Nanosecond Dynamics and Structure of a Model DNA Triple Helix in Saltwater Solution

Samantha Weerasinghe, Paul E. Smith, V. Mohan, Y.-K. Cheng, and B. Montgomery Pettitt*

Contribution from the Department of Chemistry, University of Houston, Houston, Texas 77204-5641

Received July 25, 1994[⊗]

Abstract: The structure and stability of a DNA triple helix was examined by molecular dynamics (MD) simulation using an all-atom force field. A 1.3 ns simulation was performed on a $d(CG\cdot G)_7$ triple helix in a 1 M saltwater solution. The Ewald method was used to calculate the electrostatic interactions of the system. The behavior of the DNA in the saltwater solution was determined by examining the structure, energetics, and mobility of water and ions in the system. The simulation results for the helical parameters support the validity of a model-built triplex-DNA structure. A low root mean square deviation of the dynamic structure from the initial structure demonstrates the stability of the triplex in the salt solution. The sugar pseudorotation, the backbone conformations, and the average helical parameters suggest that the conformation of strands I and III is strictly neither A-form nor B-form, whereas the conformation of strand II remains near the A-form. A higher mobility of both the cytosine strand and the triplex-forming guanine strand and also a longer residence time of water molecules in the spine of hydration were observed and are consistent with available NMR results.

1. Introduction

Under certain conditions, a third polynucleotide strand can associate with double-stranded DNA to form a three-stranded helical structure referred to as a DNA triplex.[1] These triple-helical nucleic acids have provided interesting models for alternative DNA structures and base-pair hydrogen-bonding patterns.[2,3] Theoretical and experimental studies of triplex DNA have also been stimulated by possible applications in the control of gene expression through binding of either a single strand to duplex DNA (the antigen concept)[3-6] or two strands to RNA (an extension to the antisense concept).[7-9] A substantial amount of chemical and physical evidence shows that the nucleotide bases of the third strand occupy the major groove of the target duplex and form specific hydrogen bonds with bases of the Watson–Crick duplex.[1,10-13] Depending on the orientation of the third strand, two major classes of triplexes can be identified, namely, Hoogsteen and reverse-Hoogsteen, where the third strand runs parallel or antiparallel to the purine-rich strand of the Watson–Crick (WC) duplex, respectively.[14] Even though early studies suggested that the formation of stable triplexes was limited to binding of a homonucleotide strand to the polypurine strand of a WC duplex,[1,15,16] more recently many different novel modifications of the triplex-forming oligonucleotide (TFO) have been tested in an effort to enhance the triplex-formation propensity for heterogenous DNA-duplex target sequences.[17-21] Although there have been significant developments in this area involving synthetic modifications, more understanding of the structural aspects is required in order to interpret and rationalize the experimental findings. As a first step, the determining factors responsible for the stability of homopolymeric triplexes need to be determined before the design of modified TFOs targeting heterogenous sequences is possible.

The formation and stability of a triplex are known to be dependent on the salt concentration of the solution.[22-24] For the stability of triple-helical DNA either high concentrations of monovalent ions or millimolar concentrations of divalent cation salts are often required. A reasonable initial structure for a triplex is needed to investigate the conditions for the stability of triplexes. The partial low-resolution structures of several triple helices have been determined by X-ray fiber diffraction,[25,26] and segments of triple helices have been

[⊗] Abstract published in *Advance ACS Abstracts*, February 1, 1995.

(1) Felsenfeld, G.; Davies, D. R.; Rice, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.
(2) Cheng, Y.-K.; Pettitt, B. M. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225–257.
(3) Sun, J. S.; Hélène, C. *Curr. Opin. Struct. Biol.* **1993**, *3*, 345–356.
(4) Wilson, W. D.; Tanious, F. A.; Mizan, S.; Yao, S.; Kiselyov, A. S.; Zon, G.; Strekowski, L. *Biochemistry* **1993**, *32*, 10614–10621.
(5) Hélène, C.; Toulme, J. J. *Biochem. Biophys. Acta* **1990**, *1049*, 99–125.
(6) Letai, A. G.; Palladino, M. A.; Fromm, E.; Rizzo, V.; Fresco, J. R. *Biochemistry* **1988**, *27*, 9108–9112.
(7) Giovannageli, C.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10013–10017.
(8) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 543–584.
(9) Tśo, P. O. P.; Aurelian, L.; Cheng, E.; Miller, P. S. *Ann. N.Y. Acad. Sci.* **1992**, *660*, 159–177.
(10) de los Santos, C.; Rosen, M.; Postel, G. *Biochemistry* **1989**, *28*, 7282–7289.
(11) Johnston, B. H. *Science* **1988**, *241*, 1800–1804.
(12) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, 37, 63–80.
(13) Rajagopal, P.; Feigon, J. *Nature* **1989**, *239*, 637–640.
(14) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822–823; **1963**, *16*, 907–916.
(15) Mosen, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.
(16) Hanvey, J. C.; Shimizu, M.; Wells, R. D. *J. Biol. Chem.* **1989**, *264*, 5950–5056.
(17) Radding, C. M. *Ann. Rev. Biochem.* **1978**, *47*, 847–880.
(18) Xodo, L. E.; Manzini, G.; Quadrifoglio, F. *Nucleic Acids Res.* **1990**, *18*, 3557–3564.
(19) Distefano, M. D.; Shin, J. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1991**, *113*, 5901–5902.
(20) Mergny, J. L.; Sun, J. S.; Rougée, M.; Montenary-Garestier, T.; Barcelo, F.; Chomillier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791–9798.
(21) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800–2807.
(22) Anderson, C. F.; Record, M. T., Jr. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 423–465.
(23) Mills, P. A.; Rashid, A.; James, T. L. *Biopolymers* **1992**, *32*, 1491–1501.
(24) Reddy, M. R.; Rossky, P. J.; Murthy, C. S. *J. Phys. Chem.* **1987**, *91*, 4923–4933.

0002-7863/95/1517-2147$09.00/0

observed in tRNA X-ray crystal structures at higher resolution.[27,28] However, with this limited data it is difficult to produce an accurate atomic model for a triplex, such as that required in molecular dynamics (MD) simulations. Therefore, in the present study, we have used a structure consisting of seven base planes arranged as determined by a previous model-built triplex.[2] The MD simulation method has been used successfully in the studies of many biological systems, including proteins[29] and duplex DNA[30] in aqueous solutions. In principle, the MD simulation should be able to provide evidence regarding the quality of the initial structure of the triplex, as well as providing a structural and dynamical description of the behavior of the triplex DNA in the salt solution.

Here we report the results of a 1.3 ns long molecular dynamics trajectory of a $d(CG\cdot G)_7$ triplex in 1 M aqueous NaCl solution. This triplex simulation is with explicitly unconstrained counterions and without any assumption concerning their initial preferred positions in the system. We anticipate that solvation and specific cation coordination are important in differentiating the stabilities of various possible conformations. Furthermore, by analyzing the MD trajectory, we are able to follow the dynamics of the furanose pucker and backbone conformation, which are also of recent experimental interest.[4,6,31]

The layout of this paper is as follows. In section 2 we present the method of simulation along with a description of the triplex-DNA model and the solvent environment. In section 3 we give the results of the MD simulation of triplex DNA. This section is divided into two parts. The first part describes the structure and the energetics of the DNA, while the second describes the behavior of the solvent and ions around the triplex. Section 4 gives the discussion of the results from the simulation, and section 5 presents the conclusions drawn from the present work, together with the possible direction of future work in the area of heteropolymer triplexes.

2. Methods

Model-built structures from other work in this laboratory have been utilized here. Briefly, the initial structure of the $d(CG\cdot G)_7$ triple helix was derived from the atomic coordinates of Arnott's X-ray fiber diffraction data for UAU[32] as described previously.[33,34] The first and second strands form a duplex-DNA helix, $(CG)_7$, with Watson–Crick[35] base pairing. The third strand has the reverse-Hoogsteen base-pairing scheme with the duplex DNA and was oriented antiparallel to the second strand. Thermodynamic conditions, base composition, and base order apparently are all involved in determining the orientation of the third strand.[2,34] The initial DNA 7-mer triplex was built with a helical rise of 3.04 Å and a helical twist of 32.7°, and with sugar puckers initially placed in the A-form. The charge on the DNA solute was $-e$ per phosphate group. The solute was placed in a cubic volume of length 32.0 Å with the helical axis parallel to the z-direction. Accounting for the volume of the DNA, 890 water molecules were needed to maintain a water density of 1 g/mL in the simulation box near the desired pressure.[36] Twenty-one random water molecules were replaced with Na^+ ions to obtain electroneutrality of the system, and an additional 16 Na^+ and 16 Cl^- ions were added to bring the system to approximately 1 M in salt concentration by randomly replacing another 32 water molecules. The final system then consisted of the 7-mer triplex DNA, 837 water molecules, 37 Na^+, and 16 Cl^- ions.[36]

An all-atom force field[37] (parameters from CHARMM 22 release) was used in this simulation. The three-site simple point charge (SPC/E) model was used for water–water interaction potential,[38] and the Lennard-Jones parameters for Na^+ and Cl^- ions were taken from Chandrasekhar *et al.*[39] We simulated the system in the microcanonical [N, V, E] ensemble. The initial velocities were taken from a Maxwell–Boltzmann distribution at an initial temperature of 300 K. Periodic boundary conditions were imposed in all directions. All electrostatic interactions were calculated using the Ewald summation scheme[40,41] as opposed to using a long-range cutoff. The velocity version of the Verlet algorithm[42,43] was employed to integrate the equations of motion with a 2 fs time step and SHAKE bond constraints.[44]

Ten steps of steepest descent minimization was used to remove the strain in the initial structure prior to the MD trajectory. During the initial equilibration part of the trajectory, the DNA was kept rigid while the solvent molecules and counterions were equilibrated in the vicinity of the DNA. This 100 ps long simulation achieved equilibration of saltwater solution around the rigid DNA, and the results have been discussed elsewhere.[36] After the equilibration period, the solvent molecules and ions were fixed at their respective positions, and the DNA was allowed to relax for 1 ps. Then the motion of DNA was constrained, and solvent molecules and ions were allowed to move for another picosecond. This process was continued, alternatively clamping and unclamping the DNA, for total of a 40 ps, increasing the length of the simulation time of the alternative segments toward the end. All the particles were then allowed to move for 10 ps, during which the velocities were scaled to maintain a temperature near 300 K. The simulation was continued for another 20 ps without any constraints or velocity scaling. At the end of this segment we observed that the total energy, the temperature, and the pressure of the system were stable. From this point (a total of 170 ps) we started our production run; a 1.155 ns long trajectory of homopolymer $(CG\cdot G)_7$ triple-helical DNA in saltwater solution. Positions and velocities of all the atoms, together with the dihedral angles of the triplex, were saved at every 0.1 ps for analysis.

(25) Arnott, S.; Selsing, S. *J. Mol. Biol.* **1974**, *88*, 509–521.
(26) Arnott, S.; Bond, P. J.; Selsing, B. E.; Smith, P. J. C. *Nucleic Acids Res.* **1976**, *3*, 2459–2470.
(27) Kim, S. H.; Suddath, F. L.; Quigley, G. J.; McPharson, A.; Sussman, J. L.; Wang, A. J. H.; Seeman, N. C.; Rich, A. *Science* **1974**, *185*, 435–440.
(28) Robertus, J. D.; Ladner, J. E.; Finch, J. T.; Rhodes, D.; Brown, R. S.; Clark, B. F. C.; Klug, A. *Nature* **1974**, *250*, 546–551.
(29) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*, 1st ed.; Cambridge University Press: Cambridge, 1987.
(30) Swaminathan, S.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5027–5040.
(31) Ouali, M.; Letellier, R.; Sun, J.-S.; Akhebat, A.; Adent, F.; Liquier, J.; Taillandier, E. *J. Am. Chem. Soc.* **1993**, *115*, 4264–4270.
(32) Laughton, C. A.; Neidle, S. *J. Mol. Biol.* **1992**, *223*, 519–529.
(33) Van Vlijmen, H. W. T.; Ramé, G. L.; Pettitt, B. M. *Biopolymers* **1990**, *30*, 517–532.
(34) Cheng, Y.-K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465–4474.
(35) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738.
(36) Mohan, V.; Smith, P. E.; Pettitt, B. M. *J. Phys. Chem.* **1993**, *97*, 12984–12990.
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–198.
(38) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6274.
(39) Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903–910.
(40) Ewald, P. *Ann. Phys. (Leipzig)* **1921**, *64*, 253–264.
(41) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London* **1980**, *373A*, 27–39.
(42) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 1st ed.; Oxford University: New York, 1987.
(43) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.
(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–349.

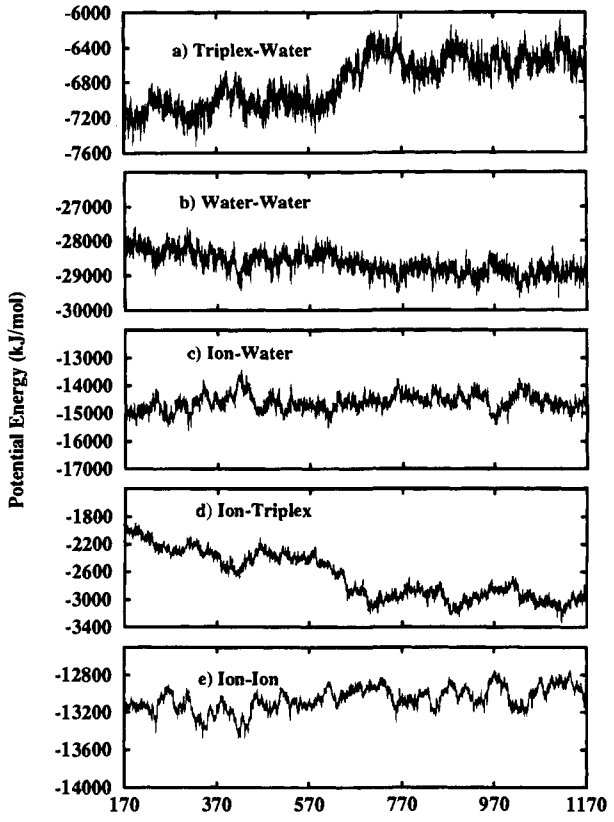

Figure 1. Time histories of the (a) triplex–water, (b) water–water, (c) ion–water, (d) ion–triplex, and (e) ion–ion interaction energies.

The moment of inertia tensor of the triplex was calculated and diagonalized for each configuration. Then the directional vector of the smallest principle moment of inertia was taken as the instantaneous global helical axis. The direction of the 3′-end of strand I was chosen as the positive direction of the global helical axis for our analysis.

3. Results

3.1. Structure and Stability of the Triplex. During the first 100 ps with the DNA rigid, it was previously observed that the solvent molecules and ions appeared to obtain a stable distribution around the triplex.[36] Motion of the DNA allowed a further relaxation of this distribution, which proceeded to establish a new equilibrium. In Figure 1, we give the potential energies for the DNA–water, water–water, ion–water, ion–DNA, and ion–ion interactions over the 1.155 ns simulation (production) time. Several features describe the path taken by the system to equilibrium. Figure 1a shows an increase in the DNA–water potential energy between 570 and 670 ps. During the same time period, there is a decrease in the interaction between the ions and the DNA, as given in Figure 1d. We have observed, by examining configurations of the system during this time interval, that six water molecules left the vicinity of the DNA as four new Na^+ ions came to within 2.5 Å of various atoms of the DNA. The potential energy decrease in the water–water interaction (Figure 1b) during the same interval corroborates the replacement of water bound to DNA by Na^+ ions. Overall, the increase in the potential energy of the DNA–water interaction was compensated for by a decrease in the potential energy of the water–water and the DNA–ion interactions.

The root mean square deviation of the triplex with respect to the initial structure was calculated, and the results are summarized in Table 1. The results were split into two categories. The first considered all the base planes of the triplex while the second neglected the two base planes at both ends of the triplex in an effort to quantify end effects. The data presented in Table 1 indicate that a significant amount of the total rms deviation originates from the base planes of both ends of the triplex. Therefore it is more relevant to consider the properties of the central part of the structure instead of considering the full triplex, to study and predict the structural and dynamical properties of a longer triple helix.

Table 1. Root Mean Square Deviation from the Initial Structure for Selected Groups of Atoms in the Triplex

	rms deviation (Å)	
group	full triplex	middle part
strand I	1.88 ± 0.36	1.74 ± 0.36
strand II	1.45 ± 0.11	1.15 ± 0.14
strand III	2.24 ± 0.52	1.76 ± 0.17
base	1.23 ± 0.21	1.00 ± 0.11
sugar	1.95 ± 0.39	1.69 ± 0.21
phosphate	2.90 ± 0.44	2.38 ± 0.32
all	1.89 ± 0.33	1.58 ± 0.19

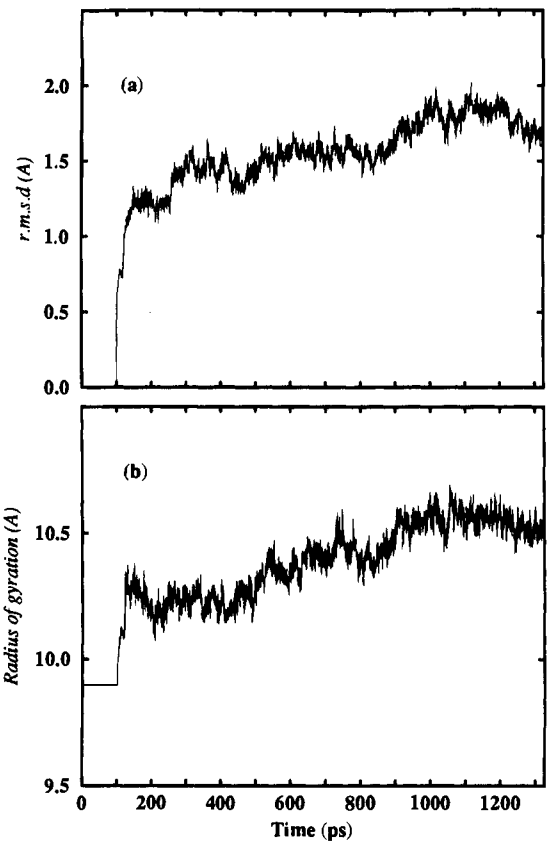

Figure 2. A plot of (a) the rms deviation and (b) the radius of gyration of the triplex as a function of the simulation time.

Strand II displays the least deviation, in part because it is sandwiched between the other two strands which show about the same rms deviation (see the last column of Table 1). The bases show the least mobility while the phosphate backbone displays the largest rms deviation. Laughton *et al.* have reported a similar (lower than 2.0 Å) rms deviation of the time-averaged MD structure of a low-pH DNA triple helix, $d(TC)_5d(GA)_5d(C^+T)_5$, from their initial model.[32] They have also reported that the phosphate group possessed the largest rms deviation.

The overall rms deviation of the central part of the triplex as a function of time is given in Figure 2a, and the radius of gyration of the DNA is presented in Figure 2b. Both the rms deviation and the radius of gyration indicate that the triplex has passed through three distinct structural stages over the simulation

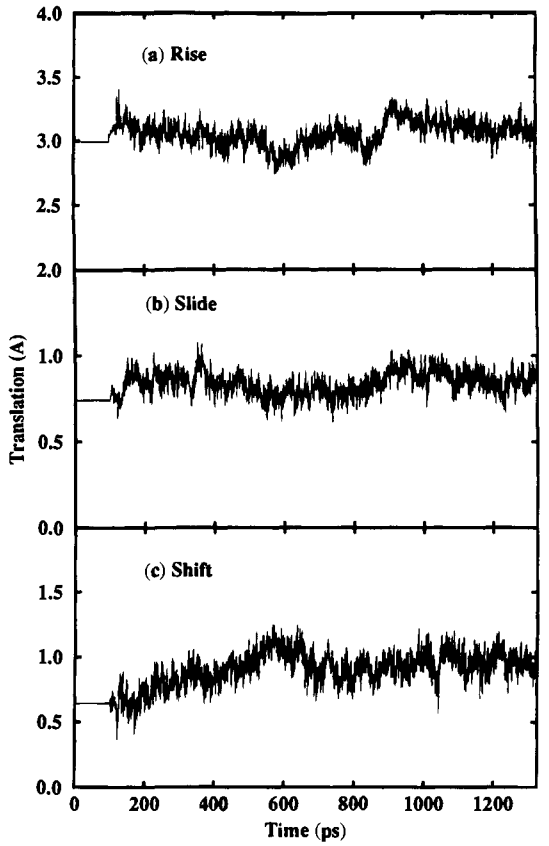

Figure 3. A plot of the average (a) helical rise, (b) slide, and (c) shift (in Å) between two adjacent base planes as a function of the simulation time.

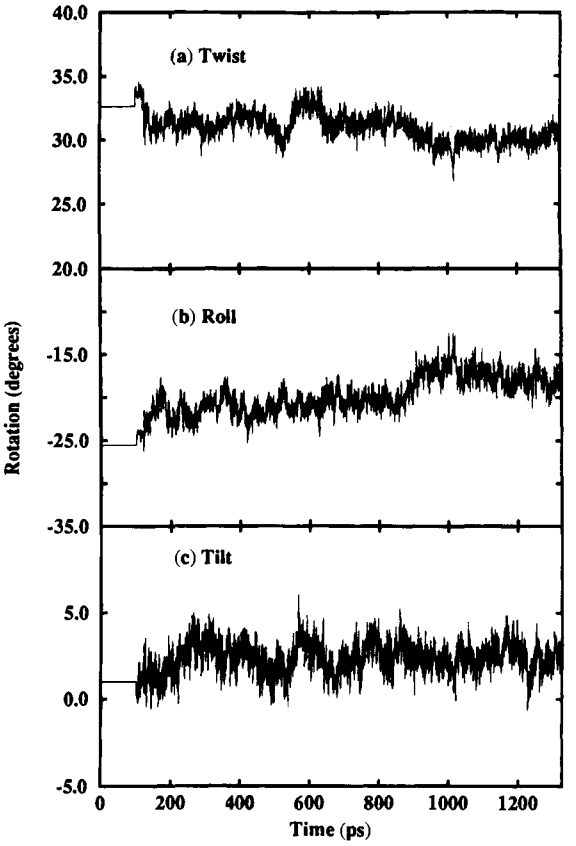

Figure 4. A plot of the average (a) twist, (b) roll, and (c) tilt (in degrees) of two adjacent base planes as a function of the simulation time.

time reported here. As we have mentioned earlier, during the first 100 ps only the solvent and ions in the system had the freedom to move, giving zero rms deviation and a constant radius of gyration. Most of the deviation from the model-built structure occurred during the 70 ps equilibration of the triplex after 100 ps of DNA rigid simulation. The average rms deviation of the DNA from the model over the last 400 ps of the trajectory, 1.71 ± 0.19 Å, is reasonably small and illustrates the stability of the triplex in the salt solution. The deviation measured from the structure obtained after the equilibration is over an angstrom smaller.

The helical parameters associated with the triplex as a function of time are a good measure of the structure and the stability of the DNA in saltwater solution. Figure 3 gives the variation of the three translational parameters, rise, slide, and shift, and Figure 4 presents the three rotational parameters, twist, roll, and tilt, of the triplex as a function of the simulation time. Appendix A describes the procedure used for the calculation of the helical parameters considered in this study. Only the central five base planes were taken into account to calculate the average helical parameters given in Figures 3 and 4. The initial helical rise given in the figure is less than 3.04 Å, the input parameter for the initial model. This discrepancy between the initial helical rise seen in Figure 3a and the input parameter is due to the steepest descent minimization of the initial system before the start of the MD trajectory.

The time-averaged value for the helical rise, 3.05 ± 0.10 Å, is close to the initial model-built rise before minimization and indicates that the DNA has some features in common with an A-DNA structure. Average structure details given in Table 2 do show some differences in backbone conformation from duplex DNA crystal structure.[45] The other translational parameters, the relative slide and the relative shift, were also given in Figure 3, panels b and c, respectively. The average slide, 0.84 ± 0.07 Å, is about the same as its initial value, while the shift has changed from its initial value of 0.68 Å to a different equilibrium value with an average of 0.92 ± 0.12 Å. The average twist angle as a function of time given in Figure 4a indicates that this triplex DNA has a structure with a twist intermediate between that of the A-form and that of the B-form. The average twist angle over the total simulation time is 30.85 ± 1.06°, but the final value is closer to 30°. The other relative rotational parameters, roll and tilt, are given in Figure 4b,c. The average roll and tilt are −19.66 ± 2.12° and 2.41 ± 0.93°, respectively. A sudden increase and subsequent decrease in twist and tilt at around 600 ps is due to a fluctuation in the hydrogen bonds between cytosine and guanine of strands I and II in the second base-triplet plane. We will return to this point later.

The average helical rise of 3.05 Å and the average helical twist of 30.85° yield an average of 11.7 base planes per turn, and hence a 35.7 Å full-turn height for CG·G triplex DNA. These results are in the general range expected from other MD results of different triplex[32,46] and duplex[30,47,48] systems.

The interbase hydrogen bonds are important for recognition and stability of the triplex. There are five possible hydrogen bonds between three bases in a reverse-Hoogsteen antiparallel CG·G base plane, ignoring out of plane geometries. Three of the H-bonds form between the CG Watson−Crick pair, and the

(45) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.
(46) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Colvin, O. M.; T'so, P. O. P. *Anti-Cancer Drug Des.* **1990**, *5*, 159−167.
(47) Shibata, M.; Zielinski, T. J.; Rein, R. *Biopolymers* **1991**, *31*, 211−232.
(48) Miaskiewicz, K.; Osman, R.; Weinstein, H. *J. Am. Chem. Soc.* **1993**, *115*, 1526−1537.

Table 2. Average Dihedral Angles (and Fluctuations) of the Phosphate Backbone and Sugar Pucker of Individual Residues in the Triplex

	dihedrals (deg)						P
	$\alpha(2)^a$	$\beta(0)^a$	$\gamma(3)^a$	$\delta(2)^a$	$\epsilon(0)^a$	$\zeta(2)^a$	(deg)
strand I							
initial dihedral	−64	174	55	85	−157	−75	
CYT2	−79(13)	164(14)	57(10)	110(24)	−160(10)	−85(14)	124(56)
CYT3	−71(10)	171(11)	60(10)	106(21)	−158(10)	−83(12)	116(42)
CYT4	17(27)	171(10)	57(9)	97(18)	−160(10)	−82(15)	84(47)
CYT5	−74(10)	168(10)	58(9)	89(18)	−158(10)	−73(11)	49(50)
CYT6	−70(10)	172(9)	60(9)	87(15)	−156(10)	−74(10)	44(44)
strand II							
initial dihedral	−62	165	80	88	−150	−75	
GUA9	−79(13)	170(8)	61(12)	90(11)	−162(10)	−78(10)	87(21)
GUA10	−76(10)	170(8)	62(10)	80(10)	−158(10)	−74(10)	55(25)
GUA11	−73(10)	166(8)	66(8)	81(10)	−160(10)	−64(10)	18(15)
GUA12	−78(10)	173(8)	67(9)	84(10)	−152(10)	−62(11)	11(13)
GUA13	−74(10)	173(9)	54(10)	75(10)	−151(10)	−66(10)	17(10)
strand III							
initial dihedral	−114	130	58	74	47	107	
GUA16	−116(14)	148(15)	55(10)	86(10)	−136(53)	−88(49)	10(14)
GUA17	−73(18)	167(16)	65(9)	125(30)	−57(65)	144(42)	148(56)
GUA18	−86(25)	133(17)	54(13)	85(10)	−148(54)	−94(43)	23(27)
GUA19	−71(16)	163(15)	62(9)	76(10)	60(14)	98(17)	5(28)
GUA20	−118(18)	150(19)	50(14)	105(32)	−115(73)	−154(62)	83(75)

[a] For each residue, n is given in parentheses: n = the multiplicity of the intrinsic dihedral angle potential; $n = 0$ implies no intrinsic dihedral barrier, $n = 2$ gives minima at 90° and −90°, while $n = 3$ gives minima at 60°, −180°, and −60°.

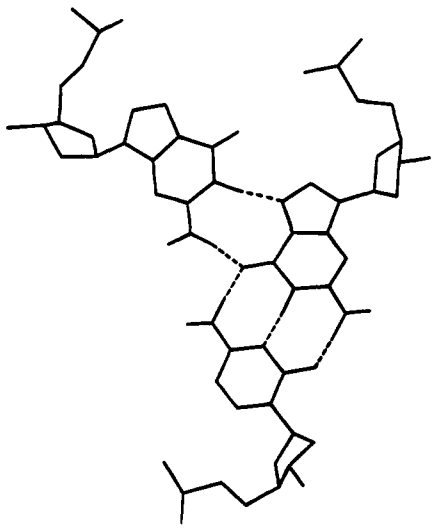

Figure 5. CG·G base-triplet plane with the possible, interbase H-bonds. The three hydrogen bonds on the lower right are between the CG Watson–Crick pair, and the upper-left two H-bonds are between strand II and the triplex-forming third strand.

other two are between the triplex-forming strand III and strand II. Following a familiar notation,[45] the first three H-bonds between Watson–Crick pairs were denoted as G_{II}N2:C_IO2, G_{II}N1:C_IN3, and C_IN4:G_{II}O6, and the two reverse-Hoogsteen H-bonds between strands II and III were denoted as G_{III}N2:G_{II}O6 and G_{III}N1:G_{II}N7. The subscripts indicate the strand numbers. Figure 5 shows a sketch of the five possible H-bonds. The distances between the heavy atoms associated with these H-bonds were calculated as a function of the simulation time. The average distance of a H-bond was 2.99 ± 0.20 Å. The average H-bond distance for a CG pair was 2.98 ± 0.19 Å, while that for a GG pair was 2.99 ± 0.20 Å. (The first and last base-triplet planes were not included in the average.)

Except for the C_IN4:G_{II}O6 H-bond in the second base-triplet plane, all the monitored distances indicate stable H-bonds with small fluctuations. Figure 6 gives the time history of the five H-bonds in the second base-triplet plane. In the first three panels of the figure we display the H-bond distances between the Watson–Crick pair, while the last two panels show the distances

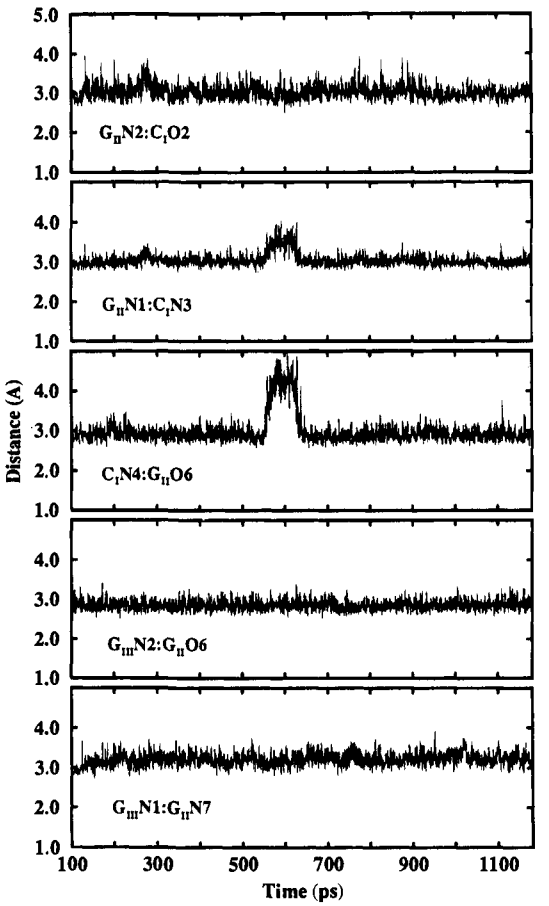

Figure 6. The time history of the five possible H-bond distances of the second base-triplet plane (see the text).

of the H-bonds formed between strands II and III. As given in the third panel of Figure 5, the C_IN4:G_{II}O6 H-bond was broken for about 100 ps, even though the two guanine bases remained in contact throughout the simulation time. As we have reported earlier,[49] there is a spine of water formed between the cytosine

(49) Mohan, V.; Smith, P. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1993**, *115*, 9297–9298.

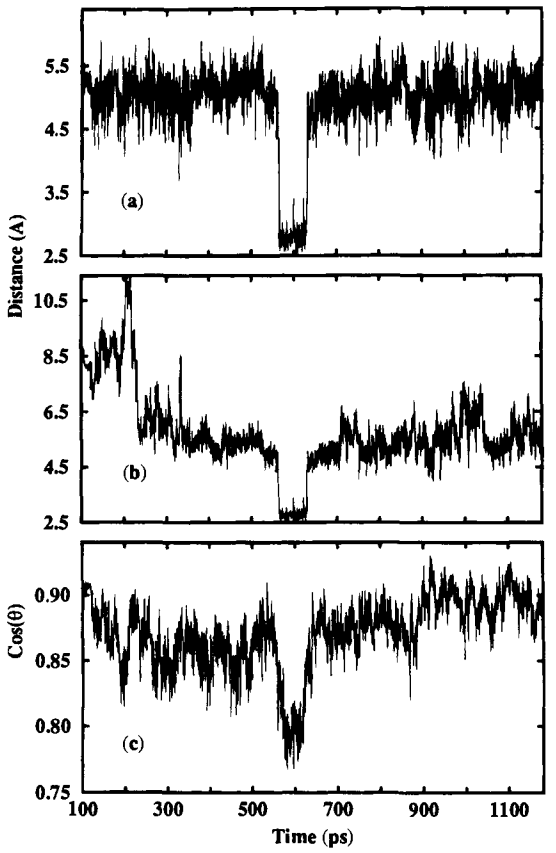

Figure 7. The time history of (a) the distance between the O6 atom of G_{II} in the second base-triplet plane and the oxygen atom of the closest water molecule, (b) the distance between the O6 atom of G_{II} in the second base-triplet plane and the oxygen atom of the water molecule involved in the H-bond breaking, and (c) the cosine of the angle between two normals of the C_I and G_{II} in the second base-triplet plane.

strand and the triplex-forming third (guanine) strand. A water molecule close to the NH_2 of the cytosine located in the second base-triplet plane moved toward the O6 of the guanine base in strand II, causing the cytosine base to move away temporarily. In Figure 7a we show the time history of the distance between the O6 of the second base plane guanine of strand II and the oxygen atom of the water molecule closest to it. This indicates that there is always a water molecule within a 6.0 Å range from the O6 of the guanine and that, during the breakdown of the H-bond, a water molecule moved closer to the O6 oxygen. Since the solvent and the ions in the system are in a dynamical equilibrium with the DNA, this type of temporary exchange of water molecules from the spine of hydration along the groove between strands and the water and ions in the bulk might be expected. In Figure 7b we present the time history of the distance between the O6 of the guanine and the oxygen atom of the water molecule which is involved the breaking of the H-bond between cytosine and guanine in the second base-triplet plane. Together, panels a and b of Figure 7 further illustrate the equilibrium between the solvent and the solute. Figure 7c shows the cosine of the angle between the normals of the cytosine and guanine planes, indicating that the two bases were essentially coplanar for about 500 ps. When the water molecule moved closer to the O6 atom in guanine base, the relative angle between two bases increased to about 40°. The last two panels of Figure 6 clearly indicate that the two guanine bases in the base-triplet plane maintained intact H-bonds during the increase of the relative angle between cytosine and guanine. In this case, only the cytosine base tilted when the water molecule hydrogen bonded with the O6 atom of the guanine base. This is confirmed by determination of the angles between the global helical axis and the normals to the two base planes (data not shown). Even though the insertion of water into the helix caused a temporary breakage of H-bonds within the base plane, the helical structure of the triplex was preserved. To our knowledge, the transient insertion of water between bases of a DNA-base plane has not been observed in any other MD simulation study on duplex or triplex DNA, although there is theoretical evidence for the insertion of water into networks of amino acid hydrogen bonds.[50]

The variation of the torsional angles of the phosphate backbone, $P^{\alpha}-O5^{\beta}-C5^{\gamma}-C4^{\delta}-C3^{\epsilon}-O3^{\zeta}-P$, also provides useful information on the mobility of the triplex DNA. The torsional angle, δ, correlates strongly with the sugar puckering conformation. Average torsional angles of the phosphate backbone, along with the sugar pucker angles, are presented in Table 2 (the results for terminal residues are not given). The β, γ, and δ dihedrals remained close to their initial values and displayed small fluctuations. The α dihedrals of the first and second strands also remained close to their initial values, with the exception of CYT4, which underwent several transitions between two different states, one at $\simeq -30°$ and one at $\approx 30°$. This suggests that this dihedral has an equilibrium between distinct conformations, and the barriers between conformations are low enough to permit transitions on the time scale of hundreds of picoseconds although the equilibrium constant always favors the conformation at $-30°$. The ϵ and ζ dihedrals of the third strand deviated significantly from the initial values. Interestingly, this was achieved by cooperative rotations around the ϵ and ζ, which retained the overall integrity of the triplex. The deviation of these angles from their initial model-built values and the resulting equilibrium are probably results of the inclusion of explicit solvent, which can strongly affect the relative free energies (population) of the different conformational states.[51] Large standard fluctuations (given in parentheses) for pseudorotation and the δ values of residues indicate the higher mobility of the cytosine and TFO strands. The smaller fluctuations of the δ and P values of the residues in strand II indicate a less mobile structure with a broadly defined A-DNA-like conformation. Agreement between the reported X-ray fiber diffraction results for the backbone torsional angles of the third strand of triplex DNA[52] and the results presented in Table 2, with the exception of ϵ and ζ, suggest that our initial model of triplex DNA is reasonable and that small structural corrections are accessible via simulation.

The average furanose pseudorotation, on a per strand basis, is given in Table 3. These results indicate the change in conformation of the cytosine strand between C3′-endo and C2′-endo. Strand II adopts a C3′-endo (A-form) conformation, with strand III also preferring a C3′-endo form over C2′-endo. The larger mobility of the cytosine strand in the d(CG·G) triplex is also observed by NMR[53] and has been discussed in more detail elsewhere.[54] It should be noted that some spectroscopic data has been interpreted to indicate the absence of such an equilibration between pucker states[55,56] while other data has been interpreted to indicate the opposite.[53,54]

(50) DiCapua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 6145−6155.
(51) Smith, P. E; Pettitt, B. M. *J. Phys. Chem.* **1994**, *98*, 9700−9711.
(52) Arnott, S.; Bond, P. J.; Selsing, B. E.; Smith, P. J. C. *Nucleic Acids Res.* **1976**, *10*, 2759−2770.
(53) Dittrich, K.; Gu, J.; Tinder, R.; Hogan, M.; Gao, X. *Biochemistry* **1994**, *33*, 4111−4120.
(54) Cheng, Y.-K.; Weerasinghe, S.; Mohan, V.; Smith, P. E.; Ramé, G. L.; Pettitt, B. M.; Johnson, K. H.; Dittrich, K. A.; Tinder, R.; Hogan, M. E. *J. Am. Chem. Soc.*, submitted for publication.
(55) Radhakrishnan, I.; Patel, D. J. *Structure* **1993**, *1*, 135−152.
(56) Liquier, J.; Coffinier, P.; Firon, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1991**, *9*, 437−445.

Table 3. Pseudorotation (P) of Furanose[a]

strand	form	average	%
I	A	18.1 ± 10	34.1
	B	161.4 ± 11	21.1
	O4′	92.1 ± 10	11.9
	other	131.2 ± 87	32.8
II	A	15.1 ± 10	53.4
	B	162.5 ± 12	4.7
	O4′	88.6 ± 15	14.5
	other	142.2 ± 99	27.4
III	A	13.9 ± 10	44.0
	B	169.4 ± 11	18.1
	O4′	86.0 ± 10	0.8
	other	272.8 ± 99	37.1

[a] $\nu_0 \rightarrow C_{4'}-O_{4'}-C_{1'}-C_{2'}$, $\nu_1 \rightarrow O_{4'}-C_{1'}-C_{2'}-C_{3'}$, $\nu_2 \rightarrow C_{1'}-C_{2'}-C_{3'}-C_{4'}$, $\nu_3 \rightarrow C_{2'}-C_{3'}-C_{4'}-O_{4'}$, $\nu_4 \rightarrow C_{3'}-C_{4'}-O_{4'}-C_{1'}$, and $\tan P = [(\nu_4 + \nu_1) - (\nu_3 + \nu_0)]/[2\nu_2(\sin 36° + \sin 72°)]$.

Table 4. Locations and Magnitudes of the Extrema in the Pair Correlation Functions and Coordination Numbers for Water Pairs and Water–Ion Pairs

	first peak			
$g(r)$	height (Å)	max (Å)	min (Å)	coord no.
O–O	3.24	2.78	3.38	4.2
O–H	1.69	1.78	2.38	1.5
H–H	1.52	2.38	2.98	4.9
Na^+–O	9.67	2.38	3.12	5.1
Na^+–H	3.38	2.99	3.63	12.7
Cl^-–O	4.61	3.28	3.98	7.5
Cl^-–H	4.21	2.29	3.06	6.9

The helical parameters of the triplex, hydrogen bonds between strands and the pseudorotation of furanose in the backbone do not show drastic changes after equilibration of the system. Next we discuss the behavior of solvent and salt ions around the triplex DNA.

3.2. Behavior of the Solvent and Counterions. The arrangement of the solvent molecules and counterions around the DNA solute depends sensitively on the system details and the composition of the overall solution environment. The equilibrium structure and the diffusional properties of solvent and ions may be strongly influenced by the presence of a solute. We have calculated the radial distribution functions for water oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen pairs, as well as for Na^+–water and Cl^-–water pairs. The results are summarized in Table 4. These results are in reasonable agreement with previously reported data,[57-59] and the water–water results show little change at this concentration. The average number of water oxygens around a Na^+ ion is 5.0, which is to be compared to a value of 6.0 for an isolated sodium ion as reported by Impey *et al.*[60] The large number of interactions between Na^+ ions and the 21 negative sites on the DNA solute decrease the coordination number for Na^+ ions. The coordination number of Cl^- ions is far less sensitive and shows good agreement with reported simulation results not containing DNA.[60]

Thus the size, shape, and charge of the triplex DNA has an influence on the distribution of many components in the solution. Unfortunately, the traditional spherical radial distribution function is not particularly informative in this situation. Hence, to further study the solvation of DNA in the saltwater solution,

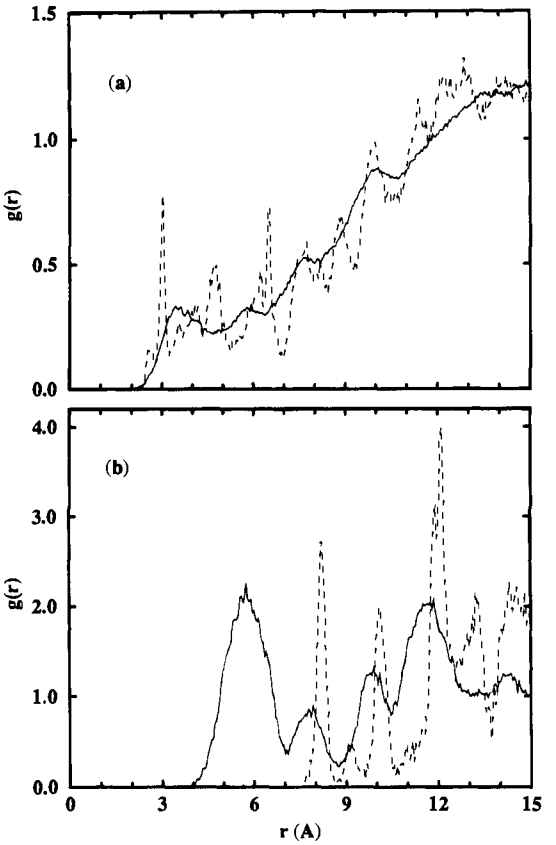

Figure 8. Cylindrical distribution of (a) water oxygen (b) Na^+ ions with respect to the global helical axis (solid line, this study; dashed line, rigid DNA simulation (ref 36)).

we have calculated the distribution of the water and the ions radiating perpendicularly from the helical axis. We first define the instantaneous global helical axis as described in the methods section. Then we define a cylindrical coordinate system with the center coincident with the center-of-mass of the triplex DNA and the major axis coincident with the helical axis. Since the average length of the triplex is only 18.3 Å, and to avoid end effects, we have considered a cylinder with a length of 17.0 Å around the instantaneous helical axis for the evaluation of the cylindrical distribution functions. However, it should be noted that, in general, the integration of the cylindrical distribution function will not give the total number of water molecules and ions in the system. In Figure 8a, we present the cylindrical radial distribution of water oxygens with respect to the helical axis. Results for rigid DNA (dashed) and for flexible DNA (solid) are shown. Motion of the DNA tends to broaden the observed features in the distribution functions. The first peak at around 3.3 Å in Figure 8a is due to the spine of water molecules bound to the triplex and is similar in physical origin to that reported elsewhere for the rigid DNA simulation.[36,49] The second broad peak at 5.7 Å is mainly due to the solvent molecules associated with the phosphate and sugar groups. Overall, the distribution of water is very smooth compared to that in the rigid DNA simulation (dashed line). The fluctuations of the triplex appear to have an influence on the ordering of solvent layers around the triplex, although averaging over an order of magnitude longer trajectory may have also provided better statistics for the distribution.

Figure 8b shows the cylindrical distribution of Na^+ ions around the triplex DNA (solid line). The integration of the area up to the end of the second peak ($r = 8.8$ Å) in Figure 8b gives a coordination number of 3.6. This indicates that there are several Na^+ ions within 9.0 Å from the helical axis. It has

(57) Smith, P. E.; Dang, L. X.; Pettitt, B. M. *J. Am. Chem. Soc.* **1991**, *113*, 67–73.
(58) Forester, T. R.; McDonald, I. R. *Mol. Phys.* **1991**, *72*, 643–660.
(59) van Gunsteren, W. F.; Berendsen, H. J. C.; Geursten, R. G.; Swinderman, H. R. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 287–297.
(60) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071–5083.

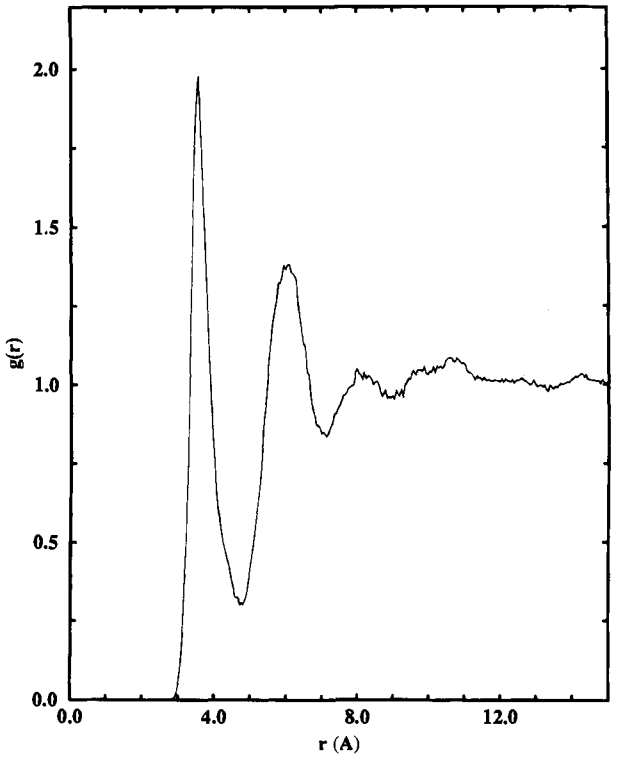

Figure 9. Radial distribution function, $g_{Na^+-Na^+}(r)$.

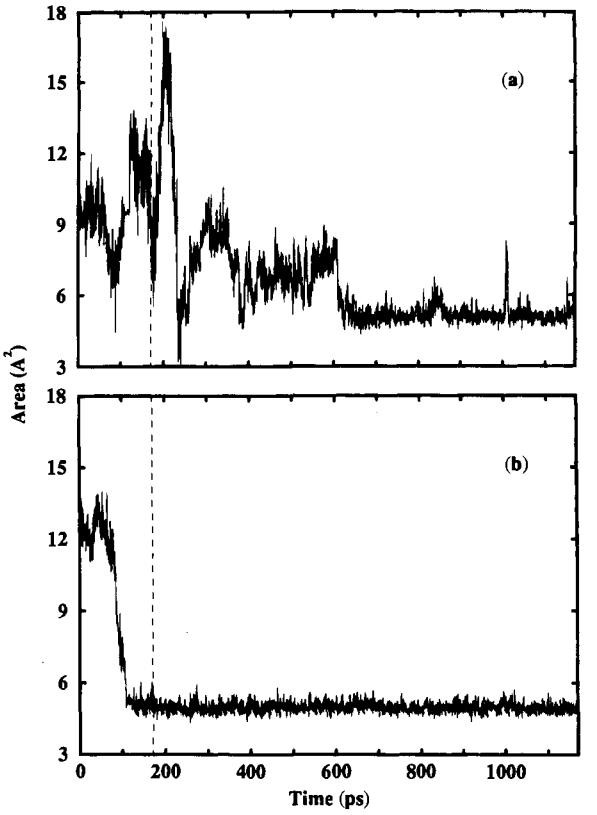

Figure 10. The time history of (a) AS-I and (b) AS-II (see text).

been found that the most favorable positions for Na^+ ions to interact with the triplex are the basic N7 atoms of the guanines in strand III (N7 atoms in strand II are already involved in reverse-Hoogsteen base paring with strand III).[36] A simple distance calculation shows that on average there are about 5 to 6 Na^+ ions associated with the guanine N7s of the entire triplex. The distribution of the Na^+ ions around the rigid DNA (dashed line) is also given. The pronounced peak at 5.5 Å is due to the coordination of Na^+ ions to N7 of guanine in strand III. The comparison of the two distributions indicates that the flexibility of the DNA has a major influence on the movement of the ions, enabling them to diffuse closer to the triplex.

The spherical radial distribution of Na^+ ions around a Na^+ ion is given in Figure 9. The second peak at around 5.8 Å indicates that there are Na^+ ion pairs which interact with each other via their first hydration shell. This feature has been reported previously by Forester *et al.*[58] However the first sharp peak at around 3.6 Å indicates that there were also positive ion pairs in the system which, despite the direct large repulsive interaction between them, were stabilized by bridging solvent molecules or solute atoms. This was also observed in the MD simulation of saltwater solution around rigid triplex DNA,[36] as well as earlier simulation studies of duplex DNA.[58,61] We found Na^+-Na^+ ion pairs in the system in which both ions were coordinated to phosphate groups. One of the pairs was associated with a phosphate group close to the 3′-end of strand I, and another pair was found closer to the middle phosphate of strand II. To examine the details of the coordination, we have monitored the area of the triangle generated by the two Na^+ ions and the P atom in the corresponding phosphate group, rather than monitor three different interatomic distances. We will denote the area corresponding to the ion pair and the P atom in strand I by AS-I, and the area generated by the other pair with the P atom in the middle of strand II by AS-II. Figure 10a gives the area AS-I and Figure 10b shows the area AS-II as a function of the simulation time. The data obtained during the preparation of the system is also given in the graph from 0 to 170 ps (up to the dashed line). The first panel shows that the AS-I triatomic "cluster" formed and dissociated several times up to 600 ps and thereafter maintained a reasonably stable complex. On the other hand, the second ion pair formed a stable "cluster" during the preparation period and then remained intact throughout the rest of the trajectory. We found several short-lived phosphate$-Na_2^+$ complexes (or clusters) during the trajectory.

Negative charges on the solute, especially the N7 atoms in the guanine bases of strand III together with the adjacent two phosphate groups from strands II and III, make a favorable environment for counterions. Figure 11a shows the spherical radial distribution of Na^+ ions around phosphate oxygen atoms of the backbone, and Figure 11b gives the radial distribution of Na^+ ions around N7 atoms in strand III. Calculation of coordination numbers from the distribution curves indicates that there are about 6−7 Na^+ ions close to N7 atoms of strand III and approximately 16 Na^+ ions close to phosphate oxygen atoms. However, calculation also shows that there are about 13 Na^+ ions which remain within a 3.0 Å range of the triplex. Binding of sodium ions to the N7 of guanine was often accompanied by simultaneous binding to a nearby phosphate group. In order to determine a correct estimate for the number of ions coordinated to the N7 base atoms and also to phosphate groups, a triplet distribution function is required. The strong association of some Na^+ ions with the solute also explains the lower coordination number for the Na^+−water oxygen distribution. That is, on average, these ions may be directly associated with phosphate oxygen atoms in the backbone or nitrogen atoms of the bases, and then further surrounded by three to five water molecules to complete their first hydration shell.

To examine the possible existence of Na^+ ions which associated with both N7 atoms and phosphate oxygens, we have generated a conditional triplet probability density[62] of Na^+ ions

(61) Lee, W. K.; Gao, Y.; Prohofsky, E. W. *Biopolymers* **1984**, *23*, 257–270.

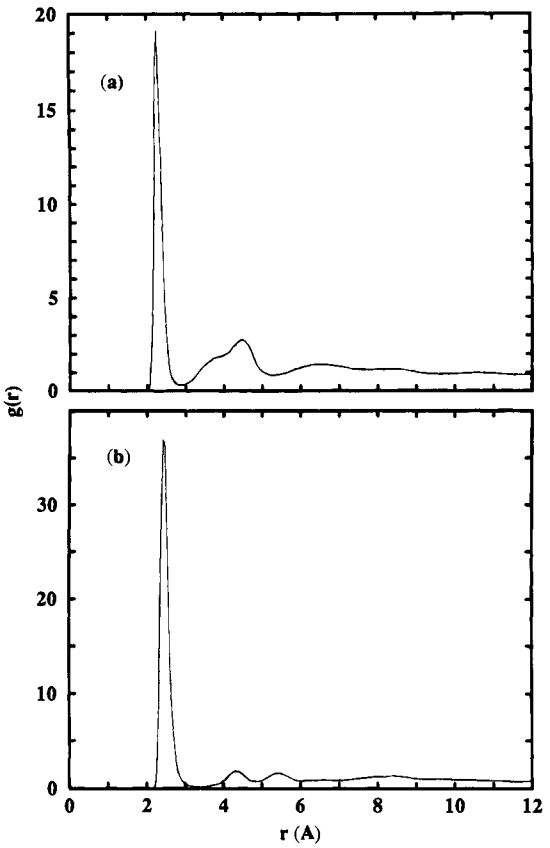

Figure 11. Radial distribution of Na^+ ions around (a) phosphate oxygen atoms and (b) N7 atoms of guanine bases in strand III.

with respect to the N7 atoms of strand III and the phosphate oxygen atoms of the backbone. The calculated conditional distribution of Na^+ ions as a function of N7–Na^+ and OP–Na^+ distances is given in Figure 12. Parts a and b of Figure 12 present the front and rear views of the conditional distribution, respectively. The derivation of the distribution is described in Appendix B. Since we have used all possible pairs of N7 and OP atoms, the distribution in Figure 12 also possesses features related to the helix periodicity. The first sharp peak at (2.8, 2.4) Å is an indication of the association of Na^+ ions with both phosphate oxygens in the backbone and N7 atoms in strand III. The secondary peak along the diagonal (Figure 12b), approximately 2.5 Å away from the first peak, suggests an association of cations with both N7 and OP atoms via a bridging water, a part of the hydration shell of the hexaquo ion complex. Further along the N7–Na distance near the OP contact a pronounced zone of exclusion exists after the contact peak. The rest of the features are small in comparison with those along the OP–Na distance near the N7 contact. This excluded region is due to the tight binding and the steric constraint of base stacking which precludes finding another N7 until the next base plane. The phosphates, by contrast, are more flexible (see Table 1) and do not show such a feature along the OP–Na direction. The reason for a larger ridge at the N7 contact along the OP–Na side is due to having more phosphate oxygens than N7 atoms in the system in a restricted volume.

The description of the structure of the solvent and the salt solution around the triplex DNA presented so far suggests that there may be an impact on the mobility of the solvent molecules and counterions due to the presence of the solute. Strong interactions may greatly reduce the average diffusion coefficient of the solvent and the counterions. The self-diffusion coefficient

(62) Ben-Naim, A. *Water and Aqueous Solutions: Introduction to a molecular theory*, 1st ed.; Plenum Press: New York, 1974.

Table 5. Diffusion Coefficients of Water and Ions in the System

	this study ($\times 10^9$ m^2 s^{-1})	rigid DNA[a] ($\times 10^9$ m^2 s^{-1})	studies without DNA ($\times 10^9$ m^2 s^{-1})
H_2O			
average	1.29	1.46	2.3,[b] 2.5[c]
bound	0.69	0.54	
unbound	1.54	1.68	
Na^+			
average	0.33	0.54	1.2,[b] 1.0,[d] 1.3[e]
bound	0.11	0.14	
unbound	0.52	0.66	
Cl^-			
average	0.95	0.74	1.52,[b] 2.3,[d] 1.3[e]

[a] Reference 36. [b] Reference 63. [c] Reference 38. [d] Reference 58. [e] Reference 64.

D can be obtained from the mean square displacements via[42]

$$\lim_{t\to\infty} 2tD = {}^1/_3\langle|\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2\rangle \qquad (1)$$

where t is the time and $\mathbf{r}_i(t)$ is the position vector of the ith particle at time t. The diffusion coefficients of water and Na^+ and Cl^- ions were calculated from the slopes of the mean-square displacements vs time. The results from the present work are shown in the second column of Table 5, and data from the simulation of saltwater solution with rigid DNA[36] are given in the third column. The average diffusion coefficient of water was significantly lower than that for bulk SPC/E water,[38] as well as that for the rigid-DNA/saltwater simulation.[36] This also indicates that the presence of the solute has a clear effect on the movement of the water molecules in the system. Interestingly, the dynamics of the solute have generally lowered the diffusion of solvent and counterions near the triplex. This can be seen from the difference in the diffusion constants between this study and the MD simulation of saltwater solution around rigid DNA.[36] We have classified the solvent and Na^+ ions closer than 3.5 Å to the DNA as bound species and the rest as unbound species. As given in Table 5, the diffusion coefficient of the unbound water was higher by more than a factor of 2 over that of water near to the triplex DNA. The diffusion coefficient of unbound Na^+ was higher by about a factor of 4. The last column of Table 5 gives the diffusion coefficients of water, Na^+, and Cl^- without a macromolecule in the system.[63–66] The noticeably lower diffusion coefficients obtained from the current simulation illustrate the large influence of the DNA triple helix on the mobility of solvent and counterions and, by implication, the bulk viscosity.

The NMR rotational correlation time for the water molecules which form the spine of hydration along the M1 groove between strands I and III was 30 ps,[49] which is somewhat higher than the corresponding value in the bulk solution of 12 ps and much higher than that observed for pure salt water of 6 ps.[67] In addition, we find that there is a preferred orientation for water molecules when close to the global helical axis. The solid line in Figure 13 gives the cosine of the average angle between the dipole vector of a water molecule and the helical axis as a function of the distance from the axis to the water molecule. The dotted line indicates the direction of the electric field generated by the charges on the atoms in the triplex as a function of the distance from the helical axis. The dashed line is the cosine of the angle between the average vector of dipole

(63) Smith, P. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1991**, *113*, 6029–6037.
(64) Berkowitz, M.; Wan, W. *J. Chem. Phys.* **1987**, *86*, 376–382.
(65) Reddy, M. R.; Berkowitz, M. *J. Chem. Phys.* **1988**, *88*, 7104–7110.
(66) Trullàs, J.; Giró, A.; Padró, J. A. *J. Chem. Phys.* **1990**, *93*, 5177–5181.
(67) Smith, P. E.; Pettitt, B. M. *J. Chem. Phys.* **1991**, *95*, 8430–8440.

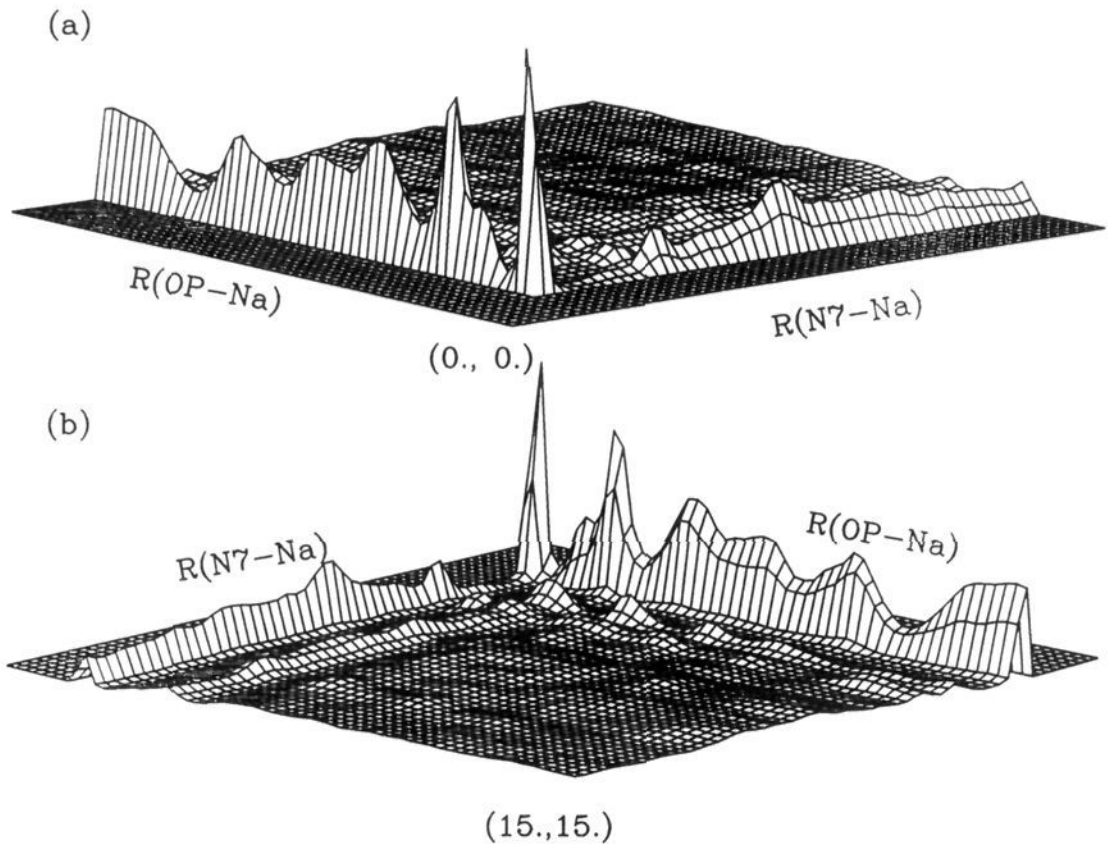

Figure 12. The conditional distribution of Na^+ ions as a function of the distances from N7 atoms in strand III and the phosphate oxygen atoms in the backbone: (a) front view and (b) rear view.

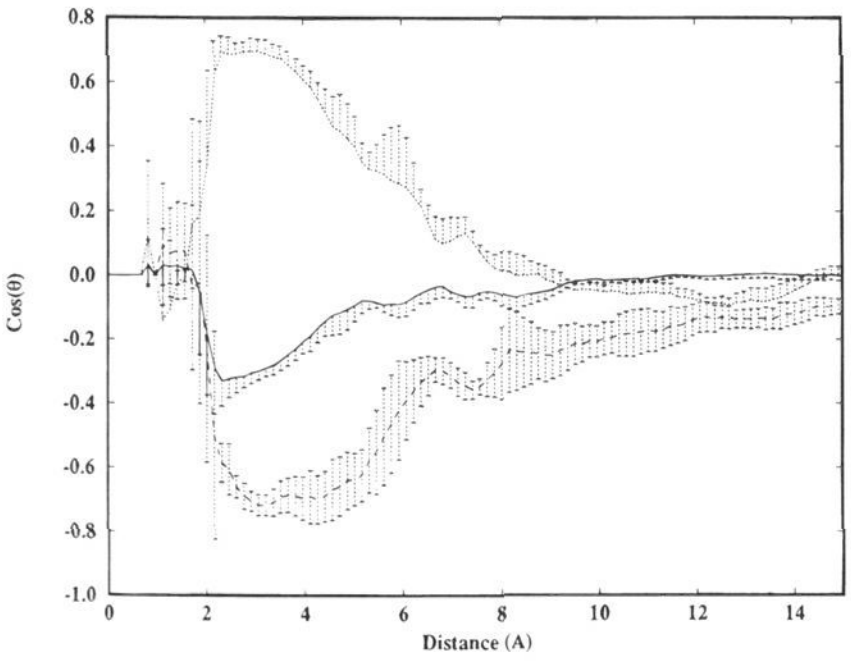

Figure 13. The cosine of the average angle between the dipole vector of water and the global helical axis (solid line), between the direction of the electric field and the helical axis (dotted line), and between the dipole vector and the electric field (dashed line) as a function of the distance from the helical axis.

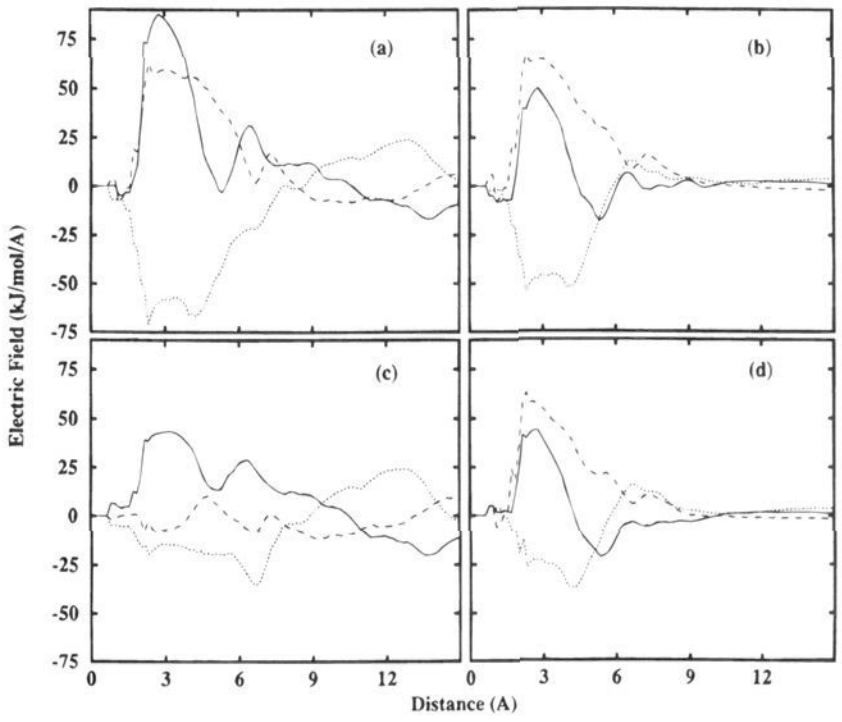

Figure 14. The electric field (a) due to the atoms in the triplex, (b) due to the atoms in the bases only, (c) due to the atoms on phosphate groups, and (d) due to the atoms in the closest base as a function of the distance from the helical axis (solid line, x-component; dotted line, y-component; dashed line, z-component).

moments on water molecules and the electric field vector as a function of the distance. The error bars indicate the statistical error of the calculation. This plot indicates that the average orientation of water molecules is zero at large distances, but molecules within 5 Å of the helical axis show a definite preferred orientation. The average maximum angle is about 110°. The direction of the electric field at that point is about 50° and suggests that the dipole vector and the electric field vector are nearly collinear and opposite to each other. These results imply that water molecules near the helical axis tend to place one of the OH bonds parallel to the global helical axis with the O−H bond pointing in the positive direction of the helical axis.

The orientation is due to the electric field generated by the charges on the atoms in the triplex. Since one of the O−H bonds of the spine waters is, on average, parallel to the z-axis (helical axis), the z-component of the electric field should have a major influence on the preferred orientation. This orientation of water molecules is most prominent in the groove between strands I and III. We now consider which chemical moieties contribute most to the observed behavior. For the following discussion, the global helical axis is taken as the z-axis with the positive direction pointing toward the 3′-end of strand I. As we have discussed earlier, this helical axis is the directional vector of the lowest principal moment of inertia of the triplex. The two directional vectors of the other two principal moments of inertia were taken as the x- and y-axes so that together with the helical axis they form a right-handed coordinate system. In Figure 14 we present the three components of the electric field decomposed into (a) the total field, (b) the field due to the atoms in all bases, (c) the field due to the atoms in phosphate groups, and (d) the field due to the atoms in the closest base as a function of the distance from the helical axis. The x-, y-, and z-components of the electric field were denoted by solid, dotted, and dashed lines, respectively, in all the panels of Figure 14. Figure 14c shows that the phosphate-induced z-component of

the field is nearly 0, indicating that phosphate groups do not appear to have an appreciable direct influence on the average orientation of the water molecule in the spine. Inspection shows that the z-component of the electric field is generated from the charges on the atoms in the base closest to the water molecule.

4. Discussion

During the simulation we observed the temporary disruption of an interbase hydrogen bond between strands I and II due to the transient insertion of a water molecule accompanied by the twisting of the cytosine base out of the base plane. This persisted for approximately 100 ps, during which the integrity of the remaining triplex was maintained. Significant penetration of water and ions into the grooves of the helix was observed, which increased on changing from a rigid to a flexible helix. An increase in water penetration, beyond the classical solvent-accessible surface, has been observed during a simulation of myoglobin.[68] Ion penetration within flexible biomolecules does not appear to have been observed before. These observations could have important consequences for the use of continuum solvent approximations, such as Poisson–Boltzmann techniques, for investigating the solvation of nucleic acids and especially triplexes, since specific molecule effects are not included in those approximations.

Several sodium ions were found to bind simultaneously to the N7 atom of a guanine base in strand III and a nearby phosphate group. However, we did not observe any phosphate sodium ion pairs arranged such that the sodium ions were located on the bisector of the O–P–O angle in a bidentate fashion. In addition, at any one instant not all phosphate groups interact with a sodium ion. On average the number of sodium ions within 4.2 Å of a phosphate group was found to be approximately 11. This is to be compared with counterion condensation theory,[69] which predicts a total of 18 condensed sodium ions. The simulation results appear low in comparison with theory. However, Manning theory is known to overestimate the number of condensed ions observed for duplex DNA by as much as 25%,[69] and therefore the discrepancy between theory and simulation appears reasonable.

We also observed a spine of hydration where several individual water molecules directly solvate the polar groups of the nucleotide bases. These waters of hydration were strongly orientated within the groove by the base to which they were bound. Many of the spine of hydration waters had residence times on the order of 500 ps or longer. This observation is consistent with reported 2D-NMR results for triplexes.[70] Overall, our MD simulation results of the structure and the hydration of the triplex are consistent with the results of NMR experiments.[53,55,70] However, due to a lack of accessible nuclei, the spine of hydration predicted by our simulations is not as yet probed by these experiments.[55] Recently, other workers have postulated that the monovalent ion binding sites seen in this and our previous work[36] are also plausible for pentacoordinated divalent cations.[71]

Our data demonstrate the instantaneous heterogeneity expected, even in homo-oligomers. Each sugar, each base, and each strand display instantaneous differences. It is the ensemble averages from this and other similar calculations which contain the most reliable information for the model system. From the average data presented in the tables, it is clear that, while well-defined averages were achieved, significant local fluctuations are possible.[30] Some of the structural transitions analyzed above thus contribute to the quoted rms values. Convergence of rms fluctuations and other high-order moments of the total distributions are problematic.[42]

5. Conclusions

We have performed a nanosecond time scale MD simulation of a reverse-Hoogsteen antiparallel $(CG{\cdot}G)_7$ model DNA triple helix in 1 M saline solution. The system was simulated without any constraints enforcing the hydrogen bonding between base pairs or any assumptions concerning the positions of ions surrounding the triplex. The triplex remained intact during the simulation with a low deviation (1.6 Å) from the initial model-built starting structure, even though it is significantly less than one full helix turn. The final helix possessed a rise of 3.05 Å and a twist of 30°. Strand II was observed to undergo small fluctuations with backbone dihedrals and sugar puckers indicative of A-form DNA. In contrast, strands I and III displayed significant mobility of the backbone dihedrals and sugar puckers with population of A, B, and other forms.

After the initial rapid relaxation of about 1 Å rms fluctuation from the model-built structure, most of which occurred within 50 ps, a slower relaxation to the final equilibrium occurred. This longer time process, which took about 500–600 ps, was associated with the response of the ions and water to the flexible helix. This is also the time scale for exchange of the tightest bound water molecules in the spine of hydration and N7 of guanine binding sites seen in this and our earlier works.[36] After the prolonged relaxation from our model-built structure, the overall rms fluctuations were relatively small. This occurred even though the cytosine bases and the phosphate backbone had substantial fluctuations. Deviations from our initial model-built structure were largest for ϵ and ζ dihedrals; however, relaxations from the model-built state retained the overall integrity of the triplex structure. The patterns seen here are different from those found in simulations of duplex DNA.[30] The rigidity of a section of triple helix on a long piece of duplex DNA would have pronounced conformational effects on the duplex. This may be an important aspect of the mechanism and efficacy of these systems as artificial repressors.

Acknowledgment. The authors would like to thank the Texas Coordinating Board, the Robert A. Welch Foundation, the NIH, and the National Science Foundation for partial support of this research. We further thank Dr. John Perkyns and Prof. M. Hogan for many interesting conversations.

Note Added in Proof: While this article was in press, the following article appeared, which experimentally determined the stability of triple helices in various salt solutions: Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376–10382.

Appendix A

In this section we outline the procedure for the calculation of helical parameters of triplex DNA. As described in the text, the directional vector corresponding to the smallest principal moment of inertia of the triplex was taken as the global helical axis. The direction of the 3′-end of strand I was chosen as the positive direction of the helical axis.

Only the helical parameters with respect to adjacent base triplet plane pairs, namely, rise, slide, shift, twist, roll, and tilt, were considered here. First, the coordinates of the center of

(68) Lounnas, V.; Pettitt, B. M.; Findsen, L.; Subramaniam, S. *J. Phys. Chem.* **1992**, *96*, 7157–7159.

(69) Manning, G. S. Electric and Elastic Instabilities of DNA. In *Theoretical Biochemistry & Molecular Biophysics: DNA*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: New York, 1991; Vol. 1.

(70) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 395–405.

(71) Potaman, V. N.; Soyfar, V. N. *J. Biomol. Struct. Dyn.* **1994**, *11*, 1035–1040.

mass of each base-triplet plane were calculated, and then the principal moments of inertia and their directional vectors of those base-triplet planes were evaluated. From each of these directional vectors, the vector which made the smallest angle with the global helical axis was chosen as the z-axis of the corresponding base-triplet plane. According to the usual nomenclature[72] the x-axis is directed toward the major groove of duplex DNA. Since the third strand of the triplex is along the major groove of the duplex, we have chosen a directional vector directed toward the third strand from the remaining two vectors as the x-axis of the base-triplet plane. Then the cross product of the z-axis and the x-axis gives the y-axis of the triplet plane.

Now consider two adjacent base-triplet planes i and j. Let O_i and O_j denote the centers of mass of base-triplet planes i and j. Then let $[\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i]$ and $[\mathbf{a}_j,\mathbf{b}_j,\mathbf{c}_j]$ be the unit vectors of right-handed axis systems (defined above) on O_i and O_j, the centers of mass of base-triplet planes i and j. Finally, let $\mathbf{r}_i$ and $\mathbf{r}_j$ be the position vectors of O_i and O_j with respect to the origin of the lab-fixed frame, where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, and $\mathbf{u}$ is the unit vector along the global helical axis. Following are the definitions that we have used for the calculation of helical parameters given in this paper, where only the rise parameter gives a global description of the DNA while the others present a local picture of the structure of the triplex DNA:

translational parameters

$$\text{(a) rise: } |\mathbf{u} \cdot \mathbf{r}_{ij}|$$

$$\text{(b) slide: } |\mathbf{b}_j - \mathbf{b}_i|$$

$$\text{(c) shift: } |\mathbf{a}_j - \mathbf{a}_i|$$

rotational parameters

$$\text{(d) twist: } \frac{\pi}{2} - \cos^{-1}\left[\frac{\mathbf{a}_i \cdot (\mathbf{a}_j \times \mathbf{c}_i)}{|\mathbf{a}_i||(\mathbf{a}_j \times \mathbf{c}_i)|}\right]$$

$$\text{(e) roll: } S_1 \tan^{-1}[|(|\mathbf{c}_i \cdot (\mathbf{r}_{ij} + \mathbf{a}_j - \mathbf{a}_i)| - |\mathbf{c}_i \cdot \mathbf{r}_{ij}|)|]$$

$$\text{(f) tilt: } S_2 \tan^{-1}[|(|\mathbf{c}_i \cdot (\mathbf{r}_{ij} + \mathbf{b}_j - \mathbf{b}_i)| - |\mathbf{c}_i \cdot \mathbf{r}_{ij}|)|]$$

where

$$S_1 = \begin{cases} 1.0 & \text{if } |\mathrm{c}_i \cdot (\mathrm{r}_{ij} + \mathrm{a}_j - \mathrm{a}_i)| < |\mathrm{c}_i \cdot \mathrm{r}_{ij}| \\ -1.0 & \text{otherwise} \end{cases}$$

and

$$S_2 = \begin{cases} 1.0 & \text{if } |\mathrm{c}_i \cdot (\mathrm{r}_{ij} + \mathrm{b}_j - \mathrm{b}_i)| > |\mathrm{c}_i \cdot \mathrm{r}_{ij}| \\ -1.0 & \text{otherwise} \end{cases}$$

give the sign of the roll and twist.[72]

Appendix B

In this section we outline the calculation of the conditional distribution of the Na^+ ions as a function of N7–Na^+ and OP–Na^+ distances. Following Ben-Naim[62] the conditional probability density of finding a Na^+ ion at $\mathbf{r}_{Na^+}$ when the both $\mathbf{r}_{N7}$ and $\mathbf{r}_{OP}$ are known can be written as

$$\varrho(\mathbf{r}_{Na^+}|\mathbf{r}_{N7},\mathbf{r}_{OP}) = \frac{\varrho^{(3)}(\mathbf{r}_{Na^+},\mathbf{r}_{N7},\mathbf{r}_{OP})}{\varrho^{(2)}(\mathbf{r}_{N7},\mathbf{r}_{OP})} \quad \text{(A1)}$$

When $\mathbf{r}_{N7}$ and $\mathbf{r}_{OP}$ approach infinity, eq A1 becomes

$$\varrho(\mathbf{r}_{Na^+}|\mathbf{r}_{N7},\mathbf{r}_{OP}) \rightarrow \frac{\varrho(\mathbf{r}_{N7}|\mathbf{r}_{Na^+})\varrho(\mathbf{r}_{OP}|\mathbf{r}_{Na^+})\varrho^{(1)}(\mathbf{r}_{Na^+})}{\varrho^{(1)}(\mathbf{r}_{N7})\varrho^{(1)}(\mathbf{r}_{OP})} = \frac{\varrho(\mathbf{r}_{Na^+}|\mathbf{r}_{N7})\varrho(\mathbf{r}_{Na^+}|\mathbf{r}_{OP})}{\varrho^{(1)}(\mathbf{r}_{Na^+})} \quad \text{(A2)}$$

Taking the fact that $g(\mathbf{r}_{Na^+\text{-}OP})$ and $g(\mathbf{r}_{Na^+\text{-}N7})$ approach unity for large $\mathbf{r}_{Na^+\text{-}N7}$ and $\mathbf{r}_{Na^+\text{-}OP}$, one can show that eq A2 reduces to the density of Na^+ ions in the system when $\mathbf{r}_{Na^+\text{-}N7}$ and $\mathbf{r}_{Na^+\text{-}OP}$ approach infinity. The calculation of the conditional distribution of Na^+ ions, given the positions of N7 atoms in strand III and the phosphate oxygen atoms in the backbone, is then straightforward. Consider a Na^+ ion at $\mathbf{r}_{Na}$. Let the distance from the N7 atom at $\mathbf{r}_{N7}$ to the Na^+ ion be r_1 and the distance of the OP atom at $\mathbf{r}_{OP}$ to the Na^+ ion be r_2. Now consider a sphere with the center at the position of the N7 atom with a radius of r_1. Consider a second sphere, the center of which is located at the position of a phosphate oxygen atom, with a radius of r_2. The two spheres will overlap if the distance between N7 and OP, $r < r_1 + r_2$, and the Na^+ ion remains at one point of the intersection of two spheres. Now we can imagine a second set of spheres with radii $r_1 + \delta r_1$ and $r_2 + \delta r_2$ from the corresponding centers. Then $\delta r_1 \times \delta r_2$ is the area element at the position of the Na^+ ion, as given in Figure 15. The accessible conditional volume for a Na^+ ion will be the volume generated by rotating the area element by 2π around the $r_1 - r_2$ axis while keeping the distance from the Na^+ ion to the $r_1 - r_2$ line, y, constant. In normal $g(r)$ calculations, the evaluation of the density by dividing the number of particles in each bin by the corresponding volume is the last step. In contrast, for the conditional distribution function, one has to calculate the density for each volume element and bin that density *element* as a function of the two conditional distances, namely, r_1 and r_2, since the volume element is a function of all three distances, r_1, r_2, and r.

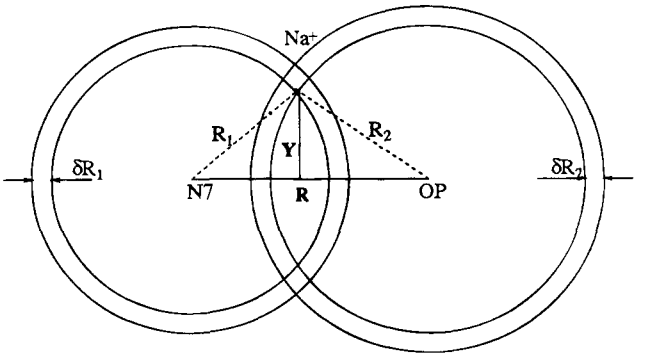

Figure 15. Bipolar coordinates for the integration to calculate the volume element for the conditional distribution of Na^+ ions when the distances r_1, r_2, and r are known.

JA942440U

(72) Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627–634.