3550–3556 *Nucleic Acids Research, 1999, Vol. 27, No. 17* *© 1999 Oxford University Press*

Probing the role of structural water in a duplex oligodeoxyribonucleotide containing a water-mimicking base analog

Dorina Kosztin, Richard I. Gumport[1] and Klaus Schulten[2,*]

Department of Chemistry, [1]Department of Biochemistry and [2]Department of Physics, University of Illinois at Urbana-Champaign, 405 North Matthews, Urbana, IL 61801, USA

Received April 23, 1999; Revised June 25, 1999; Accepted July 8, 1999

ABSTRACT

Molecular dynamics simulations were performed on models of the dodecamer DNA double-stranded segment, [d(CGCGAATTCGCG)]$_2$, in which each of the adenine residues, individually or jointly, was replaced by the water-mimicking analog 2'-deoxy-7-(hydroxymethyl)-7-deazaadenosine (hm^{7}c^{7}dA) [Rockhill,J.K., Wilson,S.R. and Gumport,R.I. (1996) *J. Am. Chem. Soc.*, 118, 10065–10068]. The simulations, when compared with those of the dodecamer itself, show that incorporation of the analog affects neither the overall DNA structure nor its hydrogen-bonding and stacking interactions when it replaces a single individual base. Furthermore, the water molecules near the bases in the singly-substituted oligonucleotides are similarly unaffected. Double substitutions lead to differences in all the aforementioned parameters with respect to the reference sequence. The results suggest that the analog provides a good mimic of specific 'ordered' water molecules observed in contact with DNA itself and at the interface between protein and DNA in specific complexes.

INTRODUCTION

DNA, the central informational molecule, is found in a variety of conformations, A-form and B-form DNA being the best known, and several detailed structures are available from X-ray crystallographic and NMR spectroscopic studies (1–8). The structure of DNA in crystals is critically influenced by temperature, crystal packing forces and hydration. Transition between A- and B-form DNA in solution is governed, in a complex way, by intrinsic and extrinsic parameters, with water having a determining role in the stability of the DNA double helix itself (3,4,9–11).

DNA hydration has been studied extensively, both theoretically and experimentally, and several recent review articles summarize the present state of knowledge (12–14). Computational studies on the hydration of DNA showed conformation-dependent differences in both geometry and the extent of hydration for the major and minor grooves (15–20). Three classes of water molecules surround DNA. In the first or inner layer, water molecules interact directly with phosphates, accessible sugar oxygens and the exposed polar groups of the bases. This shell was considered impermeable to ions, but recent studies suggest the possibility of cations in the first hydration shell (7,15,21). The second layer undergoes fast exchange of water molecules and ions with the bulk solvent water and is partially ordered with respect to the third layer, which is bulk solvent. Although NMR and thermodynamic methods provide valuable global information about the dynamics of DNA molecules at different humidities and ionic strengths, these methods fail to provide specifics concerning the positions of individual solvent molecules. The position of water molecules in the first and second hydration shells can be obtained from X-ray crystallography (3,4,7,21), theoretical studies (14,22–24), and computer simulations (15,17–20,25). For example, detection of a 'spine of hydration' in the minor groove of the A tract B-form DNA established the role of solvent in helping to determine the fine structure of DNA (3,7,26).

The role of water-mediated hydrogen bonds in protein–DNA interactions (27–30) and the potential contribution of water to the thermodynamics of DNA–ligand binding indicates the need to know the precise location of water molecules bound in the minor and major grooves of DNA. For example, the crystal structure of the *trp* repressor/operator complex, as well as other specific protein–DNA complexes (29), revealed the read-out of the DNA sequence through water-mediated contacts to the sugar phosphate backbone and water molecules positioned between specific groups on the bases and amino acid sidechains (27). The importance of certain of these water molecules in DNA recognition is supported by a second structure (31) and by mutation and chemical modification of the DNA (32). For instance, replacing the amino acids of the water-mediated interactions with amino acids not capable of hydrogen bonding to the water molecules in the *trp* repressor resulted in repressors with reduced affinities for the operator sequence (32).

Nucleoside analogs offer an alternative means to explore the role of water molecules in DNA and in protein–DNA interactions. Rockhill *et al.* (33) suggested that the analog 2'-deoxy-7-hydroxymethyl-7-deazaadenosine (hm^{7}c^{7}dA) (Fig. 1a) might mimic a bound water molecule complexed to an adenine residue in the major groove of DNA. By analogy, a cyclic urea that incorporates an analogous 'structural' water has been

*To whom correspondence should be addressed. Tel: +1 217 244 2212; Fax: +1 217 244 6078; Email: kschulte@ks.uiuc.edu

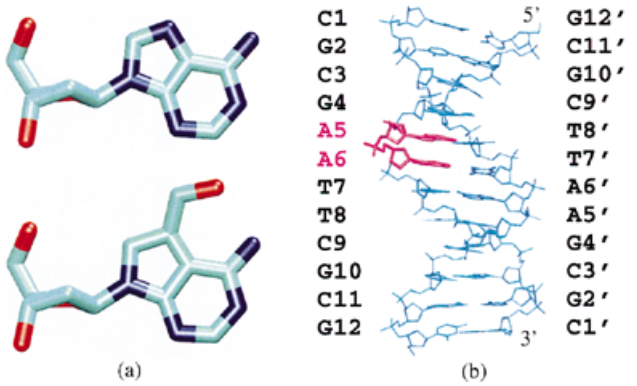

Figure 1. (**a**) Adenosine (top) and 2'-deoxy-7-hydroxymethyl-7-deazaadenosine (hm^7c^7dA) (bottom). Coloring scheme for atoms: red, oxygen; cyan, carbon; blue, nitrogen. (**b**) Crystal structure of the DNA dodecamer used in the simulation and numbering scheme for the DNA bases. The adenine bases substituted individually or jointly with the analog are represented in red.

synthesized and shown to be an inhibitor of an HIV protease (34). The nucleotide analog hm^7c^7dA replaces the hydrogen bond between one hydrogen of a water molecule and N7 of the purine ring (as found in the crystal structure of the *trp* repressor/operator complex) with a covalent bond between the carbon of a methylene group and the C7 of the deazapurine ring. The remaining O-H group of the water molecule is replaced by the hydroxyl group of the analog. In using the analog to study the role of water molecules in DNA and in protein–DNA interactions, one postulates that the oxygen of the hydroxymethyl group is functionally equivalent, as a hydrogen bond acceptor, to the oxygen of the water molecule and that it will be similarly located. Other assumptions are that the analog will neither disrupt base pair stacking nor hydration properties of normal B-form DNA. This molecular dynamics study was initiated to investigate these assumptions.

The presence and nature of defined hydration sites on the surface of DNA, as well as the dynamic structural properties of the modified DNA, and the influence of the analog on DNA hydration were examined. We used the B-form duplex dodecamer $[d(CGCGAATTCGCG)]_2$ (4), and analog-substituted derivatives. This DNA molecule has been studied extensively both structurally and theoretically (1–4,7,19,35,36). The analog hm^7c^7dA was modeled into the DNA dodecamer, substituting the two adenine bases in one strand (A5 and A6) individually or jointly (Fig. 1b). The effect of the substitution on the stability and conformation of the DNA was assessed. Structural deviations of DNA from the X-ray crystal structure (4) of the unmodified DNA were evaluated as root mean square deviations (RMSD). The conformation of DNA in each simulation was analyzed by evaluating the inter-base pair parameters of roll, tilt and twist. The hydration of the DNA was assessed by determining the extent to which the dynamics of the solvent particles were modified in the vicinity of the substituted adenine bases. We compare our results with those from calculations of the normal DNA sequence and correlate them with experimental data from X-ray crystallography and NMR studies.

MATERIALS AND METHODS

The simulations were based on the X-ray crystallographic structure of the duplex dodecamer d(CGCGAATTCGCG) at 1.9 Å resolution (4) (entry BDL020 in the Nucleic Acid Databank). All simulations were performed using version 22 of the CHARMM force field (37,38) and the molecular dynamics program NAMD (39) that provides the option of calculating full electrostatic interactions through the use of a multipole expansion algorithm, namely the program DPMTA (40). The charge distribution for the hm^7c^7dA analog was obtained by means of GAUSSIAN-94 (41) at the Hartree–Fock level with a 6-31G* basis set, using the coordinates of heavy (non-hydrogen) atoms from the crystal structure [available as supplementary material for (33) at http://pubs.acs.org] with hydrogens generated by the program QUANTA (42). The equilibrium bond lengths, angles, torsional angles and force constants for the hm^7c^7dA analog were derived from its coordinates and force field parameters of molecules with similar chemical structure available in the CHARMM22 force field.

Four distinct systems, based on the DNA dodecamer d(CGCGAATTCGCG) sequence, were constructed: the dodecamer itself, labeled dna; the dodecamer with the adenine base at position 5 replaced by the hm^7c^7dA analog, labeled dna^5; the dodecamer with the adenine base at position 6 replaced by the analog, labeled dna^6; and the dodecamer with both A5 and A6 bases replaced by the analog, labeled $dna^{5,6}$ (see Fig. 1b for numbering scheme). Note that, although the dodecamer is a palindrome, substitutions were simulated in one strand only of the duplex. The protocols used for solvating and simulating the four systems were identical. Hydrogen bonds were added using X-PLOR (43) followed by 1000 steps of minimization. The integrity of the DNA structure depends sensitively on the local environment; accordingly, a water bath together with counter-ions are required to stabilize the DNA conformation during the simulations (44–47). The DNA was surrounded by a 15 Å layer of water molecules. To counter-balance the negative charge on the DNA backbone, 15 sodium ions were added by replacing 15 water molecules with the highest electrostatic energies of the oxygen atom. By adding 15 sodium ions, 75% of the DNA charge was neutralized, in accordance with Manning's theory (48). Each of the four systems contained ~11 500 atoms. The systems were further equilibrated for 50 ps with the atoms of the terminal base-pairs harmonically constrained. For each system, 1 ns of dynamics was performed: 200 ps of dynamics with soft harmonic constraints on the terminal base-pairs of the DNA and 800 ps of free dynamics, without constraints.

Structural deviations of the DNA were assessed as RMSD from the initial X-ray crystal structure using a least-square fitting algorithm (49). To describe the solvation of DNA bases, radial distribution functions $g(r)$ of the water molecules in the vicinity of the relevant atoms of the bases were calculated using a routine available in X-PLOR (43). The function $g(r)$ is defined as the average water-oxygen or water-hydrogen density on a sphere with radius r around a given site in the DNA. Conformational changes of DNA in each simulation were analyzed using the Molecular Dynamics Analysis Toolchest (50,51). The terminal base pairs were excluded from all analyses. Hydrogen bond interactions between DNA and water molecules were analyzed using the following convention:

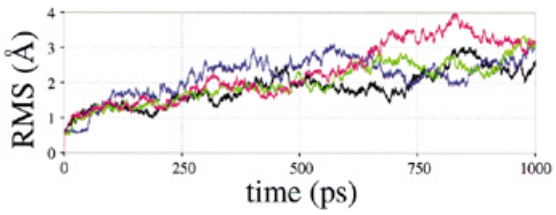

Figure 2. RMS deviation for all heavy atoms of DNA in the simulations dna (black line), dna[5] (blue line), dna[6] (green line) and dna[5,6] (red line).

two atoms were considered to form a hydrogen bond (A···H-D) if the acceptor–donor distance was <3.5 Å and if the A-H-D angle was between 120° and 180°.

RESULTS

Simulations of the DNA sequences in dna[5], dna[6], and dna[5,6] were compared with that of the unmodified normal duplex DNA segment, in dna. The results will be described in two sections covering the dynamic stability of the simulated DNA segments along their trajectories and the hydration of the DNA.

Trajectory and dynamic stability of simulated DNA segments

The stability of structures in molecular dynamics simulations is often characterized by the RMSD of atomic positions from a suitable reference structure. The time evolution of the RMSDs for the four simulated systems, calculated as described in Materials and Methods, is presented in Figure 2. For the first 400 ps, the RMSDs for DNA in the dna, dna[5] and dna[6] simulations remains below 2 Å, increasing slightly to values <3 Å until the end of the 1 ns simulation. For dna[5,6] the RMSD values are higher, reaching values of 4 Å around 840 ps simulation time and dropping back to 3.3 Å at the end of the 1 ns simulation. These results indicate that dna, dna[5] and dna[6] are all stable, i.e. behave similarly during the simulations and that the control dna behaves as observed in other simulations (16,19,20).

Most of the conformational changes in the various trajectories, for all four systems, were local fluctuations. To illustrate this, DNA structures for each system, at intervals spaced at 50 ps, were superimposed. The results are presented in Figure 3. As shown, the backbone and the terminal base pairs exhibited large scale movements, partially as a result of their extensive contacts with the surrounding solvent molecules. Fraying of the terminal base pair ends is observed in all simulations, but most prominently in the dna[5,6] simulation. Such effects are well known in DNA simulations and more study is needed to firmly establish their causes. The hydrophobic bases are partially buried inside the DNA molecule and therefore the DNA is most stable near the center of the molecule. One notable event during the simulations is the bending of the DNA ends, primarily at the 5' termini. In all simulations, the DNA axis has a gently writhed conformation that is a consequence of generally non-zero roll angles at individual base steps. Analysis of the axis bending indicates that the bending regions are at or near the junction between AATT and CGCG tracts, as found in the majority of the crystal structures of this DNA sequence. The bend is primarily due to displacement in the roll parameter toward the major groove at G2-C3 and G4-A5 steps at the 5' end and at G10-C11 at the 3' end. Pronounced oscillations of the DNA axis arise for dna[5,6] simulation and both dna[5] and dna[6] appear more 'flexible' than the reference duplex, as presented in Figure 3.

The behavior of the substituted base can be characterized by its hydrogen bonds. Analysis of the crystal structure of the hm[7]c[7]dA analog, shown in Figure 1, indicates a hydrogen bond between the amino-proton on N6 and the O of the hydroxymethyl group. NMR data also support the existence of this hydrogen bond (33). This hydrogen bond is well maintained during the entire simulation time for all the systems with substituted bases. Occasionally, the hydroxymethyl group will rotate <90° and contact acceptors on the neighboring bases. These events take place on a femtosecond time scale over long time intervals

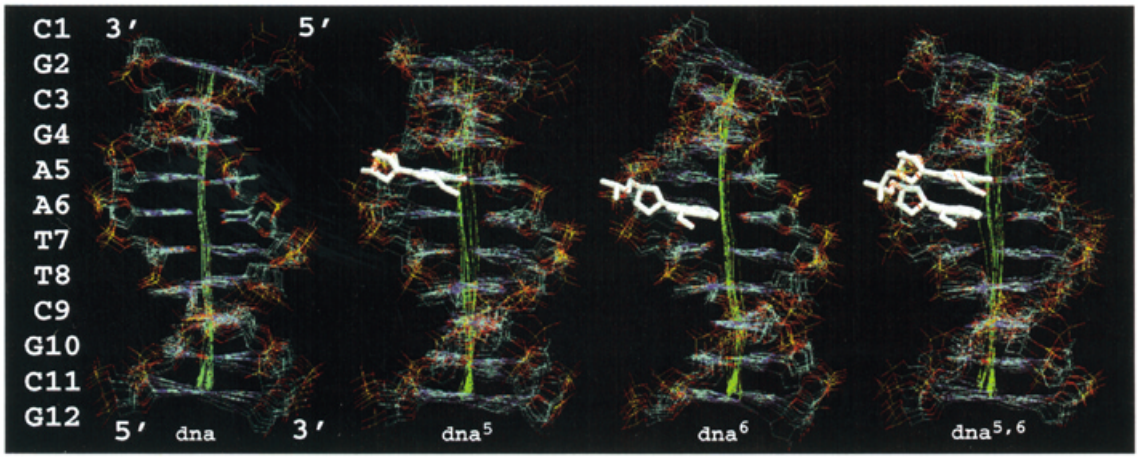

Figure 3. Superposition of DNA structures and DNA axis (snapshots taken at 50 ps intervals) in the simulations dna, dna[5], dna[6] and dna[5,6]. The substituted bases in each system are shown in white in a liquorice representation.

and therefore may be reasonably considered as fluctuations of the hydroxymethyl group around its equilibrium position.

The average values for the DNA backbone and the sugar torsion angles, are, with a few exceptions, near those of the crystal structure. Angles (α, β and γ) all oscillate around the values observed in the crystal. The angles describing the dynamic sugar conformation, specifically the glycosidic torsion angle χ, the sugar pucker P, and the backbone angle δ, show higher deviations from the crystal values, due to the significant fluctuations occurring during the simulations. These fluctuations reflect motions on a time scale that are averaged out by X-ray crystallography. The C2'-endo structure (sugar pucker phase angle between 144° and 180°) is characteristic of B-form DNA and is favored in all simulations. The sugar rings show transient puckering to the C3'-endo value (sugar pucker phase angle between 0° and 36°) characteristic of A-form DNA, as confirmed by NMR experiments (5).

The average major and minor groove widths are similar for all four structures. For the major groove, distances between a phosphate on one strand and a phosphate five steps away on the other strand were measured since these phosphates define the shortest distance between strands; for the minor groove the analogous, shortest distances involved phosphates removed by two residue positions. The major and minor groove distances (in Å) are presented in Tables 1 and 2. The results imply that the helix is capable of undergoing significant groove width fluctuations within the framework of the simulations. The minor groove in the AATT region failed to narrow in our simulations as it did in the simulations reported by Young *et al.* (15) and Duan *et al.* (19). This difference may be a consequence of the unwinding of the DNA during our simulations as well as of different conditions and parameters used for the above mentioned simulations, e.g., the use of periodic boundary conditions.

Table 1. Average values for the major groove distances (Å) in the four simulated systems: dna, dna^5, dna^6 and dna^{5,6}

	dna	dna^5	dna^6	dna^{5,6}
G2-T7'	19.40	17.35	15.82	15.97
C3-T8'	19.85	20.46	19.33	20.50
G4-C9'	19.39	20.46	19.33	19.22
A5-G10'	15.84	16.02	16.93	15.90
A6-C11'	14.17	15.20	15.32	13.84
T7-G12'	19.61	20.70	21.51	21.17

The distances were determined between phosphates of individual residues, e.g., the phosphate of G2 on the 5' to 3' strand and the phosphate of T7' on the other DNA strand (see text).

The A and B conformations of DNA have distinct geometric parameters that can reliably distinguish between the two forms: the base pair X displacement (XDP), i.e., the distance from the central helical axis to the center of the base pairs, and the inclination of the base pairs (INC) with respect to the helical axis. In typical A-DNA, the X displacement measures ~5 Å and the INC parameter is ~19°, whereas in B-DNA XDP measures zero or slightly negative and the INC parameter is approximately –6°. The average values and standard deviations of the XDP and INC parameters indicate that the DNA remains in the B conformation in all four simulations throughout the entire trajectory.

Table 2. Average values for the minor groove distances (Å) in the four simulated systems: dna, dna^5, dna^6 and dna^{5,6}

	dna	dna^5	dna^6	dna^{5,6}
A5-G12'	13.09	14.90	16.24	14.95
A6-C11'	14.17	15.20	15.32	13.84
T7-G10'	14.01	14.69	14.86	14.15
T8-C9'	13.07	14.12	14.01	14.95
C9-T8'	12.58	12.34	12.27	12.52
G10-T7'	11.97	13.15	11.64	14.17
C11-A6'	12.86	13.76	13.11	13.43
G12-A5'	13.88	14.81	14.70	13.19

The distances were determined between phosphates of individual residues, e.g., the phosphate of A5 on the 5' to 3' strand and the phosphate of G12' on the other DNA strand (see text).

The twist angle (helix repeat) characterizes the relative rotation of two adjacent base pairs along the DNA axis with an average value of 34.3° (10.5 bp/turn) in canonical B-form DNA. In our simulations, the helix repeat eventually exceeded 11.0 bp/turn, indicating an underwound conformation of the DNA. The observed unwinding of the DNA during simulations beyond the solution value of 10.5 bp/turn (52) is expected since DNA typically becomes underwound in such simulations (53,54). As a result of the unwinding, there are kinks in the DNA backbone. The unwinding of the DNA helical axis is a slow process compared with rise and roll processes: unwinding at one step overwinds its neighboring step. Therefore, the observed kinks might disappear if the simulation time were increased.

DNA hydration

The dominant interactions between water molecules and DNA are caused by hydrogen bonding between the DNA polar groups and water molecules. Therefore, a detailed analysis was performed for water molecules hydrogen bonded to the DNA. The negatively charged phosphate oxygens bind water more tightly than do the deoxyribose oxygens or the heteroatoms of the bases, as shown in studies of DNA at relative low humidities where water is located mainly around phosphate groups (55). The precise positions of individual water molecules around the phosphate oxygens are not unique, being much less well defined than those in the grooves. Hydration of phosphate groups is highly dynamic and relatively few ordered water molecules are found occupying the first coordination shells of phosphate oxygens.

Both major and minor grooves are filled with water molecules. As observed in crystal structures, the hydration sites most often lie in or near the plane of the base pairs. In order to assess the influence of the substituted base on the overall DNA hydration, special attention was accorded not only to the substituted base(s), but also to the neighboring ones, in particular all the

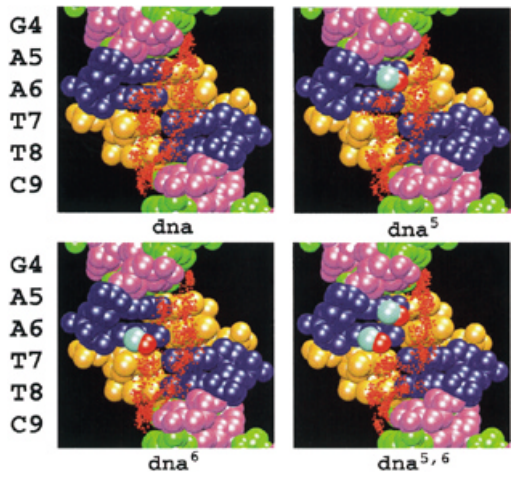

Figure 4. The position of water molecules within hydrogen bonding distance (proton–acceptor distances <2.5Å and angle donor–proton–acceptor >120°) of the acceptor/donor groups of the GAATTC bases were selected and are shown superimposed (for a better representation the oxygen atoms of the water molecules are shown as vdW sphere with a reduced radius). Each sphere represents a water molecule observed sometime during the simulation. The DNA atoms are in vdW representation, with each base represented by a different color. The C10 and O11 atoms of the analog are shown in cyan and red, respectively.

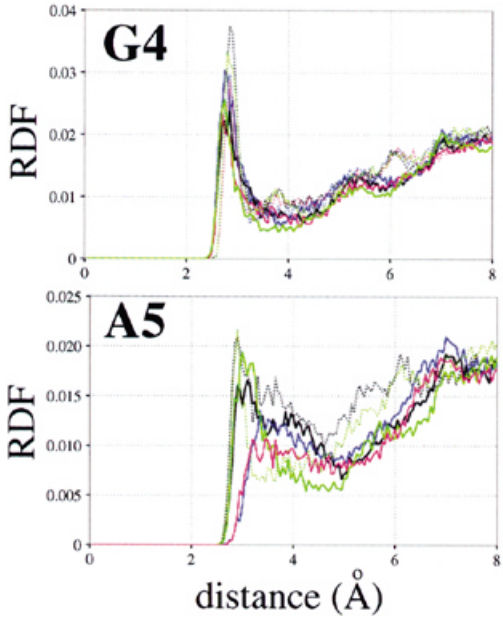

Figure 5. Unnormalized radial distribution functions *g(r)* of water-oxygen density around O6 (solid line) and N7 (dotted line) groups of guanine 4 (top) and N6 (solid line) and N7 (dotted line) groups of adenine 5 (bottom). Colors are defined as in Figure 2.

bases in the 5'-GAAT-3' region. During the simulations, donors and acceptors of the DNA bases were hydrogen bonded to water molecules between 35 and 85% of the time, depending on the identity of the DNA base group. Study of the dynamics of individual water molecules shows that, in the major groove, water molecules that form hydrogen bonds to specific groups on the DNA bases have short residence times. This mirrors the conclusions from NMR studies of the *trp* operator (56). One water molecule is replaced by another in such a way that the hydrogen bond is essentially maintained. Some water molecules visit the same site several times during the 1 ns simulation, and others migrate from site to site. Analogously, fluctuating water positions were observed in *Antennapedia* homeodomain–DNA complex (57). Water molecules forming transient hydrogen bond bridges between two and even three donors/acceptors of adjacent bases were detected but for short periods of time. The hydration of G and A bases is very similar; the main hydration sites being found, as in X-ray crystallography (23), in or close to the plane of the base within hydrogen bonding distance of their polar atoms. The hydration site for O6 is distinct from the hydration site of N7 in guanines: each group hydrogen bonds water molecules individually for 75–85% of the simulation time. The two hydration sites for N6 and N7 of adenine bases are discernible during the entire simulation time in all systems. Two distinct water molecules are found within hydrogen bonding distance of the two groups for 35–55% of the simulation time. Occasionally, for short time periods (~1–5 ps), one water molecule bridges between the two groups in the A and G bases. The hydration sites on the C and T bases are also very similar to one another. One difference is that hydration of the thymine base major groove is influenced by its methyl group, which moves the hydration site closer to the paired adenine. It is noteworthy that the presence of the analog fails to affect the hydration of the paired thymine bases.

The hydration sites for the 5'-GAAT-3' bases in the four simulations are shown in Figure 4. Because of the dynamics of the bases, these water molecules are sometimes not at hydrogen-bonding distances from the donor/acceptor atoms although the same rules (described in Materials and Methods) were applied for identifying them as hydrogen-bonded molecules. One should remember that the DNA bases fluctuate and the location of the water molecules is influenced by the base position at each moment. The hydration of the substituted bases themselves presents unique features. Analysis of all the water molecules within hydrogen bonding distance of the analog indicates that it gives rise to two hydration regions: one site close to the backbone of the DNA and the other site near the partner thymine. The presence of the hydroxymethyl group of the analog affects the hydration of only the substituted base and thereby interrupts the 'chain' of free water molecules lying along the edges of the bases in the major groove.

Precise description of the solvation of the backbone, or of specific parts of a base, can also be obtained from radial distribution functions (RDFs) of the water molecules in the vicinity of the relevant sites. To elucidate sequence effects in the DNA–water distribution, as well as the effect of the base analog on hydration, the RDF was decomposed into contributions corresponding to each base pair. The effect of the substituted analog in the major groove can be seen in the RDF of water around the N6 group in base A5, shown in Figure 5. For dna[5] and dna[5,6] the hydration of N6 in A5 (the site of the analog) is obviously hindered by the presence of the hydrogen bond between the N6 amino-proton and the O of the hydroxymethyl group. This is reflected by the more diffuse peaks present in the RDF for water around the N6 of A5 as opposed to the sharp peaks in the dna and dna[6] systems that indicate the existence of a clearly

defined solvation shell around the N6 group of A5. The RDF of water around the N6 group of A6 is relatively unaltered by the presence of the analog at base A5. The RDFs of water near the O6 and N7 groups of G4 are very similar in all systems, indicating that the substitution of the analog does not significantly alter hydration of the G4 base. These results allow the important conclusion that analog substitution should be without effect on the hydration pattern of immediately adjacent bases.

As might have been anticipated, substitutions of a functional group in the major groove have no direct effect on minor groove hydration. However, structural perturbations introduced by the analog might affect minor groove geometry and hence hydration. Calculations and experiments indicate that each base has one hydration site in the minor groove with the exception of the central AATT region, where water densities of neighboring bases fuse together and create five hydration sites, the 'spine of hydration' (3,4). In our simulations, this spine was present at the beginning of the simulations and remained, more or less, in a stable state for the entire time of the simulation.

CONCLUSIONS

What have we learned about DNA and DNA hydration from molecular dynamics simulations? Several MD simulations have been reported on DNA *in vacuo* or in water, including ions explicitly or implicitly through the use of a reduced charge on the phosphates of the DNA backbone, or by a salt-dependent potential of mean force between phosphates (16,19,20,35, 44,46,53,58). These studies indicated that accurate, nanosecond all-atom molecular dynamics simulations including water and ions explicitly can achieve a realistic description of the DNA double helix in solution. The simulations yielded useful information about the structure, dynamics and solvation of the double helix. Even though molecular dynamics simulations of DNA (53,54), using different methodologies and parameter sets, have shown that a given force field may favor one DNA form over another, Cheatham and Kollman showed that careful MD simulations of DNA in various environments can reproduce environment-dependent structural changes with reasonable success (59).

We have expanded past studies by performing a detailed MD simulation study on a DNA containing a water-mimicking base analog. The 1 ns molecular simulations presented in this paper have explored the structural and dynamic features of B-form DNA, as well as modified DNA, in aqueous medium. Overall, the results show that the simulated systems have structures that remain remarkably similar to that of the crystal structure (4). Helix bending toward the major groove at G-A and T-C steps, with the most prominent effect in dna[5,6], and some degree of unwinding were observed in all systems, as found in other molecular dynamics simulations of this sequence (16,19,20). Analog substitution of both adenine bases led to higher fluctuations of the DNA (reflected in the RMSD and the DNA axis fluctuations) than did single-base replacements. The simulations revealed that the postulated intramolecular bond between the exocyclic amino group of the analog and the hydroxymethyl group was maintained during the entire simulation. Rotations of the hydroxyl group were observed over short times. The ability of the hydroxymethyl group to rotate may provide the flexibility necessary to adapt to conformations required for proteins that bind specifically to it (33). Molecular modeling of the analog as a deoxynucleoside by J.Rockhill (personal communication) indicated that the hydroxymethyl group adopts different low energy conformations; however, the preferred conformations have the hydroxymethyl group pointed toward and forming a hydrogen bond with the amino proton. The comparatively fewer rotations observed in our simulations may reflect the simulation assumptions or result from the analog being in a DNA helix.

We found the hydration of the DNA to be a highly dynamic process, especially in the major groove, with average residence times for water molecules varying from 10 to 200 ps. Analysis of hydration of the major groove showed that the hydroxymethyl group of the analog replaces the hydration site of the substituted adenine base. This leads to a redistribution of water molecules around the substituted base without affecting the hydration of the neighboring bases. NMR studies on water residence times in the DNA major groove find values significantly less than 1 ns at 4°C (17,36).

Although sequence-specific effects cannot be entirely ruled out, this analysis suggests that substituting one base with the analog fails to significantly affect the structural properties of the DNA. Overall, our results with respect to the stability and hydration of DNA are consistent with those obtained in other molecular dynamics simulations (16,19,20). Detailed comparisons are complicated because different programs were used to analyze the DNA parameters and different criteria were used to define hydration (60). In addition, although of the same sequence, the simulations used different crystal structures as starting points.

The arrangement of water molecules lining the DNA grooves is both sequence and structure dependent (14,22). Hence, depending on their residence times, such water molecules might be used as specific recognition elements of the DNA target by a sequence-specific binding protein. By replacing an adenine base with the (hm^7c^7dA) analog in protein–DNA complexes that have been proposed to use water-mediated interactions, a permanently bound 'water molecule' would be achieved. A 'locked' water molecule could then be studied using the variety of biochemical and physical techniques previously employed. The hydroxymethyl analog may make it possible to distinguish a specific water molecule from the remainder and to assess its effect on specific protein–DNA interactions, as previously achieved with an analogous enzyme inhibitor (34).

This paper demonstrates the potential for molecular dynamics to aid in the investigation of DNA sequences and their recognition by proteins, especially when relevant crystallographic structures are unavailable. Molecular dynamics simulations can guide experiments, indicating in the present case, for example, that mono-substitutions of bases by the hm^7c^7dA analog do not affect DNA structure and its hydration shell, whereas double substitutions involving contiguous bases may alter the system, thereby rendering interpretation of experimental data more difficult. Experimental observations often provide either incomplete information or global parameters, e.g., binding affinities or information on limited structural features, whereas MD simulations provide a more complete picture of individual atomic arrangements and the associated dynamics. However, the problem is that they are limited in their degree of realism: systems can presently be described in simulations over periods of only a few nanoseconds and the force fields employed are not completely realistic. The main problem in this regard is the slight unwinding of DNA caused by an imperfect force field;

better estimates of stacking interactions including polarizabilities of the conjugated π electrons in the bases may remedy this problem. In addition, computer power is likely to increase to a level that would permit analysis of longer time scales relevant for the rearrangement of ions and of more geometries. Nevertheless, at present, with the caveats mentioned and when analyzed in close conjunction with experimental observations, simulations promise to be a valuable tool in DNA research.

ACKNOWLEDGEMENTS

We would like to thank Jason Rockhill for helpful discussions and a critical reading of the manuscript and R. Brunner and J. Phillips for help with NAMD. We also thank J. Carey, Princeton University, for helpful comments. This work was supported by grants from the National Institutes of Health (PHS 5 P41 RR05969-04), the National Science Foundation (BIR-9318159, BIR 94-23827 EQ), the MCA 93S028P computer-time grant at Pittsburgh Supercomputing Center and the Roy J. Carver Charitable Trust. The figures in this paper were created with the molecular graphics program VMD (61) (http://www.ks.uiuc.edu/Research/VMD).

REFERENCES

1. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itikura,K. and Dickerson,R.E. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 2179–2983.
2. Dickerson,R.E. and Drew,H.R. (1981) *J. Mol. Biol.*, **149**, 761–786.
3. Drew,H.R. and Dickerson,R.E. (1981) *J. Mol. Biol.*, **151**, 535–556.
4. Westhof,E. (1987) *J. Biomol. Struct. Dyn.*, **5**, 581.
5. Alam,T.M., Orban,J. and Drobny,G.P. (1991) *Biochemistry*, **30**, 9229–9237.
6. Cognet,J.A.H., Boulard,Y. and Fazakerley,G.V. (1991) *J. Mol. Biol.*, **246**, 209–226.
7. Shui,X., McFail-Isom,L., Hu,G.G. and Williams,L.D. (1998) *Biochemistry*, **37**, 8341–8355.
8. Egli,M. (1998) *Antisense Nucleic Acid Drug Dev.*, **8**, 123–128.
9. Franklin,R.E. and Gosling,R. (1953) *Acta Cryst.*, **6**, 673–677.
10. Ivanov,V.I., Minchenkova,L.E., Frank-Kamentskii,M.D. and Schyolkina,A.K. (1974) *J. Mol. Biol.*, **87**, 817–833.
11. Sharp,K.A. and Honig,B. (1995) *Curr. Opin. Struct. Biol.*, **5**, 323–328.
12. Beveridge,D.L. and Ravishanker,G. (1994) *Curr. Opin. Struct. Biol.*, **4**, 246–255.
13. Louise-May,S., Auffinger,P. and Westhof,E. (1996) *Curr. Opin. Struct. Biol.*, **6**, 289–298.
14. Berman,H.M. (1997) *Biopolymers*, **44**, 23–44.
15. Young,M.A., Jayaram,B. and Beveridge,D.L. (1997) *J. Am. Chem. Soc.*, **119**, 59–69.
16. Young,M.A., Ravishanker,G. and Beveridge,D.L. (1997) *Biophys. J.*, **73**, 2313–2336.
17. Denisov,V.P., Carlstrom,G., Venu,K. and Halle,B. (1997) *J. Mol. Biol.*, **268**, 118–136.
18. Tippin,D.B. and Sundaralingam,M. (1997) *Biochemistry*, **36**, 536–543.
19. Duan,Y., Wilkosz,P., Crowley,M. and Rosenberg,J.M. (1997) *J. Mol. Biol.*, **272**, 553–572.
20. Bonvin,A.M.J., Sunnerhagen,M., Otting,G. and van Gunsteren,W.F. (1998) *J. Mol. Biol.*, **282**, 859–873.
21. Shui,X., Sines,C.C., McFail-Isom,L., VanDerveer,D. and Williams,L.D. (1998) *Biochemistry*, **37**, 16877–16887.
22. Schneider,B., Cohen,D. and Berman,H.M. (1992) *Biopolymers*, **32**, 725–750.
23. Schneider,B. and Berman,H.M. (1995) *Biophys. J.*, **69**, 2661–2669.
24. Feig,M. and Pettitt,B.M. (1998) *Structure*, **6**, 1351–1354.
25. Feig,M. and Pettitt,B.M. (1999) *J. Mol. Biol.*, **286**, 1075–1095.
26. Dickerson,R.E. (1992) *Methods Enzymol.*, **211**, 67–111.
27. Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) *Nature*, **335**, 321–329.
28. Feng,J., Johnson,R.C. and Dickerson,R.E. (1994) *Science*, **263**, 348–355.
29. Schwabe,J.W.R. (1997) *Curr. Opin. Struct. Biol.*, **7**, 126–134.
30. Woda,J., Schneider,B., Patel,K., Mistry,K. and Berman,H.M. (1998) *Biophys. J.*, **75**, 2170–2177.
31. Lawson,C.L. and Carey,J. (1993) *Nature*, **366**, 178–182.
32. Joachimiak,A., Haran,T.E. and Sigler,P.B. (1994) *EMBO J.*, **13**, 367–372.
33. Rockhill,J.K., Wilson,S.R. and Gumport,R.I. (1996) *J. Am. Chem. Soc.*, **118**, 10065–10068.
34. Lam,P.Y., Pu,Y., Jadhav,P.K., Aldrich,P.E., DeLucca,G.V., Eyerman,C.J., Chang,C.H., Emmett,G., Holler,E.R., Daneker,W.F. *et al.* (1996) *J. Med. Chem.*, **39**, 3514–3525.
35. Prevost,C., Louise-May,S., Ravishanker,G., Lavery,R. and Beveridge,D.L. (1993) *Biopolymers*, **33**, 335–350.
36. Phan,A.P., Leroy,J.L. and Gueron,M. (1999) *J. Mol. Biol.*, **286**, 505–519.
37. MacKerell,A.D.Jr, Bashford,D., Bellott,M., Dunbrack,R.L.Jr, Evanseck,J., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S., Joseph,D., Kuchnir,L., Kuczera,K., Lau,F.T.K., Mattos,C., Michnick,S., Ngo,T., Nguyen,D.T., Prodhom,B., Roux,B., Schlenkrich,M., Smith,J., Stote,R., Straub,J., Watanabe,M., Wiorkiewicz-Kuczera,J., Yin,D. and Karplus,M. (1992) *FASEB J.*, **6**, A143.
38. MacKerell,A.D.Jr, Bashford,D., Bellott,M., Dunbrack,R.L.Jr, Evanseck,J., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S., Joseph,D., Kuchnir,L., Kuczera,K., Lau,F.T.K., Mattos,C., Michnick,S., Ngo,T., Nguyen,D.T., Prodhom,B., Reiher,I.W.E., Roux,B., Schlenkrich,M., Smith,J., Stote,R., Straub,J., Watanabe,M., Wiorkiewicz-Kuczera,J., Yin,D. and Karplus,M. (1998) *J. Phys. Chem. B*, **102**, 3586–3616.
39. Nelson,M., Humphrey,W., Gursoy,A., Dalke,A., Kalé,L., Skeel,R.D. and Schulten,K. (1996) *Int. J. Supercomput. Appl. High Perform. Comput.*, **10**, 251–268.
40. Rankin,W. and Board,J. (1995) *IEEE Symposium on High Performance Distributed Computing*. In press. [Duke University Technical Report 95-002.]
41. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Gill,P.M.W., Johnson,B.G., Robb,M.A., Cheeseman,J.R., Keith,T., Petersson,G.A., Montgomery,J.A., Raghavachari,K., Al-Laham,M.A., Zakrzewski,V.G., Ortiz,J.V., Foresman,J.B., Peng,C.Y., Ayala,P.Y., Chen,W., Wong,M.W., Andres,J.L., Replogle,E.S., Gomperts,R., Martin,R.L., Fox,D.J., Binkley,J.S., Defrees,D.J., Baker,J., Stewart,J.P., Head-Gordon,M., Gonzalez,C. and Pople,J.A. (1995) Gaussian 94, Revision B.3, Gaussian Inc., Pittsburgh, PA.
42. Molecular Simulations Inc., Burlington, Massachusetts (1994) QUANTA 4.0.
43. Brünger,A.T. (1992) X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR, The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University.
44. Seibel,G.L., Singh,U.C. and Kollman,P.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 6537–6540.
45. York,D.M., Darden,T. and Deerfield,D. (1992) *Int. J. Quantum Chem. Symp.*, **19**, 145–166.
46. Miaskiewicz,K., Osman,R. and Weinstein,H. (1993) *J. Am. Chem. Soc.*, **115**, 1526–1537.
47. Jayaram,B., Aneja,N., Rajasekaran,E., Arora,V., Das,A., Ranganathan,V. and Gupta,V. (1994) *J. Sci. Indust. Res.*, **53**, 88–105.
48. Manning,G.S. (1978) *Q. Rev. Biophys.*, **2**, 179–246.
49. Kabsch,W. (1976) *Acta Cryst.*, **2**, 922–923.
50. Ravishanker,G., Swaminathan,S., Beveridge,D.L., Lavery,R. and Sklenar,H. (1989) *J. Biomol. Struct. Dyn.*, **6**, 669–699.
51. Sklenar,H., Etchebest,C. and Lavery,R. (1989) *Proteins Struct. Funct. Genet.*, **6**, 46–60.
52. Klug,A. and Rhodes,D. (1980) *Nature*, **286**, 573–578.
53. Cheatham,T.E. and Kollman,P.A. (1996) *J. Mol. Biol.*, **259**, 434–444.
54. Yang,L.Q. and Pettitt,B.M. (1996) *J. Phys. Chem.*, **100**, 2564–2566.
55. Milton,J.G. and Galley,W.C. (1986) *Biopolymers*, **25**, 1673–1684.
56. Sunnerhagen,M., Denisov,V.P., Venu,K., Bonvin,A.M.J., Carey,J., Halle,B. and Otting,G. (1998) *J. Mol. Biol.*, **282**, 847–858.
57. Billeter,M., Güntert,P., Luginbühl,P. and Wüthrich,K. (1996) *Cell*, **85**, 1057–1065.
58. Swaminathan,S., Ravishanker,G. and Beveridge,D.L. (1991) *J. Am. Chem. Soc.*, **113**, 5027–5040.
59. Cheatham,T.E. and Kollman,P.A. (1997) *Structure*, **15**, 1297–1311.
60. Babcock,M.S. and Olson,W.K. (1994) *J. Mol. Biol.*, **237**, 98–124.
61. Humphrey,W.F., Dalke,A. and Schulten,K. (1996) *J. Mol. Graphics*, **14**, 33–38.