_3T_2A_3CGA_4$, $A_3T_2AGACGATA_2$ (plus caps). The 14 bp structures were taken from work of Segal and Widom where it was found that nucleosome occupancy is connected both to the length of the poly-A sequence and to its purity (fraction of tract that is A).

As an initial step in this project, we validated the ability of three popular force fields to describe the water of hydration in an A-tract through comparison with X-ray results for the A_3T_3 sequence. While all three force fields gave structural water, there was considerable variation in the stability of these structures, with BMS and Amber generally being better than CHARMM. Since Amber is better treating the backbone degrees of freedom and is a popular force field for simulating nucleic acid properties, we chose this for the rest of our simulations. Our conclusions about CHARMM27 are in contrast to earlier studies of base-pair flipping in DNA,²² where it was found that this force field was preferred over Amber and BMS. More recent versions of CHARMM have been developed,⁵⁸ however these refer to increased sampling of BII configurations (defined in Supporting Information) that, based on the results in Fig. S6, do not play an important role in our work.

Our studies of the spine of hydration for the 6 bp structures demonstrated a number of distinct trends, most notably that the stability of the structural water is highest for T-water-T, a little lower for A-water-T and T-water-A and much lower for A-water-A. These rules not only explain the water occupancy results and lifetimes, but also the inter-basepair flexibility. The same trends also describe the 14 bp results, and in a few cases we found other trends,

such as a lack of stability of C-water-A and G-water-G. In addition, we found trends in the variation of the minor groove width with bp sequence that are consistent with water stability and are in agreement with X-ray results.

We found that these trends with respect to structural water also connect to the *roll* and *slide* variable, where we found that A-tracts resist having the negative *roll* and positive *slide* values that are needed to form nucleosomes (where the minor groove narrows at the point where the DNA bends, such that the minor groove points towards the histone complex). More importantly, we found that the bending stability is highest for the poly-A sequence, and becomes less noticeable as one inserts other bases (T, G, C) into this sequence. Thus the MD calculations support the assertion that the resistance of A-tracts to nucleosome formation arises from the spine of hydration as the spine favors structures that are closer to the WC structure, and makes these structures more rigid than is found for random sequences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the PS-OC Center of the NIH/NCI (grant 1U54CA143869-01) and was also supported in part through the computational resources and staff contributions provided by Information Technology at Northwestern University as part of its shared cluster program, Quest. Jonathan Widom played a key role in getting us interested in doing this project.

References

1. Segal E, Widom J. Trends in Genetics. 2009; 25:335–343. [PubMed: 19596482]
2. Raveh-Sadka T, Levo M, Shabi U, Shany B, Keren L, Lotan-Pompan M, Zeevi D, Sharon E, Weinberger A, Segal E. Nature Genetics. 2012; 44 743-U163.
3. Segal E, Fondufe-Mittendorf Y, Chen LY, Thastrom A, Field Y, Moore IK, Wang JPZ, Widom J. Nature. 2006; 442:772–778. [PubMed: 16862119]
4. Anderson JD, Widom J. Molec. Cell. Biol. 2001; 21:3830–3839. [PubMed: 11340174]
5. Segal E, Widom J. Curr. Opin. Struct. Biol. 2009; 19:65–71. [PubMed: 19208466]
6. MacKerell AD. J. Comp. Chem. 2004; 25:1584–1604. [PubMed: 15264253]
7. Yuan GC, Liu JS. Plos Comput Biol. 2008; 4:164–174.
8. Sekinger EA, Moqtaderi Z, Struhl K. Mol Cell. 2005; 18:735–748. [PubMed: 15949447]
9. Peckham HE, Thurman RE, Fu YT, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng ZP. Genome Res. 2007; 17:1170–1177. [PubMed: 17620451]
10. Iyer V, Struhl K. Embo J. 1995; 14:2570–2579. [PubMed: 7781610]
11. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. Nature. 2009; 458 362-U129.
12. Battistini F, Hunter CA, Moore IK, Widom J. J. Mol. Biol. 2012; 420:8–16. [PubMed: 22472421]
13. Morozov AV, Fortney K, Gaykalova DA, Studitsky VM, Widom J, Siggia ED. Nucl. Acids Res. 2009; 37:4707–4722. [PubMed: 19509309]
14. Prytkova T, Zhu X, Widom J, Schatz GC. J. Phys. Chem. B. 2011; 115:8638–8644. [PubMed: 21639136]
15. Furse KE, Corcelli SA. J. Phys. Chem. B. 2010; 114:9934–9945. [PubMed: 20614919]
16. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J. Chem. Phys. 1983; 79:926–935.
17. Ryckaert JP, Ciccotti G, Berendsen HJC. J. Comp. Chem. 1977; 23:327–341.
18. Tillo D, Hughes TR. BMC Bioinformatics. 2009; 10:442. [PubMed: 20028554]

19. Locke G, Tolkunov D, Moqtaderi Z, Struhl K, Morozov AV. *Proc. Natl. Acad. Sci. US.A.* 2010; 107:20998–21003.
20. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD. *J. Chem. Theory Comp.* 8:348–362.
21. Sanghani SR, Zakrzewska K, Harvey SC, Lavery R. *Nucl. Acids Res.* 1996; 24:1632–1637. [PubMed: 8649979]
22. Koo HS, Crothers DM. *Biochemistry.* 1987; 26:3745–3748. [PubMed: 3651409]
23. Drew HR, Dickerson RE. *J. Mol. Biol.* 1981; 151:535–556. [PubMed: 7338904]
24. Woods KK, Maehigashi T, Howerton SB, Sines CC, Tannenbaum S, Williams LD. *J. Am.Chem. Soc.* 2004; 126:15330–15331. [PubMed: 15563130]
25. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. *PLoS Comput. Biol.* 2008; 4:e1000216. [PubMed: 18989395]
26. Young MA, Jayaram B, Beveridge DL. *J. Am.Chem. Soc.* 1997; 119:59–69.
27. Hamelberg D, McFail-Isom L, Williams LD, Wilson WD. *J. Am.Chem. Soc.* 2000; 122:10513–10520.
28. McConnell KJ, Beveridge DL. *J. Mol. Biol.* 2000; 304:803–820. [PubMed: 11124028]
29. Madhumalar A, Bansal M. *Biophys. J.* 2003; 85:1805–1816. [PubMed: 12944294]
30. Ababneh AM, Ababneh ZQ, Large CC. *J. Theo. Biol.* 2008; 252:742–749.
31. Petersen HG. *J. Chem. Phys.* 1995; 103:3668–3679.
32. Wang JM, Cieplak P, Kollman PA. *J. Comp. Chem.* 2000; 21:1049–1074.
33. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. *Biophys. J.* 2007; 92:3817–3829. [PubMed: 17351000]
34. Foloppe N, MacKerell AD. *J. Comp. Chem.* 2000; 21:86–104.
35. Langley DR. *J. Biomolec. Str. & Dyn.* 1998; 16:487–509.
36. Nose S. *J. Chem. Phys.* 1984; 81:511–519.
37. Sanchez MI, Vazquez O, Vazquez ME, Mascarenas JL. *Chem. Comm.* 2011; 47:11107–11109. [PubMed: 21909542]
38. Andac CA, Miandji AM, Hornemann U, Noyanalpan N. *Int. J. Biological Macromolecules.* 2011; 48:531–539.
39. Banerjee D, Makhil A, Pal SK. *J. Fluorescence.* 2009; 19:1111–1118.
40. Majid A, Smythe G, Denny WA, Wakelin LPG. *Chem. Res. Tox.* 2009; 22:146–157.
41. Mitra RK, Sinha SS, Maiti S, Pal SK. *J. Fluorescence.* 2009; 19:353–361.
42. Sarkar R, Pal SK. *Biopolymers.* 2006; 83:675–686. [PubMed: 16983651]
43. Shaikh SA, Ahmed SR, Jayaram B. *Arch. Biochem. and Biophys.* 2004; 429:81–99. [PubMed: 15288812]
44. Hoover WG. *Phys. Rev. A.* 1985; 31:1695–1697. [PubMed: 9895674]
45. Feller SE, Zhang YH, Pastor RW, Brooks BR. *J. Chem. Phys.* 1995; 103:4613–4621.
46. Priyakumar UD, MacKerell AD. *J. Chem. Theory Comp.* 2006; 2:187–200.
47. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD. *J. Chem. Theory Comp.* 2011; 8:348–362.
48. El Hassan MA, Calladine CR. *J. Mol. Biol.* 1998; 282:331–343. [PubMed: 9735291]
49. Yonetani Y, Kono H. *Biophys. J.* 2009; 97:1138–1147. [PubMed: 19686662]
50. Cheatham TE, Young MA. *Biopolymers.* 2000; 56:232–256. [PubMed: 11754338]
51. Ricci CG, de Andrade ASC, Mottin M, Netz PA. *J. Phys. Chem. B.* 2010; 114:9882–9893. [PubMed: 20614923]
52. Lankas F, Sponer J, Hobza P, Langowski J. *J. Mol. Biol.* 2000; 299:695–709. [PubMed: 10835278]
53. Lankas F, Sponer J, Langowski J, Cheatham TE. *Biophys. J.* 2003; 85:2872–2883. [PubMed: 14581192]
54. Perez A, Lankas F, Luque FJ, Orozco M. *Nucl. Acids Res.* 2008; 36:2379–2394. [PubMed: 18299282]

55. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T, Dixit S, Jayaram B, Lankas F, Laughton C, et al. Nucl. Acids Res. 2010; 38:299–313. [PubMed: 19850719]
56. Jana B, Pal S, Bagchi B. J. Phys. Chem. B. 2010; 114:3633–3638. [PubMed: 20170102]
57. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB. J Mol Biol. 2007; 371:725–738. [PubMed: 17585938]
58. Nadeau JG, Crothers DM. Proc. Natl. Acad. Sci. U.S.A. 1989; 86:2622–2626. [PubMed: 2704739]

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1.

Top: Schematic of water bridge associated with the spine of hydration. Bottom: Structure of the spine of hydration, which each water molecule represented as a red sphere.

Figure 2.
Distribution of RMSD of non-hydrogen atoms excluding end base pairs with respect to crystal structure 1S2R.

Figure 3.
Average interstrand phosphate distances for minor groove (top) and major groove (bottom).

Figure 4.
RMSF of backbone atoms in main strand of DNA, parmbsc0 in black, BMS in red and CHARMM27 in green.

Figure 5. Fraction and lifetime of one-water bridge in minor groove of DNA, parmbsc0 in black, BMS in red and CHARMM27 in green.

Figure 6. Minor groove first shell hydration pattern in the A₃T₃ region, overlapped with crystal water in the 1S2R structure (cyan). Contours show water oxygen (purple) density at level of 32 water molecules per nm³ from a grid with 0.25 Å resolution.

Figure 7.

Left panels: Water bridge occupancies for the 6 bp A-tracts A_3T_3 (black), A_6 (blue), $A_2T_2A_2$ (red), $(AT)_3$ (green). Middle panels: average lifetime of water in bridge sites for the same A-tracts. Right panels: Inter-basepair flexibility as measured by V_{step} for the same structures.

Figure 8. Water density in the minor groove for A₃T₃(black), A₆(blue), and A₂T₂A₂(red). MD results for parmbc0 are in cyan, showing the 40 water/nm³ (66 mol/L) contour. Nucleotides are colored blue for adenine and purple for thymine. Red spheres indicate waters in the A₃T₃ crystal structure.

Figure 9. Water bridge occupancies (left) and inter-basepair flexibility as measured by V_{step} (right) for PAT0 (black), PAT2 (red), PAT4 (green) and PAT6 (blue)

Figure 10.

Top: minor groove width for PAT0 (black), PAT2 (red), PAT4 (green) and PAT6 (blue). Middle: *slide* versus bp position for the same sequence. Bottom: *roll* versus bp position for the same sequences.

Figure 11.
Helical stiffness parameters (left: roll, right:slide) for PAT0 (black), PAT2 (red), PAT4 (green) and PAT6 (blue).

Scheme 1. Sequence and numbering for the CGCA₃T₃GCG dodecamer.

Scheme 2.

Structures of the 6 bp A-tracts, showing possible water bridge locations.

PAT0	A	A	A	A	A	A	A	A	A	A	A	A	A	A
PAT2	A	A	A	A	T	A	A	A	A	G	A	A	A	A
PAT4	A	A	A	T	T	A	A	A	C	G	A	A	A	A
PAT6	A	A	A	T	T	A	G	A	C	G	A	T	A	A

Scheme 3.

14 bp sequences used for this work. Note that the 5'-end is capped with CGC while the 3'-end is capped with GCG.

Table 1

Hydration numbers at specific sites in DNA with different force fields

Location	parmbc0	BMS	CHARMM27
Major groove	27.8±0.5	28.8±0.4	26.5±0.5
Major groove (A ₃ T ₃)	16.4±0.3	15.9±0.1	14.4±0.2
Minor groove	22.6±0.3	23.2±0.1	25.7±0.4
Minor groove (A ₃ T ₃)	10.5±0.2	11.2±0.1	12.4±0.3
Backbone oxygens (strand 1)	74.3±0.4	73.6±0.5	72.1±0.7
Backbone oxygens (strand 2)	74.1±0.2	73.8±0.4	71.9±0.7