Molecular Dynamics Simulation of a PNA·DNA·PNA Triple Helix in Aqueous Solution

George C. Shields,[‡,§] Charles A. Laughton,[⊥] and Modesto Orozco*[,‡]

Contribution from the Departament de Bioquímica i Biología Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, and Cancer Research Laboratories, Department of Pharmaceutical Sciences, University of Nottingham, Nottingham NG7 2RD, U.K.

Received July 14, 1997. Revised Manuscript Received February 13, 1998

Abstract: Molecular dynamics simulations have been used to explore the conformational flexibility of a PNA·DNA·PNA triple helix in aqueous solution. Three 1.05 ns trajectories starting from different but reasonable conformations have been generated and analyzed in detail. All three trajectories converge within about 300 ps to produce stable and very similar conformational ensembles, which resemble the crystal structure conformation in many details. However, in contrast to the crystal structure, there is a tendency for the direct hydrogen-bonds observed between the amide hydrogens of the Hoogsteen-binding PNA strand and the phosphate oxygens of the DNA strand to be replaced by water-mediated hydrogen bonds, which also involve pyrimidine O2 atoms. This structural transition does not appear to weaken the triplex structure but alters groove widths and so may relate to the potential for recognition of such structures by other ligands (small molecules or proteins). Energetic analysis leads us to conclude that the reason that the hybrid PNA/DNA triplex has quite different helical characteristics from the all-DNA triplex is not because the additional flexibility imparted by the replacement of sugar−phosphate by PNA backbones allows motions to improve base-stacking but rather that base-stacking interactions are very similar in both types of triplex and the driving force comes from weak but definate conformational preferences of the PNA strands.

Introduction

Polyamide nucleic acid (PNA) was designed using computer modeling as an uncharged polymer analogue of DNA and RNA.[1] PNA is able to bind single strands of RNA in a sequence-specific manner,[2] which opens up the possibility of using PNAs as antisense drugs. Furthermore, binding of PNA to DNA is also very tight and occurs with a specificity similar to that observed in the formation of DNA duplexes.[3] In fact, PNA·DNA duplexes are much more stable that DNA·DNA duplexes, and the introduction of a suitable strand of PNA can displace the corresponding DNA strand from a preexisting DNA duplex.[1] This opens up the possibility of applying PNAs in the field of antigene therapy.

Several authors have noted[3−7] that in certain circumstances PNA can form a 2:1 complex with DNA. Structural studies have shown that such 2:1 complexes involve the formation of a PNA·DNA·PNA triple helix, where an additional strand of PNA binds to the PNA·DNA duplex by Hoogsteen-type hydrogen bonds, just as observed in the all-DNA triple helix [DNA(t)].

The three-dimensional structure of PNA·DNA·PNA triple helices in solution is unknown, and it is probably sequence-dependent. However, the crystal structure of a homopurine PNA·DNA·PNA triple helix has been recently obtained by Betts et al.[7] using high-resolution X-ray diffraction data of a purine DNA strand of sequence 5′-GAAGAAGAG-3′ bound to an hairpin PNA containing two pyrimidine sequences (Nterminus-CTCTTCTTC−Cterminus and Nterminus-CTTCTTCTC−Cterminus) separated by an inert peptidic loop. The crystal structure shows the DNA bound to the complementary PNA sequence by Watson Crick-type hydrogen bonds. The other PNA sequence is antiparallel to the Watson−Crick PNA (WC−PNA) strand ("parallel" to the DNA strand) and is bound to the purine strand of DNA by Hoogsteen type hydrogen bonds. The overall helix is right-handed, with the Hoogsteen PNA strand (H−PNA) in the major groove of the DNA·PNA duplex. The structural characteristics of the crystalline PNA·DNA·PNA triple helix differ notably from those of DNA(t). Thus, while the rise (3.4 Å) is close to that found in DNA(t), [8−12] the twist values are

[‡] Universitat de Barcelona.
University of Nottingham.
[§] On leave from Department of Chemistry, Lake Forest College, Lake Forest, IL 60045.

(1) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497.
(2) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. *Science* **1992**, *258*, 1481.
(3) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Nordén, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566.
(4) Almarsson, O.; Bruice, T. C.; Kerr, J.; Zuckermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7518.
(5) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. *J. Am. Chem. Soc.* **1992**, *114*, 1895.
(6) Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. *J. Am. Chem. Soc.* **1992**, *114*, 9677.
(7) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. *Science* **1995**, *270*, 1838.
(8) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. C. *Nucleic Acids Res.* **1976**, *11*, 4141.
(9) Radhakrishnan, I.; Patel, D. J. *Biochemistry* **1994**, *33*, 11405.
(10) Raghunathan, G.; Miles, H. T.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 455.

S0002-7863(97)02344-5 CCC: $15.00
Published on Web 06/24/1998

about 6−7 degrees smaller than the values suggested for DNA(t), which leads to a helical periodicity of 15−16 bases per turn in the PNA·DNA·PNA triplex compared with 12 bases per turn in the DNA(t).[8−12] Finally, the X-displacement of the bases from the helix axis is 3−4 Å larger than the values derived by NMR and MD techniques for DNA(t).[9−12] In summary, beside the general helical shape there are few similarities between all-DNA and PNA·DNA·PNA triplexes.

The origin of the differences between DNA(t) and PNA·DNA·PNA triplexes might lie in their different chemical structures, but other factors could play a part such as lattice restrictions in the crystal of PNA·DNA·PNA or the restraining effect of the peptidic loop connecting the two PNA strands in the crystal structure. Furthermore, considering the flexibility of the PNA backbone (see below), the PNA·DNA·PNA triplex might adopt a wide range of different conformations in solution at room temperature, some of them near DNA(t)-type structures, and others near the X-ray structure. It is therefore of interest to examine the dynamic characteristics of the PNA·DNA·PNA triplex in aqueous solution and try to obtain a more complete picture of the structure and reactivity of the PNA·DNA·PNA triplex under something closer to potential physiological conditions.

Molecular dynamics (MD) provides very reasonable pictures of protein structure and flexibility. Until recently MD was not so reliable in representing nucleic acid structures, due to deficiencies in the force-fields, and in the treatment of long-range electrostatic effects. However, recent improvements in force-field parametrizations and MD algorithms have made the accurate modeling of nucleic acid (DNA, RNA, and hybrid) structures possible.[13−16] Furthermore, we have recently shown that MD gives reasonable samplings of the configurational space of DNA(t) in solution, being able to drive the DNA from erroneous starting structures to the correct conformation.[12] These results strongly suggest that MD should be a reliable tool to study the structure of the PNA·DNA·PNA triplex in solution.

In this paper we present a detailed MD study of the PNA·DNA·PNA triplex in aqueous solution. Three 1.155 ns trajectories have been simulated starting from three different starting models of the PNA·DNA·PNA triplex: (i) the crystal structure (P-type),[7] (ii) the A-type structure derived from fiber diffraction of DNA(t),[8] and (iii) the B-type structure derived from NMR data on DNA(t).[10] The study allowed us to describe a large part of the configurational space accessible to PNA·DNA·PNA triplexes as well as to determine the reactive characteristics of these structures. Furthermore, the study represents an additional test of the quality of the new MD algorithms to study complex and highly charged macromolecules.

Methods

General. All the MD simulations were done with the AMBER5.0 suite of programs.[17] The analysis of the MD trajectories was performed using Curves,[18] Newhelix and in-house software. Quantum mechanical calculations were carried out using Spartan3.1 and Gaussian 94.[19] All the calculations were performed on the IBM SP2 supercomputer of the Centre de Supercomputació de Catalunya (CESCA) and on workstations in our laboratories.

Construction of the Starting Structures. The sequence chosen corresponded to that of the crystal structure, that is a purine strand of DNA: d(GpApApGpApApGpApG) and two complementary pyrimidine strands of PNA. All the cytosines of the PNA Hoogsteen strand were considered in their protonated form, mimicking the situation found in water at acidic pH. The peptide linker present in the crystal structure was omitted; instead each PNA strand was capped with the conventional methylamino and acetyl groups (Figure 1). Three starting structures were selected for MD simulation. The first (the P-type structure) was obtained from crystal structure,[7] and the only modification was to remove the conecting loop. The second was adapted from the first by using restrained energy minimizations to force the whole of the DNA strand, and the bases of the PNA strands, into the positions they would take up in a corresponding all-DNA triplex of the same sequence built up in the fiber-diffraction based conformation (the A-type structure).[8] The third starting conformation was generated in the same manner as the second one but using as the template the NMR-based structure of DNA(t) (the B-type structure).[18] Each of the three starting models was immersed in a box of 4165 TIP3P water molecules,[20]) and four sodium counterions (Hoogsteen cytosines were protonated) were added in the most negative regions found by Poisson−Boltzmann calculations[21] to obtain neutral systems.

Parametrization. The standard AMBER5.0 all atom force-field[17b] was used to describe the DNA. This force-field contains parameters for most units except the PNA residues (see Figure 1). van der Waals and bonded parameters can be easily transferred from standard peptidic or nucleic acid fragments in AMBER, but atomic charges for the new residues need to be obtained. To obtain RESP charges,[22] the PNA residues were separated into peptidic and basic fragments. *N*-Methylated bases (C,T and C+) were used to obtain charges for the bases. To obtain charges for the peptidic fragment of PNA, a model was built with the N- and C-terminii capped with the conventional acetyl and *N*-methylamino moieties and the side chain capped with NHCOMe. These capped fragments were partially optimized using the AM1-Hamiltonian[23] while the backbone geometries observed in the crystal were preserved. Electrostatic potenials used in the calculation of RESP[22] charges were obtained using the HF/6-31G* wave function. Final charges are shown in ref 24, and the corresponding "prep.in" files for AMBER calculations are available upon request to the authors.

Molecular Dynamics Protocols. A 10-step protocol was used for the heating and equilibration process, which ran for 50 ps of MD (see Figure 2). The final equilibrated structures were the starting points

(11) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 17.
(12) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463−7469.
(13) Weerasinge, S.; Smith, P. E.; Mohan, V, Cheng, Y.-K.; Pettit, B. M. *J. Am. Chem. Soc.* **1995**, *117*, 2147−2158.
(14) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.
(15) (a) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194. (b) Yang, L.; Pettit, B. *J. Phys. Chem.* **1996**, *100*, 2564−2566. (c) Weerasingle, S.; Smith, P. E.; Petitt, M. *Biochemistry* **1995**, *34*, 16269−16278.
(16) (a) Cheatham, T. E.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434−444. (b) Spector, T. I.; Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 7095−7104. (c) Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805−4825. (d) Miller, J.; Kollman, P. A. *J. Mol. Biol.* **1997**, *270*, 436−450.
(17) (a) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheetham, T. E.; DeBolt, S.; Fergurson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun*. **1995**, *91*, 1. (b) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K.; Fergurson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 51.
(18) Lavery, R.; Sklenar, J. *J. Biomol. Struct. Dyn*. **1988**, *6*, 63.
(19) (a) Spartan version 3.1; Wavefunction, Inc.: Irvine, CA, 1996. (b) Frisch, M. J.; Trucks, G. W.; Schelgel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, G. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Ciolowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; 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.; Pople, J. A. GAUSSIAN94 (Rev. A.1); Gaussian Inc.: Pittsburgh, PA, 1995.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R.; Klein, M. *J. Chem. Phys*. **1983**, *79*, 926.
(21) DELPHI Computer Program; Biosym Co.: San Diego, 1992.
(22) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem*. **1993**, *97*, 10269.
(23) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. *J. Am. Chem. Soc*. **1985**, *107*, 3902.

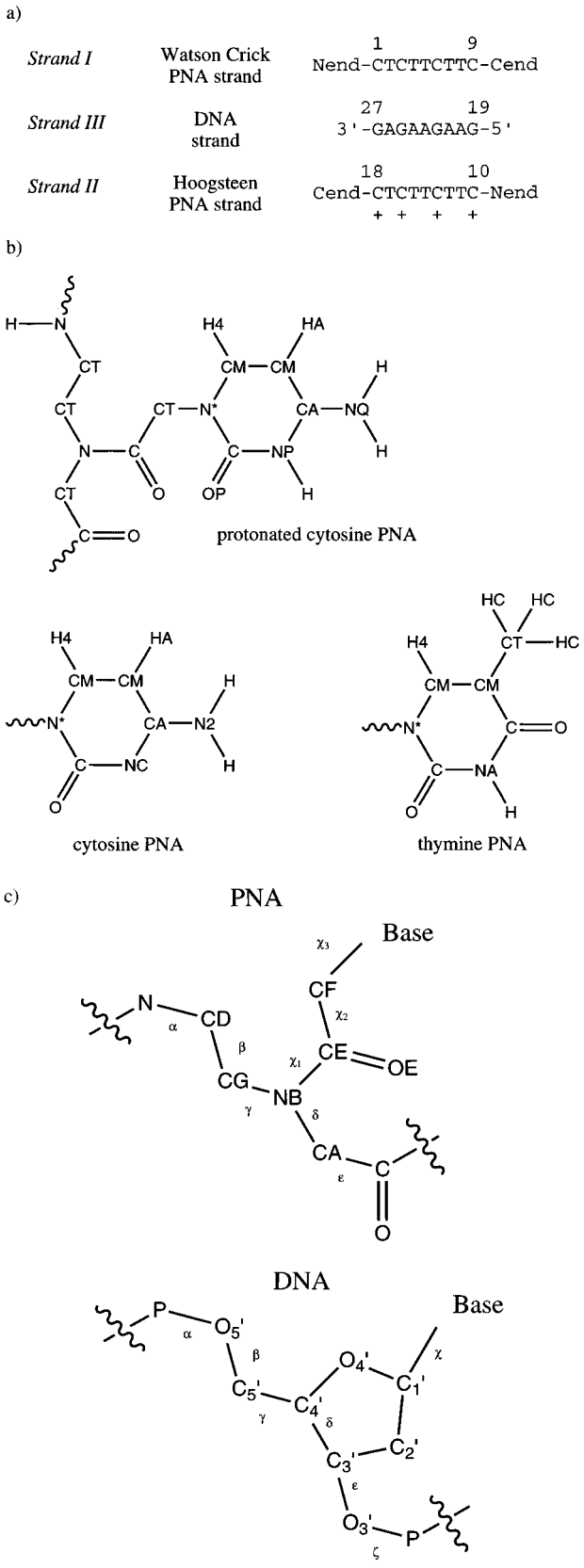

Figure 1. (a) Sequence of the PNA·DNA·PNA triplex studied, showing numbering scheme and (b) AMBER atom types used for new PNA residues. Omitted hydrogen atoms were all of type H1. (c) Definition of DNA and PNA torsion angles.

for three MD trajectories extended over 1.0 ns. All simulations were done using periodic boundary conditions and the particle mesh Ewald (PME, see ref 25) approach for evaluating long-range electrostatic effects. Simulations were carried out in the isothermic-isobaric ensemble (NPT) for $T = 300$ K and $P = 1$ atm. A time step of 2 fs was used for the integration of Newton's equations, since all the bond lengths were constrained to their equilibrium values using SHAKE.[26]

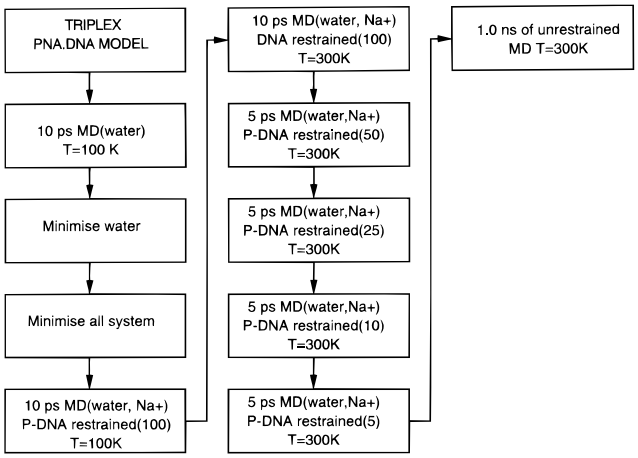

Figure 2. Molecular dynamics/energy minimization protocol for equilibration and data collection.

Time-Averaged Structures. The last 500 ps of each trajectory were used to obtain "MD-averaged" representative structures of the PNA·DNA·PNA helix. For this purpose, all the conformations sampled were coordinate-averaged after least-squares fitting, and the resulting structures were minimized using steepest descent and conjugate gradient techniques to remove the most severe of the geometrical artifacts introduced by the averaging procedure (e.g., bad C−H bond lengths in methyl groups). To avoid the minimization changing the structures too much, all heavy atoms were restrained to their initial positions with a force constant of 100 kcal/mol/Å.

Calculation of Molecular Interaction Potentials. Calculations of molecular interaction potential (MIP) were performed to determine the ability of the structures to interact with positively charged groups, such as groove binders or proteins. The MIP was computed as an electrostatic term computed for a positive charge using Poisson−Boltzmann techniques, plus a van der Waals term determined using the van der Waals parameters of a TIP3P (20) particle. AMBER charges and van der Waals parameters[17b] were used for the DNA. The Poisson−Boltzmann equation was solved numerically using the nonlinear formalism, as implemented in DELPHI[21] for a medium with an ionic strength of 0.1 M. A dielectric response of 2 and 80 was assigned to the triplex and water, respectively.

Calculation of Stacking Interactions. Stacking calculations were performed for three MD-averaged conformations as well as for the two possible types of DNA(t) and for the PNA·DNA·PNA crystal structure. In all the cases the backbones were removed, and the charge on CF or C1′ was adjusted to have integer charge in the bases (0 or +1). AMBER5.0 charges and van der Waals parameters were used for all the atoms except for CF, C1′ and the methyl group of thymines which were described using an united atom model. The stacking energy was computed as two contributions: (i) the intrastrand term due to

(24) Charges for the PNA backbone are as follows: N: −0.4943, HN: 0.2758, CD: 0.0982, HD1,2: 0.0354, CG: 0.1750, HG1,2:-0.0405, NB: −0.0470, CA: −0.4452, HA1,2: 0.1968, C: 0.6797, O: −0.5883. Charges for neutral (PNA) cytosine are as follows: CE: 0.5746, OE: −0.5654, CF: −0.0908, HF1,2: 0.0674, N1: −0.1238, C6 0.0016, H6: 0.2038, C5: −0.5312, H5: 0.1951, C4: 0.8659, N4: −1.0039, N41,42: 0.4342, N3: −0.7541, C2: 0.8121, O2: −0.6244. Charges for protonated (PNA) cytosine are as follows: CE: 0.0.6876, OE: −0.5851, CF: −0.0978, HF1,2: 0.0916, N1: −0.0460, C6 0.0296, H6: 0.2362, C5: −0.4074, H5: 0.2194, C4: 0.6550, N4: −0.9507, N41,42: 0.4888, N3: −0.3395, H3: 0.3550, C2: 0.5180, O2: −0.4714. Charges for (PNA) thymine are as follows: CE: 0.5404, OE: −0.5466, CF: 0.0556, HF1,2: 0.0500, N1: −0.2108, C6− 0.1740, H6: 0.2247, C5: 0.0098, C7: −03459, H71,2,3: 0.1080, C4: 0.6346, O4: −0.5635, N3: −0.5507, H3: 0.3681, C2: 0.6807, O2: −0.5835.

(25) (a) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys*. **1993**, *98*, 10089. (b) Essmann, U.; Perera, L.; Berkowitz.; M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys*. **1995**, *103*, 8577−8593.

(26) Ryckaert, J. P.; Ciccote, G.; Berendsen, J. C. *J. Comput. Phys*. **1977**, *23*, 327.

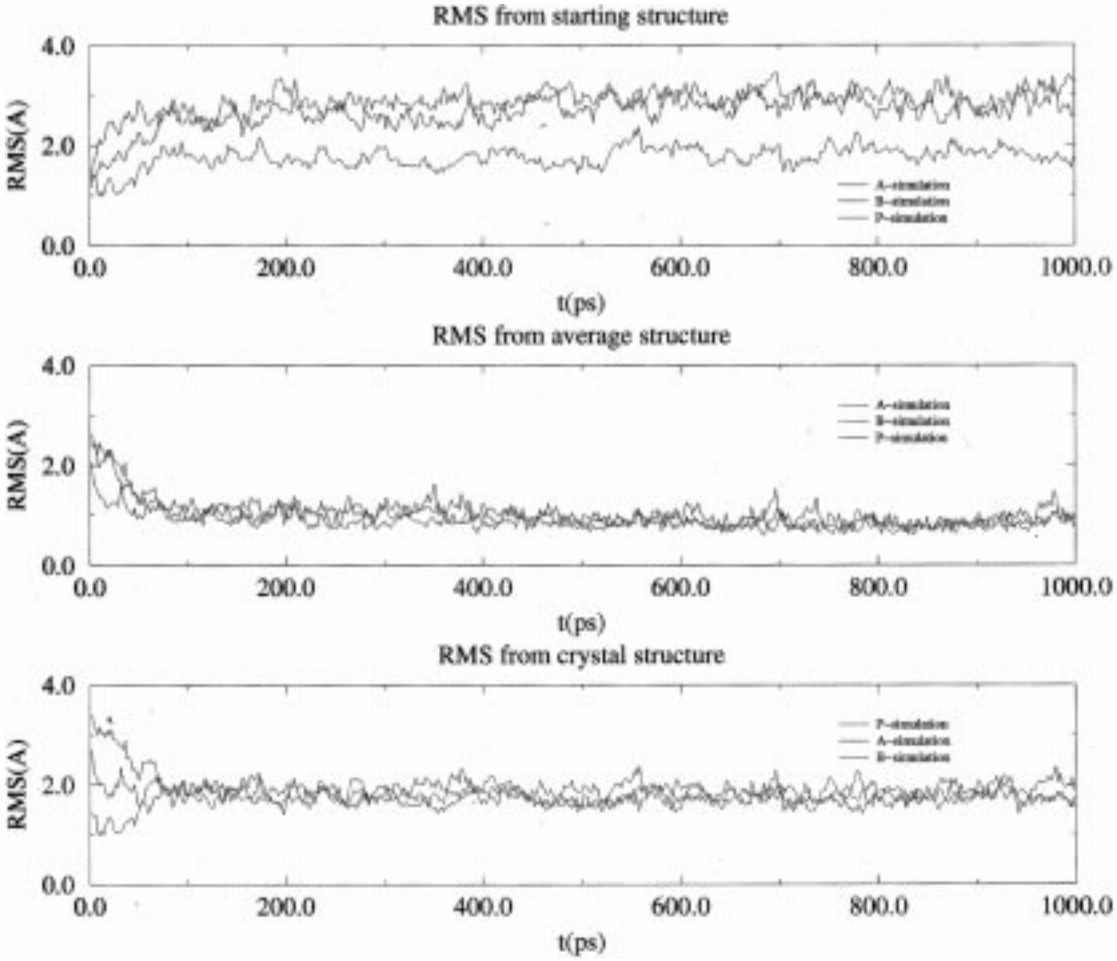

Figure 3. RMS deviation for P-, B-, and A-type trajectories. Top: from starting structures. Middle: from time-averaged structures in the P-, B-, and A-type trajectories. Bottom: from crystal structure.

interactions of bases within each DNA or PNA strand and (ii) the interstrand term due to stacking interactions between bases in different strands. To avoid bias in the results the first and last triplex steps of the helix were removed from the analysis.

Results and Discussion

Gross Structural Features. Analysis of the three trajectories showed that the equilibration process led to stable structures, as noted in macroscopic properties of the systems such as potential and total energy, the density, radii of gyration, etc. (data not shown). No major helical unfolding or transitions to double helical structures were detected during the simulations. Only a local and partial unfolding was detected in the 5′ end of the helix, as the N-terminal cytosine of the H−PNA strand (C_{10}—see Figure 1 for numbering scheme) lost its hydrogen bonding to the G(DNA):C(PNA) Watson Crick base pair during the course of all three simulations. C_{10} ends up stacked over the 5′-terminal guanine in trajectories P and A, while in trajectory B it moves into the major-major groove of the PNA, interacting edge-to-face with thymines 7 and 8 in the WC−PNA strand. Interestingly a similar structural anomaly for the equivalent of C_{10} is also found in one of the two molecules in the asymmetric unit of the crystal structure (not the one used for the intial construction of the models used here), even though in the crystal structure this cytosine does not represent the N-terminus of the PNA peptide. However, to avoid bias in the comparison analysis the triplex step defined by C_{10} was not considered in the analysis of the PNA·DNA·PNA structures.

Convergence in the Trajectories. Key tests of the reliability of MD simulations are the determination of (i) whether sampling is independent of the starting conformation and (ii) whether the experimental geometry is sampled during simulations. When these two conditions are fulfillled the MD simulation may be expected to provide a reasonable sampling of the conformational space most easily accessible to the molecule. The study of the MD trajectories then provides information on key issues such as the configurational flexibility of the helix, its flexibility on the nanosecond time scale, and even its reactive properties.

The requirement of the independence of the sampling from starting configuration means in practice that different trajectories starting from different conformations should converge to the same region of the configurational space. The convergence of MD trajectories on the nanosecond time scale is sytem-dependent and is not always simple (see for example ref 17c), due to the possible existence of metastable structures whose transitions to the most stable conformer are slow, and can be detected or not depending on technical details such as the set up of the system or the equilibration protocol used. However, other studies have reported (see example refs 16a and 17d) consistently good convergence for duplex and triplex(12) structures in MD simulations on the nanosecond time scale. These recent results allowed us to be confident about achieving reasonable convergence on the nanosecond time scale for the PNA·DNA·PNA triplex considered here.

The RMS deviations of the three trajectories with respect to their corresponding starting conformations [i.e., rms(X,X_o), X

Table 1. RMS Deviations (in Å) between Structures Sampled during Last 500 ps of the Three Trajectories (X) and the (a) Starting (X_0) or (b) Average ($\langle X \rangle$) Conformations

	A	B	P
	(a)		
A_0	2.9(0.2)	3.0(0.2)	3.7(0.3)
B_0	2.9(0.2)	2.9(0.2)	3.7(0.3)
P_0	1.7(0.2)	1.8(0.2)	1.8(0.2)
	(b)		
$\langle A \rangle$	0.8(0.1)	1.1(0.1)	1.6(0.2)
$\langle B \rangle$	1.1(0.1)	0.9(0.1)	1.6(0.2)
$\langle P \rangle$	1.5(0.2)	1.6(0.2)	0.9(0.1)

= A,B, or P] are displayed in Figure 3(top). Both A- and B-type trajectories are moving away very quickly (transition time around 100−200 ps) from the starting conformation, the average RMS deviation between the structures in the second half of the trajectories and the starting conformations [i.e., $\langle \mathrm{rms}(A,A_o) \rangle$ and $\langle \mathrm{rms}(B,B_o) \rangle$] being around 3 Å (see also Table 1). On the contrary the P-type trajectory samples regions much closer to the original conformation (the crystal one), with $\langle \mathrm{rms}(P,P_o) \rangle$ around 1.7 Å (see also Table 1). Detailed inspection of Figure 3 (top) shows that none of the trajectories are really diverging, on the contrary, after a 100−200 ps transition all the trajectories are stable, it being possible to define A-, B-, and P-type averaged structures ($\langle X \rangle$, X = A, B, or P) which represent within 0.8−0.9 Å RMS deviation [$<\mathrm{rms}(X,\langle X \rangle)>$] the structures sampled during the last half of the trajectories (see Table 1). These results suggest that neither B- nor A-type conformations represent stable structures for the PNA·DNA·PNA triple helix, and the trajectories starting from such conformations converge to other conformations. On the contrary, MD simulations suggest that the P-type structure, as determined by X-ray difraction in crystals, is quite stable in aqueous solution.

The A- and B-trajectories are converging to the same region of conformational space, as noted in the RMS deviations between one trajectory and the average conformation representing the other trajectory [$<\mathrm{rms}(A,\langle B \rangle)>$, and $<\mathrm{rms}(B,\langle A \rangle)>$], which are about 1.1 Å (Table 1). This compares with the RMS deviation of 0.8−0.9 Å for $\langle \mathrm{rms}(A,\langle A \rangle) \rangle$ and $\langle \mathrm{rms}(B,\langle B \rangle) \rangle$. Actually, the A- and B-trajectories are converging spontaneously to a region of conformational space which is very close to the region sampled in the P-trajectory. Thus, $\langle \mathrm{rms}(A,\langle P \rangle) \rangle$ and $\langle \langle \mathrm{rms}(B,\langle P \rangle) \rangle$ are small (around 1.5 Å); only around 0.6 Å larger than $\langle \mathrm{rms}(P,\langle P \rangle) \rangle$. In summary, the three MD simulations lead to slightly different average structures, but all of them are sampling the same region of conformational space. Interestingly, calculation of RMS deviations between structures from the second half of each trajectory and the crystal conformation [i.e., $\langle \mathrm{rms}(X,P_o) \rangle$, X = A, B or P] show (see Figure 3, and Table 1) that the MD-converged structures are not far from the conformation detected in X-ray experiments (RMS deviations in all cases 1.7 Å), while they are clearly far from the alternative B- and A-type conformations (RMS deviations around 3 Å). The convergence of all the simulations to a small region of configurational space illustrates the quality of the MD simulations and suggests that information on the dynamical properties of PNA·DNA·PNA triplex in solution can be obtained with some confidence by the analyis of the MD trajectories.

Analysis of Helical Parameters. The helical characteristics of the triplex determined from MD-simulations are displayed in Table 2. The helix sampled during the P-trajectory was divided for some analyses into two parts owing to the existence of conformational differences between the first and second halves of the helix. These differences arise from changes in the backbone conformation that accompany changes in the H-bonding pattern between the DNA and the H−PNA strands. These lead to changes in the dimension of the groove defined by the DNA and the H−PNA strands (see below). In general the PNA·DNA·PNA triplex in solution is defined by a rise around 3.3−3.4 Å, a twist of 20−23 degrees (suggesting a periodicity of 16−18 bases per turn), a slide between 2.3 and 2.7 Å, a shift between 0.9 and 1.0 Å, and small values of both tilt and roll. The agreement between the average helical parameters and the helical parameters for the MD-averaged structures demonstrate that the MD-averaged structures are representative of the conformations sampled during the trajectory. The helical parameters of the structures obtained during the three simulations are very similar to those found in the crystal structure and differ clearly from the values suggested for the all-DNA triple helix which were used to build up the A- and B-starting conformations.

PNA Strand Backbone Conformations. Average backbone dihedral angles for the three trajectories are displayed in Table 3. PNA backbone dihedral angles obtained in the three simulations are very similar to the X-ray structure-derived values. Furthermore, the agreement between the PNA backbone conformation in our MD simulations and PNA backbone conformation in the PNA duplex determined by X-ray crystallography[27] is also remarkable. The only discrepancy appears for the H−PNA strand in the P-type trajectory, since in this case the c_1 and c_2 values are slightly different from those obtained in other simulations or from the X-ray structures. Demonstrating the convergence of the simulations, the PNA backbone in the A- and B-trajectories matches the experimental data despite the differences in the starting conformations. The fluctuations of the torsional angles demonstrate that the PNA backbones are in general very flexible, and during the equilibrium part of the trajectories several reversible transitions were detected (data not shown). This flexibility is indirectly supported by the large variation in the different backbone torsions found between NMR and X-ray determined structures of PNA·DNA, PNA·RNA, and PNA·DNA·PNA complexes.[7,28−30] MD simulations suggest that a and e are the most flexible bonds, while g is the most rigid one, in agreement with NMR data on the PNA·DNA duplex. We observe that changes in these dihedral angles are quite concerted, canceling out to large extent their individual effects on the relative positions of the bases. This probably explains why there is not much fluctuation in the helical parameters despite the flexibility of the PNA backbones.

DNA Strand Backbone Conformations. The three MD simulations provide very similar values for the backbone dihedrals of the DNA strand despite the different starting conformations. All the torsions exhibit uniform flexibility and are equally sampled during the trajectory. The overall flexibility of the DNA backbone is clearly smaller than that of the PNA backbones. The agreement with X-ray data is perfect for a, b, and g torsions, while the largest variations are found for the d torsion, which correlates well with changes in ribose puckering with respect to the crystal structure. Thus, calculation of pseudorotation phase angles (Table 4) suggests sampling in the *South* to *Southeast* part of the pseudorotational circle, ranging from the C2′-endo to the O4′-exo conformations (average phase

(27) Rasmussen, H.; Sandholm, J. *Nature Struct. Biol.* **1997**, *4*, 98.
(28) Leijon, M.; Gräslund, A.; Nielsen, P. E.; Buchardt, O.; Nordén, B.; Kristensen, S. M.; Eriksson, M. *Biochemistry* **1994**, *33*, 9820−9825.
(29) Brown, S. C.; Thomson, S. A.; Veal, J. M.; Davis, D. G. *Science* **1994**, *265*, 777.
(30) Eriksson, M.; Nielsen, P. E. *Nature Struct. Biol.* **1996**, *3*, 410−413.

Table 2. Helical Parameters of the PNA·DNA·PNA Triplex Determined from MD Trajectories Compared with X-ray Diffraction Data[a]

simulation	shift	slide	rise	tilt	roll	twist
A	0.9(0.1)	−2.3(0.1)	3.3(0.1)	1.2(0.9)	1.5(1.3)	22.9(0.7)
	0.9	*−2.3*	*3.3*	*1.1*	*1.4*	*22.9*
B	0.9(0.1)	−2.3(0.1)	3.3(0.1)	1.6(1.3)	1.1(0.9)	22.9(0.7)
	1.0	*−2.3*	*3.3*	*1.1*	*1.5*	*22.9*
P(all)	1.0(0.1)	−2.6(0.1)	3.4(0.1)	0.6(0.9)	0.1(1.3)	20.7(0.8)
	1.0	*−2.6*	*3.4*	*0.6*	*0.0*	*20.7*
P(1st 5)	0.9(0.2)	−2.7(0.2)	3.3(0.1)	0.6(1.7)	−1.2(2.3)	20.7(1.3)
P(last 5)	1.1(0.2)	−2.5(0.2)	3.4(0.1)	0.5(1.5)	1.4(2.0)	21.3(1.4)
crystal	0.95	−2.59	3.37	−0.2	2.6	22.6

[a] Roman numbers correspond to values averaged and in parentheses the standard fluctuation. Values in italic correspond to the values obtained for the time averaged structures obtained for each simulation. The helical parameters of the first and last part of the P-type conformation were also determined. Translation parameters are in Å, and rotation parameters are in deg.

Table 3. Torsional Angles Defining the PNA·DNA·PNA Triplex in the Three MD Simulations and in the Crystal Structure[a]

sim/strand		α	β	γ	δ	ϵ	ζ	χ_1	χ_2	χ_3	ψ
A	I	−102(24)	72(11)	82(9)	96(12)	−20(24)		0(10)	−142(19)	70(14)	
	II	−106(34)	68(13)	76(8)	99(13)	−17(21)		2(10)	−170(11)	88(12)	
	III	−68(11)	174(9)	56(10)	107(16)	−171(9)	−89(11)				−122(18)
B	I	−102(17)	71(11)	82(10)	96(12)	−24(22)		1(10)	−143(20)	71(14)	
	II	−107(21)	67(11)	75(8)	99(13)	−20(21)		1(11)	−168(11)	89(12)	
	III	−69(12)	175(10)	56(10)	108(17)	−171(10)	−89(12)				−121(19)
P	I	−102(14)	70(10)	80(8)	96(13)	−24(13)		2(11)	−141(21)	69(14)	
	II	−102(33)	73(15)	81(9)	101(13)	−28(28)		2(11)	−169(12)[b]	83(11)[b]	
									−118(13)[c]	58(12)[c]	
	III	−69(12)	175(10)	56(10)	102(18)	−169(10)	−85(11)				−129(20)
X-ray	I	−103	73	70	93	−15		1	−170	89	
	II	−108	69	69	87	−5		1	−175	102	
	III	−70	173	61	77	−161	−69				−167

[a] Strand I is the Watson−Crick PNA, strand II is the Hoogsteen PNA, and strand III is the DNA (for nomenclature of the torsion angles see Figure 1). [b] Values for the first half of the helix (from C10 to T14). [c] Values for the last part (from T14 to C18) of the helix. All the values are in deg.

Table 4. Pseudorotational Phase Angle (in deg) Obtained for the Three MD-Simulations for Each Nucleotidic Step in the DNA Strand[a]

nucleotide	A-trajectory	B-trajectory	P-trajectory
G(5′)	105(33)	128(26)	124(37)
A	128(21)	120(20)	118(20)
A	95(15)	97(17)	92(13)
G	114(22)	120(20)	112(19)
A	107(21)	112(26)	104(20)
A	102(20)	106(20)	85(19)
G	113(21)	115(19)	105(19)
A	110(23)	95(16)	96(20)
G(3′)	123(22)	127(21)	114(26)
av	111	113	106

[a] The standard deviation in the average are displayed in parentheses.

angle around 110 degrees). This finding does not agree with the crystal structure for the triplex helix, which with a 2.5 Å resolution favors a C3′-endo puckering, as is suggested by NMR data for RNA·PNA duplexes.[29] The discrepancy can be related to a certain tendecy of AMBER force-field to overestimate the relative stability of the C2′-endo conformation, but we believe that it reflects a notable flexibility of the sugars. In support of this we note that the ribose puckerings found here are in perfect agreement with the recent NMR-determined PNA·DNA duplex,[30] where phase angles in the range of 88−130 degrees were observed, and with the NMR structure of a DNA triplex where the phase angles again range between very similar values.[9−11] As an additional remark, we should note that the authors of these NMR studies did not use AMBER force-field for their structure refinements.

It is difficult to explain the differences between the ribose puckering in complexes of PNA and DNA determined from NMR and MD simulations and the puckerings determined by X-ray techniques for the PNA·DNA·PNA triplex. It might be that solvent or packing effects are responsible for this difference. In any case, it is worth noting that the change in ribose puckering does not lead to significant modification in the overall backbone nor in the relative position of the bases. This suggests that, in contrast to what is generally assumed, sugar puckering may not be a key parameter in the determination of general helical structure, at least in these cases. Overall, our results are in general agreement with the idea that PNA strands are very flexible and that in DNA·PNA and RNA·PNA hybrids the PNA strands generally adapt their structure to accommodate the conformational preferences of the DNA or RNA strand.

Hydrogen-Bonding. All the Watson Crick and Hoogsteen H-bonds are preserved during the three simulations, except those binding C_{10} to the 5′-terminal G (see above). As noted above, the agreement between these results and the available X-ray data is remarkable. The crystal structure is also characterized by a second set of hydrogen bonds between the amide hydrogens of the H−PNA strand and phosphate oxygens of the DNA strand. This is generally not maintained in our simulations. The B- and A-trajectories show no backbone−backbone hydrogen-bonds, and the potential H-bond donors/acceptors (which are still close in space) are mainly exposed to the solvent. At the beginning of the P-trajectory the phosphate (DNA)-amido (Hoogsteen PNA) H-bonds found in the crystal are of course present, but during the trajectory most of them are broken leading to a structure in two sections with distinct characteristics (see Figure 4). The first half of the triplex exhibits a situation identical to that of B- and A-type simulations, without any clear intra- or interstrand H-bonds. The second half of the triple helix in the P-trajectory shows one interstrand P(DNA)-amido(H−PNA) H-bond (see Figure 5(top)), but the rest have been

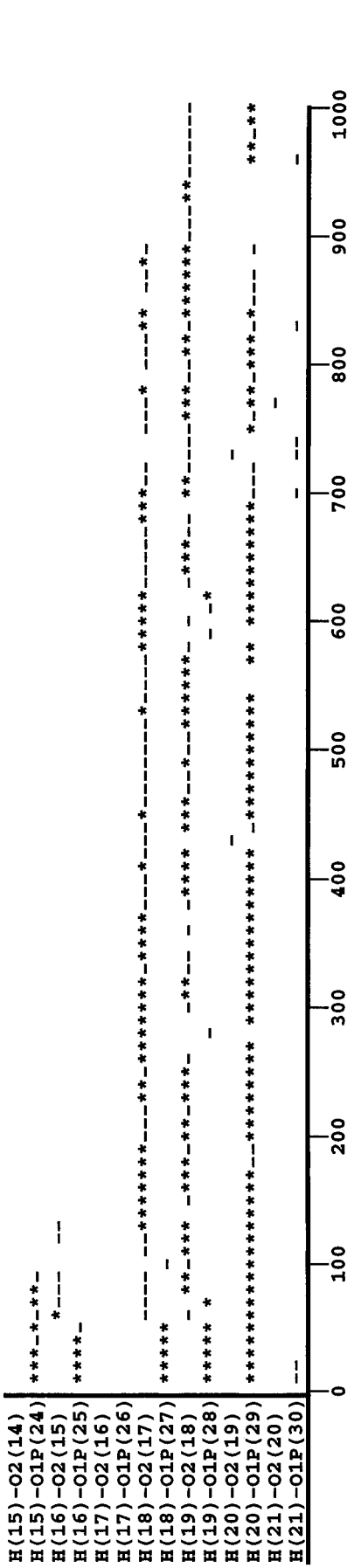

Figure 4. Time evolution of the H-bonding pattern in the minor−major groove over the P-trajectory. Asterisks indicate H-bonding distances less than 2.5 Å. Dashes indicate distances between 2.5 and 3.0 Å. Only the unrestrained period of the trajectory is included.

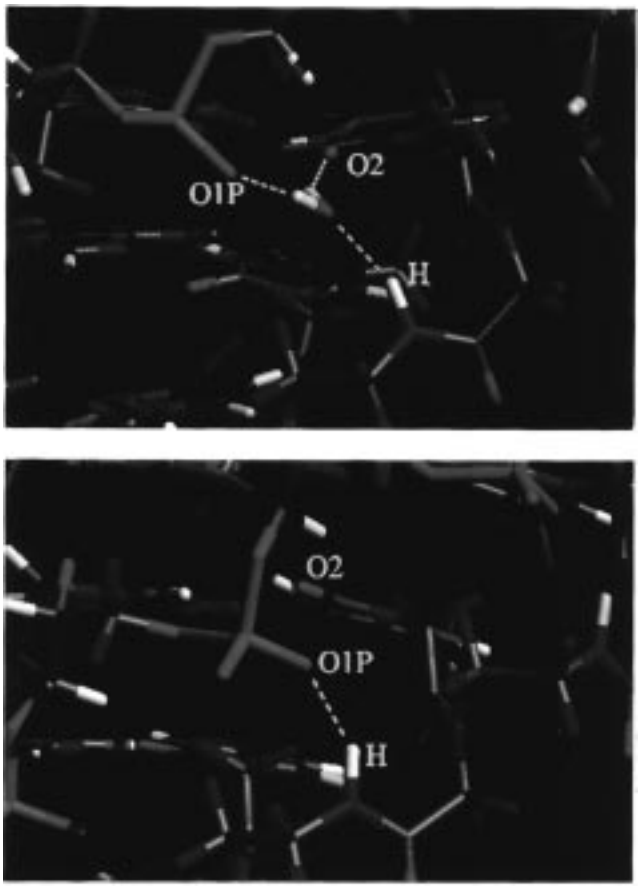

Figure 5. Snapshots taken from the P-trajectory. Top: example of water-mediated hydrogen bonding between amide NH, phosphate O1P, and pyrimidine O2 atoms in the minor−major groove. All H-bonding distances are around 1.9 Å. Bottom: example of direct H-bond between amide NH and phosphate O1P atoms. This is the type of interaction found in the crystal.

replaced by solvent-separated contacts (see Figure 5(bottom)), as previously found in other triplex simulations[15c] or by intrastrand H-bonds between the O2 of pyrimidine bases and the (i+1) amido group in the H−PNA strand. In all cases the change or loss of backbone−backbone H-bonds occurs through local movements which have only minor impact on the global helical structure but which can lead to a certain increase in the width of the groove defined by the DNA and H−PNA strands (see below).

According to all the MD simulations the pattern of H-bond interactions between H−PNA amides and DNA phosphates is not particularly stable in aqueous solution. The reason lies clearly in the competition of water molecules for the solvation of the amido hydrogens and, to an even greater extent, the phosphate groups (see below). In any case, it is clear that the pattern of H-bond interactions found in the crystal is not the only possibility, and what is more relevant, the backbone−backbone H-bonds are not crucial to mantain the overall structure of the helix, as suggested from X-ray data; it seems that water-mediated contacts can efficiently stabilize the triplex in their place. This hypothesis is indirectly supported by the fact that interstrand H-bonds involving backbones have not been detected in any other PNA or PNA·DNA(RNA) hybrid structures.[28−30]

Stacking Interactions. Both X-ray experiments and MD simulations indicate that the PNA·DNA·PNA triplex exhibits twist values (20−23 degrees), which are around 6−9 degrees smaller than that expected for triplex DNA.[8−12] Interestingly,

Table 5. Intrastrand, Interstrand, and Total Stacking Energies for the Central Portion (First and Last Triplex Steeps Were Removed, See Text) of the Triple Helix, for Different Base Triplex Arrangements[a]

structure	E_{stack}(intra)	E_{stack}(inter)	E_{stack}(total)
A-type DNA triplex	−128	−32	−160
B-type DNA triplex	−135	−34	−169
PNA·DNA·PNA crystal	−107	−48	−155
PNA·DNA·PNA (⟨A⟩)	−120	−44	−164
PNA·DNA·PNA (⟨B⟩)	−134	−34	−168
PNA·DNA·PNA (⟨P⟩)	−112	−48	−160

[a] All the values are in kcal/mol.

comparing the experimental structures of DNA·DNA and DNA·PNA duplexes[30], a similar difference of twist is found: twist for DNA·DNA is around 36 degrees, and twist for DNA·PNA is around 28 degrees. It seems then that the presence of PNA strand leads to an unwinding of around 6−9 degrees in the structure of the helix with respect to the situation found for a pure DNA polymer. This is probably due to an intrinsic lower twist of the PNA strands as noted in the very small twist (19.8 degrees) found in the high resolution (1.7 Å) X-ray structure of the PNA duplex. The PNA-induced unwinding of DNA polymers occurs without alteration of the rise, which is always similar to that in pure DNA polymers. It seems then that PNA is flexible enough as to adapt its structure to the DNA or RNA template, but in all cases it induces significant changes in the helical twist of the polymer. It is interesting then to determine if such alterations are due to (i) intrinsic characteristics of the backbone which to some extent force the twist of the bases or (ii) to the high flexibility of the PNA backbone which allows the bases to adopt a twist maximizing their stacking interactions.

Table 5 contains the average base pair stacking interaction for the central part (steps 2−7) of the PNA·DNA·PNA triplex in the crystal conformation, in each of the average conformations found during the trajectories as well as in the conformations corresponding to a DNA(t) triplex in the A- and B-models. Results show that in all the cases stacking interactions are favorable, due mainly to the stength of dispersion interactions. The intramolecular term is the most important in all cases, but the balance between intra- and interterms can change depending upon the structure. All the stacking energies are very similar, since the largest difference accounts for less than 10% of the total stacking energy. The crystal conformation shows the worst stacking interactions, probably because it is not fully energy-refined. However, the differences are small and probably lie within the error of the calculation method. What is clear is that there is no detectable advantage in terms of stacking due to the PNA-induced twist. This allows us to suggest that the twist of the PNA-containing helices does not maximize the stacking between bases and that the lower twist found in PNA-containing helices is due to the intrinsic conformational properties of the PNA backbone, as probably the twist found in the DNA and RNA is due to the intrinsic conformational characteristics of the phospho-ribose backbone.

Groove Widths. Triplexes define three grooves (see Figure 6), which according to our previous nomenclature for triplex DNA can be named as minor groove (m-groove), minor-major groove (mM-groove), and major−major groove (MM-groove). The definition of the grooves is difficult, since the usual rules for DNA structures are not valid in DNA−PNA hybrids. Inspection of the structures lead us to define the grooves using the following distances: C4′ (DNA strand)-CD(WC−PNA strand) for the minor groove; P(DNA strand)-CD(H−PNA strand) for the mM groove, and CA(WC−PNA strand)-CD-(H−PNA strand) for the MM groove. According to these criteria the MM groove is very wide (more than 20 Å), while the m and mM grooves are narrow (see Table 6). As expected, the structures sampled during the P-trajectory have a split character in terms of mM groove size: the first half of the helix show a mM groove width identical to that found in the B- and A-trajectory, while in the second half the mM groove is narrower due to the existence of intra- and interstrand H-bonds. Very interestingly, the change in the width of the mM groove occurs without significant alteration of the dimensions of the other grooves. The MM groove is wide enough to accommodate big molecules such as proteins, while the m-groove, and perhaps the mM groove, could potentially interact with small molecules akin to DNA minor groove binding drugs.

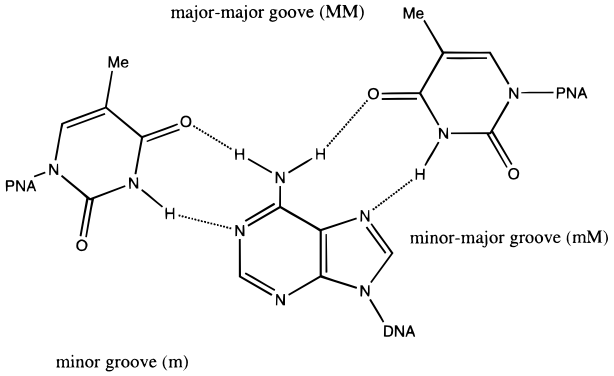

Figure 6. Definition of triplex grooves.

Table 6. Groove Sizes for the PNA·DNA·PNA Triplex in the Three Trajectories[a]

trajectory	m-groove	mM-groove	MM-groove
A-	7.7	6.7	23.8
B-	7.8	6.6	23.9
P-	7.7	5.4	23.9
first half		6.6	
last half		4.2	

[a] Values in Å. For groove definition see text.

Solvation. The triplex PNA·DNA·PNA structure is well solvated, even though the apparent density of water around the molecule is smaller (around a half) than that found for triplex DNA.[12] This is not surprising, considering the smaller density of charge in PNA·DNA·PNA triplexes. Integration of water occupancy allowed us to determine preferential hydration sites, which correspond to regions of large apparent water density. One of these regions is located inside the minor groove (Figure 8) and corresponds to strands of highly structured water. These results agree very well with the X-ray data for the PNA·DNA·PNA triple helix and also with the spine of hydration found in DNA duplexes. Detailed inspection of the MD simulation data (see Figure 8) suggests that the minor groove is better hydrated in AT regions than in GC regions, but even at GC steps the minor groove is preferentially solvated with respect to the bulk. This finding does not support the view derived from the crystal data on the PNA·DNA·PNA triplex that the string of ordered waters in the minor groove is broken at GC steps but agrees with high-resolution X-ray data for duplex DNA suggesting that the spine of hydration in the minor groove is weaker but still exists at GC steps.[31]

Another well hydrated region is the mM groove where, for A- and B-trajectories, the water density is similar to that found

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

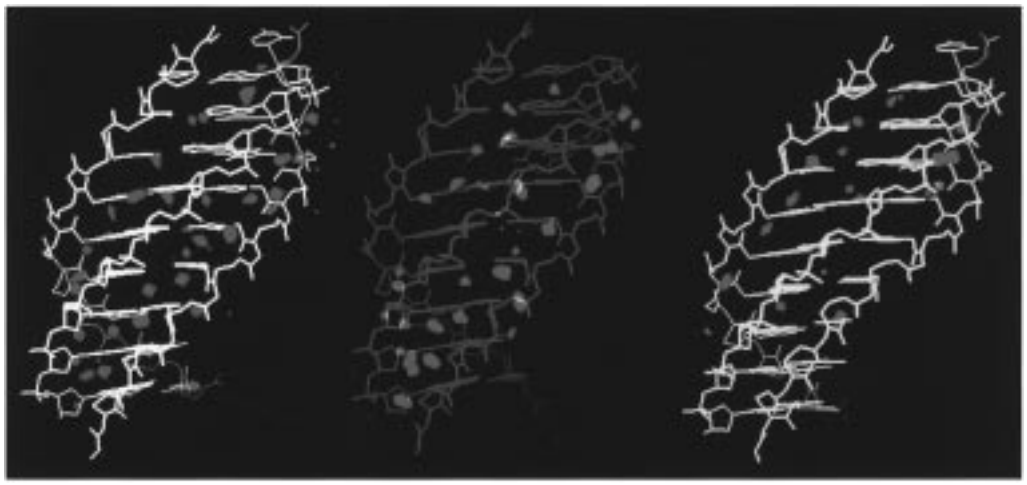

Figure 7. Water density contour plots for the average A (white), B (green), and P (yellow) structures. The contour level corresponds to an apparent water density four times that of pure water.

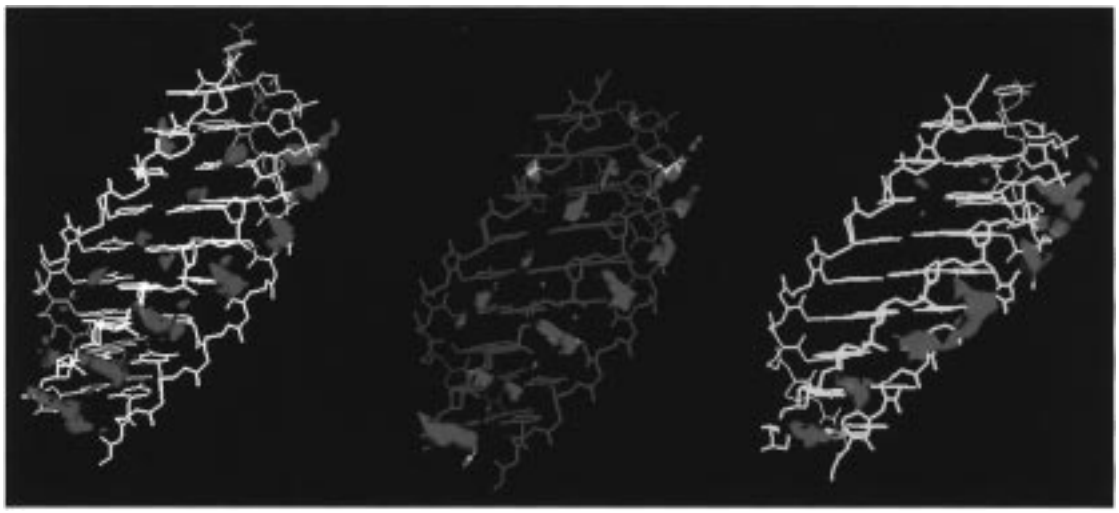

Figure 8. MIP plots or the average A (white), B (green), and P (yellow) structures. The contour shown corresponds to −3 kcal/mol.

in the m groove. In contrast, hydration of the mM groove is not as good in the P-trajectory, especially in the second part of the helix due to the narrowing of the mM groove. Inspection of the density plots reveals that water molecules in the mM groove are more on the surface of the groove than those in the m groove and are H-bonded to the O2 of pyrimidines in the Hoogsteen PNA strand as well as with the phosphates of the DNA strand. These results suggest that in contrast to the situation found in the crystal the mM groove can be well hydrated in aqueous solution by highly structured water molecules. Such solvation leads to the breaking of the interstrand P(DNA)-NH(H−PNA) H-bonds, since the water molecules trapped inside this narrow groove seem to bridge the phosphate and amino groups. Finally, small regions of high water density are located in the MM groove in the GCC steps near the amino groups of PNA cytosines. These regions must result from the presence of the protonated Hoogsteen cytosines, which should generate a strong local positive potential in these regions. The remainder of the MM groove is not preferentially hydrated, but the water density is similar to that of the bulk solvent.

Potential Interactions with Small Molecules and Proteins. Due to the size of the grooves in the PNA·DNA·PNA triplex this molecule should be able to interact with groove binders such as drugs or proteins, though such interactions have not been yet described. A simple way to detect regions of potential interaction with drugs or proteins is to analyze the regions of more negative molecular interaction potential (MIP) with positive charges. The MIP around the PNA·DNA·PNA helix shows contours as shown in Figure 8, which indicate two regions of favorable interaction with a small cation. The first region is located mainly around the phosphates of the DNA strand and extends slightly into the mM groove. The second region is somewhat less negative and appears located in the minor groove of the triplex. As expected, the MIP values are clearly smaller (around 2−3 kcal/mol in absolute values) than those found for triplex or duplex DNA. On this basis one would expect less ability for the PNA·DNA·PNA triplex to interact with cations than for parent DNA triplexes, and a smaller dependence of its structure on ionic strength, in agreement with experimental data.[3]

Conclusions

The obvious aim of molecular modeling studies is to provide information not readily accessible through experimental techniques. Before any faith can be put in such information, it is necessary to demonstrate that the theoretical studies are capable of reproducing results which have been experimentally determined. Here we have shown that an approach based on extended molecular dynamics simulations gives information on the structural features of the PNA·DNA·PNA triplex which agree very closely with what has been ascertained through experiment. Venturing beyond what has been experimentally

determined, we conclude first that the particular interstrand hydrogen bonding pattern observed in the crystal structure of the triplex may not be present in aqueous solution but that the loss of this interaction is not likely to have any detrimental effect on the stability of the triplex or result in any gross changes in structure. Our second main conclusion concerns the role of the PNA backbone in the structure of its complexes with DNA and RNA. We conclude that while the PNA backbone is much more conformationally flexible than that of DNA or RNA, it is not "neutral". The observed decrease in helical twist of 6−9 degrees that appears to accompany the replacement of a DNA or RNA strand by its PNA counterpart is not the result of a passive response of this strand to increased stacking interactions accompanying a reduced twist. Rather the stacking interactions are invariant, and the reduction in twist is driven by the active preference of the PNA strand. Our third conclusion concerns the possibilities for molecular recognition of the PNA·DNA·PNA triplex and relates to our first conclusion. The loss of inter-backbone H-bonding results in a pattern of hydration of the mM groove quite different from that observed in the crystal structure, and the MIP analysis suggests that this groove might now represent a very reasonable target for recognition by molecules structurally related to the well-known DNA minor groove binders, despite its lower electronegativity compared with the mM groove of pure DNA structures.

Acknowledgment. We thank Prof. F. J. Luque for many helpfuls suggestions and valuable comments on this manuscript. We also thank the technical support of Dr. J. L. Gelpi. G.S. thanks the Spanish Ministry of Science for support during his sabbatical period in Barcelona. This work was supported by the Spanish DGICYT (PB93-0779) and the MEC−British Council program for Spanish−British collaborative projects.

JA9723444