Molecular Dynamics Simulations of the d(T·A·T) Triple Helix

George C. Shields,*,†,§ Charles A. Laughton,*,‡ and Modesto Orozco*,†

Contribution from the Departament de Bioquímica i Biologia 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, UK

Received February 24, 1997[⊗]

Abstract: Molecular dynamics (MD) simulations have been used to study the dynamical and time-averaged characteristics of the DNA triple helix $d(T)_{10}\cdot d(A)_{10}\cdot d(T)_{10}$. The structures sampled during the trajectory resemble closely the B-type model for the DNA triplex proposed on the basis of NMR data, although there are some subtle differences. Alternative P- and A-type conformations for the triplex, suggested from X-ray experiments, are not predicted to contribute significantly to the structure of the DNA triplex in solution. Comparison with the best available experimental data supports the correctnes of the MD-generated structures. The analysis of the collected data gives a detailed picture of the characteristics of triple-helix DNA. A new and interesting pattern of hydration, specific for triplex DNA, is an important observation. The results suggest that molecular dynamics can be useful for the study of novel nucleic acid structures.

Introduction

Nucleic acid triple helices are important structures with a broad range of potential applications in gene therapy.[1] One of the most important classes of triplex is based on the "pyrimidine motif" in which a pyrimidine-rich third strand lies parallel to and interacts with the purine strand in a Watson−Crick duplex. The interactions and strand senses have led to this being termed the $Y^+\cdot R^-\cdot Y^-$ type triplex. The determination of the three-dimensional structure of DNA triple helices has been one of the challenges for structural biochemists in the last decade, but despite the many efforts directed toward this goal there is not a generally accepted model for the $Y^+\cdot R^-\cdot Y^-$ triplex. The first structural description was based on X-ray fiber diffraction data.[2] Since then, more information has been gained from fiber diffraction, and spectroscopic studies have been performed on various triplex sequences.[3−5] None of these studies have provided atomic resolution data, but have led to a certain amount of conflicting evidence as to the structure of triple-helical DNA.

Some authors, on the basis of X-ray fiber difraction results, have suggested that triplex DNA adopts a structure resembling the A-type conformation of duplex DNA.[2] In this structure the riboses are in the C3′-endo conformation, the base pairs are inclinated around 8.5° from the helix axis, and the Watson−Crick (WC) base pairs show a large displacement from the helix axis. The rise is around 3.3 Å, and the helical twist around 30°, leading to a helical periodicity of 12 base pairs per turn. Other authors,[6] on the basis of NMR and IR data,[5,7] have supported a structure for the DNA(t) resembling that of B-type duplex DNA. In this model the riboses are C2′-endo, the base pairs are almost perpendicular to the helix axis, the displacement of the WC base pairs from the helix axis is less than in the A-form model, the rise is around 3.4 Å, and the helical twist is between 27° and 32°, leading again to a helical periodicity of around 12 bases per turn. Finally, the recent determination of the structure of the PNA·DNA·PNA triple helix by high-resolution X-ray diffraction[8] has added a new possible candidate to be the structure of triplex DNA, the so called P-type conformation. The P-type helix is completely different from A- and B-type helices: this structure shows C3′-endo ribose puckerings, a very large displacement of the WC base pairs from the helix axis, an average rise of 3.4 Å, and an average helical twist of 23°, giving it a helical periodicity of 15 bases per turn.

Molecular dynamics (MD) simulations have proven to be an excellent adjunct to experimental techniques for the determination of protein structures,[9,10] but their use for the study of nucleic acids has been hampered by technical problems in the simulations. Very recently, the development of new force fields and of methods for the accurate representation of long range effects have made MD techniques extremely powerful for the study of DNA structures as well.[11−13] Thus, in a recent series of papers Kollman and co-workers have shown that extended MD simulations of solvated systems are able to reproduce

* Address correspondence to any of the authors.
† Universitat de Barcelona.
‡ University of Nottingham.
§ On leave from Lake Forest College, Lake Forest, IL 60045.
⊗ Abstract published in *Advance ACS Abstracts,* July 15, 1997.

(1) (a) Strobel, S. A.; Dervan, P. *Methods Enzymol.* **1992**, *216*, 309. (b) Grigoriev, M.; Praseuth, D.; Guieysee, A. L.; Robin, P.; Thuong, N. T.; Helene, C.; Harel-Bellan, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3501. (c) Sun, J. S.; Garestier, T.; Helene, C. *Curr. Opin. Struct. Biol.* **1996**, *6*, 327. (d) Agarwal, B. B.; Schwarz, L.; Hogan, M. E.; Rando, R. F. **1996**, *56*, 5156.
(2) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. C. *Nucl. Acids Res.* **1976**, *11*, 4141.
(3) Radhakrishnan, I.; Patel, D. J. *Biochemistry* **1994**, *133*, 1405.
(4) Dagneaux, C.; Liquier, J.; Taillandier, E. *Biochemistry* **1995**, *34*, 14815.
(5) Macaya, R. F.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1992**, *114*, 781.
(6) Raghunathan, G.; Miles, H. T.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 455.
(7) Howard, F. B.; Miles, H. T.; Liu, K.; Frazier, J.; Raghunathan, G.; Sasisekharan, V. *Biochemistry* **1993**, *31*, 10671.
(8) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. *Science* **1995**, *270*, 1838.
(9) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.
(10) York, D. M.; Woldawer, A.; Petersen, L.; Darden, T. *Proc. Natl. Acad. Sci U.S.A.* **1994**, *91*, 8715.
(11) Laaksonen, A.; Nilsson, L. G.; Jonsson, B.; Teleman, O. *Chem. Phys.* **1989**, *129*, 175.
(12) (a) Mohan, V.; Smith, P. E.; Pettit, M. *J. Am. Chem. Soc.* **1993**, *115*, 9297. (b) Mohan, V.; Smith, P. E.; Pettit, M. *J. Phys. Chem.*. **1993**, *97*, 12984. (c) Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y.-K.; Pettit, B. M. *J. Am. Chem. Soc.* **1995**, *117*, 2147.
(13) York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. *J. Am. Chem. Soc.* **1995**, *117*, 5001.

S0002-7863(97)00601-X CCC: $14.00

accurately the crystal structure of B-form duplex DNA,[14,15] a result that we have also found in control simulations in our laboratory. Furthermore, they have shown that high-level simulations can capture the difference in stability between A- and B-conformations of duplex DNA, so that the conformational transition from A- to B-form DNA can be modeled. Very recently, we have found that MD simulations can reproduce the three dimensional structure of the PNA·DNA·PNA triplex in aqueous solution, even when trajectories are started from incorrect conformations (work in progress). These results suggest that extended MD trajectories can capture certain transitions between different DNA conformations, allowing a more complete sampling of DNA conformational space. This strategy can then be used then to obtain *ab initio* structures of new forms of DNA for which experimental information is incomplete, or simply does not exist.

In this paper we summarize a systematic exploration of the conformational space of the d(T·A·T) triple helix, including the three conformations suggested experimentally for this molecule (A-, B-, and P-type). This study has led to the definition of an average structure of the poly-d(T·A·T) triple helix in aqueous solution, providing atomic-level information for the triple-helix structure which agrees well with the most recent available experimental data.

Methods

Three 1.155 ns MD trajectories have been performed for the $d(T{\cdot}A{\cdot}T)_{10}$ sequence. The first one started from an A-type conformation obtained from X-ray diffraction data.[2] The second one started from a B-type conformation derived by modeling from early NMR and IR data.[6] Finally, the third trajectory started from the helical parameters of a P-type structure derived from the X-ray data of the PNA·DNA·PNA triple helix.[8] All the starting structures were neutralized by adding 27 Na^+ ions, which were placed in the most eletronegative regions according to Poisson−Boltzman calculations. A total of 4135 water molecules were added using standard AMBER4.1 rules to hydrate the system.[17] Systems were optimized, heated slowly, and equilibrated carefully for 155 ps, before 1 ns of MD simulation without any constraints.[18] The careful equilibration procedure was designed to avoid artifactual transitions from the starting models[18] during the trajectories.

Simulations were performed in the isothermic isobaric enssemble (P = 1 atm, T = 300 K). Periodic boundary conditions and the Particle−Mesh−Ewald algorithm were used.[14] A 2 fs integration time step was used, and all bond distances were frozen using SHAKE.[19] The AMBER-95 force field was used to represent the DNA,[20] and the TIP3P model was used to represent water molecules.[21] MD calculations were carried out with the AMBER4.1 program.[17]

The last 500 ps of the B-type structure were used to generate an average structure for the triplex. For this purpose the coordinates of 1000 snapshots were averaged after least-squares fitting of the DNA atoms. Coordinate averaging in this way often results in structures with poor geometry (e.g., very short C−H bonds in thymine methyl groups), so the structure was then subjected to energy minimization, while all heavy atoms were restrained to their initial positions with a force constant of 100 kcal mol^{-1} $Å^{-1}$.

Molecular interaction potentials (MIPs) were used to obtain a picture of the reactive characteristics of triplex DNA. In this case we use a MIP defined as an electrostatic component plus a Lennard-Jones contribution. The electrostatic component was computed as the interaction with an unit positive charge, determined by solving the nonlinear Poisson−Boltzmann equation.[16] These calculations were performed using an ionic strength of 1 M, a dielectric constant of 2 for the DNA, and 80 for the water, and all the default values in DELPHI compute program.[16] The Lennard-Jones contribution was determined using the standard all-atom OPLS parameters for the DNA, and the TIP3P parameters for the probe molecule.

All calculations were carried out in the IBM SP2 of the Centre de Supercomputació de Catalunya (CESCA) and in SGI workstations in our laboratory.

Results and Discussion

After equilibration the three trajectories seem reasonably stable in terms of density, temperature, potential energy, and other macroscopic properties. The total number of Watson−Crick (WC) and Hoogsteen (H) hydrogen bonds is preserved throughout all the simulations, even though no constraints are introduced to reinforce them, and so the integrity of the three helices is maintained. This result demonstrates that our simulation conditions are robust enough to preserve the integrity of DNA and avoid artifactual unfoldings of the type observed in early MD studies of DNA.[22] Present simulations are a clear improvement with respect with pioneering MD simulations in DNA triplexes.[23] A detailed inspection of the RMS plots (from starting conformations, and computed always over all DNA atoms) displayed in Figure 1a shows that trajectories starting from A-type and P-type conformations (the A-and P-type trajectories) clearly diverge from their respective starting conformations. The divergence in the P-type trajectory is very large ($\langle RMS \rangle$ = 3.7 Å), and fast (transition time less than 0.1 ns), suggesting that the P-type structure does not contribute at all to the structure of the DNA triple helix in solution. The divergence in the A-type trajectory is slower (transition time around 0.4 ns), and smaller ($\langle RMS \rangle$ = 2.2 Å), which suggests that the A-type is more similar than the P-type to the structure of the poly(T·A·T) triple helix in solution.

The center panel of Figure 1a shows that the divergence in the B-type trajectory with respect to the starting B-type conformation is small ($\langle RMS \rangle$ = 1.4 Å) and that the RMS reaches a plateau after 150−200 ps. This means that after a fast and slight reorganization from the starting conformation the B-type trajectory reaches an equilibrium structure, which can then be characterized by averaging over the last 0.8 ns of the trajectory. The average of the structures sampled over these 0.8 ns of equilibrium trajectory is denoted as Bav structure (see Methods).

The RMS deviations from the starting B-type conformation of the P- and A-trajectories are also displayed in Figure 1a. It is clear that both P- and A-type trajectories are converging to a common structure resembling the B-type conformation. As expected, such a convergence is faster for the P-type (less than

(14) Cheetham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.
(15) Cheetham, T. E.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434.
(16) *DELPHI Computer Program*, Biosym: San Diego, CA, 1994.
(17) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Fergurson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1.
(18) The equilibration of the system consists of the following consecutive steps: (i) 10 ps MD (T = 100 K) of water, (ii) minimize water, (iii) minimize all system, (iv) 10 ps MD (T = 100 K) of system with restrains in DNA (K = 100 kcal mol^{-1} $Å^{-1}$), (v) as iv but T = 300 K, (vi) 25 ps MD (T = 300 K) of system with restrains in DNA (K = 50 kcal mol^{-1} $Å^{-1}$), (vii) as vi but K = 25 kcal mol^{-1} $Å^{-1}$, (viii) as vii but K = 10 kcal mol^{-1} $Å^{-1}$, (ix) as viii but K = 5 kcal mol^{-1} $Å^{-1}$, (x) as ix but K = 0.2 kcal mol^{-1} $Å^{-1}$. The final system was the starting point for 1.0 ns of unrestrained MD simulation (T = 300 K).
(19) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327 (1977).
(20) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(21) Jorgensen, W. L.; Chandresekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(22) (a) Swaminathan, S.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5027. (b) Miaskiewicz, K.; Osman, R.; Weinstein, H. *J. Am. Chem. Soc.* **1993**, *115*, 1526.
(23) (a) Hausheer, F. H.; Singh, U. C.; Palmer, T. C.; Saxe, J. D. *J. Am. Chem. Soc.* **1990**, *112*, 9468. (b) Laughton, C. A.; Neidle, S. *J. Chim. Phys.* **1991**, *88*, 2597. (c) Laughton, C. A.; Neidle, S. *J. Mol. Biol.* **1992**, *223*, 519. (d) Laughton, C. A.; Neidle, S. *Nuc. Acids Res.* **1992**, *20*, 6535. (e) Laughton, C. A.; Neidle, S. *Mol. Simul.* **1995**, *14*, 275.

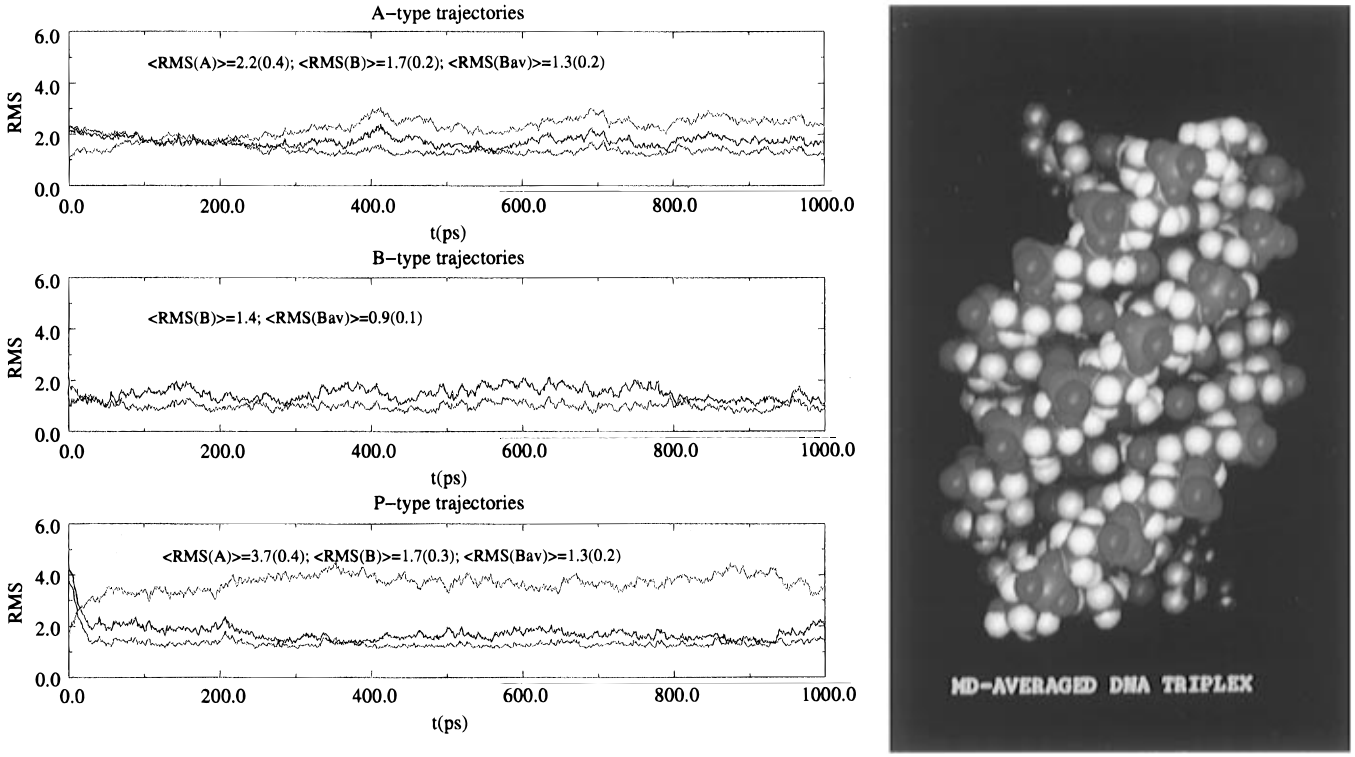


Figure 1. (a, left) Root mean square (RMS in Å) deviation for A-, B-, and P-type trajectories: [(green) RMS *vs* the initial (A or P) structure; (black) RMS *vs* the initial B-type structure; (red) RMS *vs* the Bav structure] and (b, right) average structure obtained by MD simulations.

Table 1. Values of the Different Helical Parameters for the Average Conformation and MD-Averaged Values (in Parentheses)[a]

reference	type of movement	parameters	MD values	fiber	NMR
axis	base-pair translation	*X*-displacement	−3.5 Å (−2.8 ± 0.7 Å)	−3.2 Å	−2.0 Å
axis	base-pair rotation	tip	−1.3° (2.1 ± 4.0°)	small	small
axis	base-pair rotation	inclination	−2.8° (−7.9 ± 8.0°)	−8.5°	small
base pair	base-pair translation	slide, shift	0.0 Å (0.0 ± 0.1°)	small	small
base pair	base-pair translation	rise	3.5 Å (3.5 ± 0.1 Å)	3.3 Å	3.4 Å
base pair	base-pair rotation	twist	29.3° (29.0 ± 0.9°)	30°	27−32°
base pair	base-pair rotation	roll	1.0° (1.9 ± 1.2°)	small	small
base pair	base-pair rotation	tilt	−0.3° (2.2 ± 2.0°)	small	small
base	base−base translation	stagger	−0.2 Å (−0.2 ± 0.1 Å)		
base	base−base translation	stretch, shear	0.1 Å (0.2 ± 0.1 Å)		
base	base−base rotation	opening	2.2° (2.6 ± 2.2°)		
base	base−base rotation	buckle	−2.3° (−2.1 ± 3.6°)		
base	base−base rotation	propeller twist	−8.5° (−7.3 ± 4.5°)		

[a] Standard deviations (SD) of the MD-averaged parameters are also displayed. All the helical values refers to the WC strands. Experimental estimates from fiber difraction and NMR are also displayed (if available) for comparison purposes.

0.1 ns) than for A-type (around 0.5 ns) trajectories. It is worth noting that the transition to the B-type conformation mirrors the divergence from the starting structures in both A- and P-trajectories. This increases our confidence in the reliability of these transitions, which cannot be explained as artifacts of the simulations. Even more convincing is the RMS deviation for P-type and A-type trajectories taking the Bav structure as reference (Figure 1a). In this case, all the simulations converge to the Bav structure with RMS deviations of only 0.9 (B-type trajectory), and 1.3 Å (P- and A-trajectories). In other words, three extended MD simulations starting from different reasonable conformations are spontaneously converging to the same average structure, without the introduction of any restraints.

The Bav conformation is displayed in Figure 1b, and its main helical characteristics are given in Table 1, where the MD averaged values of the different helical parameters and the experimentally available values are also displayed. As expected, the helical parameters for the average structure are not identical to the MD-averaged values, the largest discrepancies being found for very flexible parameters (see below), and for parameters depending on helix axis. However, despite slight numerical differences for some parameters the general agreement betwen the helical parameters for the average structure and the MD-averaged values is very good, which demonstrates the ability of Bav to represent the average of the ensemble of structures sampled during the MD trajectory.

The DNA(t) helix is regular, with little bending, as evidenced by a shortening of only 0.03%. Due to the presence of the third strand, the helix axis is displaced from the center of the WC pair to a position located between the WC and the H base pairs, which leads to a moderate X displacement of the WC base pairs (around −3 Å). As expected for a B-type DNA the base pairs are perpendicular to the helix axis as noted by the inclination and tip values (see Table 1). Inspection of global base-base parameters shows very little disruptions of the planarity of the base pairs, as noted in the average buckle (around 2°) , opening (around 2°), and propeller twist (around 8°) for the WC strand. These are not far away from those found for B-DNA duplexes.[24]

Calculation of the global inter-base pair-parameters for Bav reveals (Table 1) very small values of tilt and roll, confirming the planarity of the base pairs with respect to the axis and to

(24) Chandrasekaran, R.; Arnott, S. in *Landolt-Börstein, New Series*, Saenger, W., Ed.; Springer-Verlag: Berlin, 1989; Group VII, Vol. 1B, p 68.

the preceding base pair. This finding is in agreement with the behavior expected for a B-type duplex DNA. The rise of the average structure is around 3.4−3.5 Å in both WC and H strands, matching the rise value found in duplex B-DNA. The twist of the Bav is around 29°, which leads to a periodicity of 12 base pairs per turn. This sharply constrasts with the average twist of duplex B-DNA (around 36°), and the periodicity of 10 bases per turn. The entirety of the results suggests that the formation of the triple helix leads to a certain unwinding in the double helix, but not to an increase in the distance between base pairs.

The analysis of the Bav structure, and of the trajectories sampled during the simulation, shows that puckering amplitudes range from 36° to 41°, as expected for nucleic acid structures. The phase angle of the ribose ranges typically from 100° to 140° for the WC strands, while for the Hoogsteen strand the ribose phase angle typically varies between 90° and 120°. This means that most of the riboses are in the south to south-east regions. We found the same results for all the simulations (once the trajectories were stable), irrespective of starting conformation. Three different conformers contribute to the sugar puckering: C2′-endo, C1′-exo and O4′-endo. The most populated conformer is C1'-exo for the WC strands, while C1′-exo and O4′-endo are equally populated for the Hoogsten strand. This puckering is characteristic of a B-type DNA(d), where high-resolution X-ray studies have shown that C1′-exo is the most populated conformer.[14] Analysis of phase angle and puckering amplitudes reveals that sugar repuckering is one of the fastest movements in the DNA, occurring on the picosecond time scale and producing a very fast interconversion between south, south-east, and east conformations. In all cases the transitions occur by simple change of the phase angle, while the puckering amplitude remains fairly constant, as expected from pseudorotational theory.

Overall, the analyses of the trajectories reveal the high stability of the helix, which is preserved throughout the simulation. The only significant movement detected during the B-type trajectory involves the partial disruption of a Watson−Crick A-T pair in one of the ends of the triplex. This disruption happens 0.6 ns after equilibration and last for only 0.05−0.07 ns, before WC pairing is recovered. The partial disruption of the A-T WC interaction leads to very modest and local changes in a few helical parameters, which happen almost simultaneously. This shows that local fluctuations in DNA are due to combined movements, rather than to one or two key canonical movements. MD simulations suggest then that H-bonding in triplex DNA is very stable, and a local disruption of one base pair does not lead to the disruption of additional H-bonds. Furthermore, the breaking of a H-bond is probably slow, but the re-forming of such a H-bond must be fast, probably on the time scale of 0.05−0.1 ns.

Inspection of helical parameter fluctuations (see Table 1) during the trajectory gives us additional insight into amplitude of the several of the fastest movements in triplex DNA. All the different helical parameters show random variations without any clear transitions throughout the simulation time of the trajectory. It is clear that the most flexible parameters are those referring to movements of the base pairs with respect to the helix axis. Thus, standard deviations in the X- and Y-displacements are in the range of 0.6−0.7 Å, and the tip and inclination rotations exhibit standard fluctuations around 4−5°. On the contrary, the relative position of one base pair with respect to the contiguous ones is quite rigid, particularly in terms of translations, as the values of slide, shift, and rise in Table 1 indicate. Finally, there is a certain degree of flexibility in the position of one WC-paired base with respect to the other. The most flexible movements are those that affect H-bonding the least, i.e. propeller twist and buckle, which show standard fluctiations in the range of 3−4° along the trajectory.

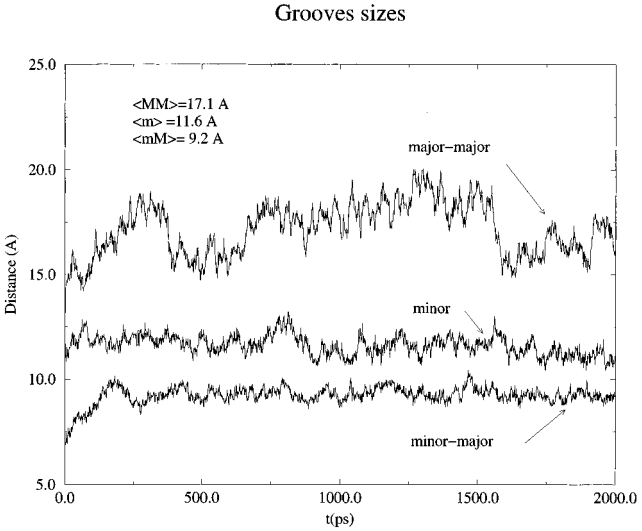
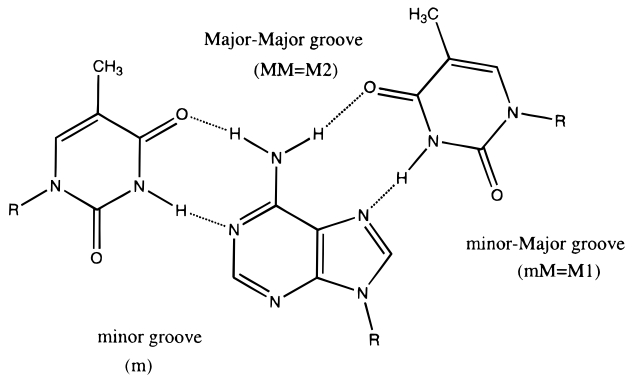


Figure 2. Grooves size (P−P distance in Å) during the B-type trajectory.

Chart 1. Schematic representation of a T·A·T triplex.



The minor groove (m-groove) of the Bav triplex structure is similar in width and depth to that of B-type duplex DNA and does not resemble at all the minor groove of an A-type or P-type triplex DNA. The shortest average distance between phosphates across this groove is around 11.6 Å throughout the trajectory (see Figure 2), which (assuming a radius of 2.9 Å for the phosphates) gives a groove width of 5.8 Å, matching typical values for B-type duplex DNA (5.9 Å).[25] However, this superficial similarity with a B-form duplex minor groove is somewhat misleading, as examination of space-filling models of the two structures shows that the groove geometry is rather different. While the B-form duplex minor groove has a U-shaped cross section, the triplex minor groove as a V-shaped cross section (see Figure 3a). This would help to explain the observation that while some established duplex minor-groove binding ligands such as netropsin[26] and distamycin[27] can bind to triple helices, they destabilize them as they bind more strongly to the duplex generated on loss of the triplex-forming strand. The major groove is quite different from that in standard B-type DNA, since the presence of the third strand widens the major groove and divides it into two asymmetric parts (see Chart 1): the minor part of the major groove (mM-groove; M1 groove according to Petitt's nomenclature), and the major part of the

(25) Dickerson, R. E. *Methods Enzymol.* **1992**, *211*, 67.
(26) Park, Y. W.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6653.
(27) Umemoto, U.; Sarma, M. H.; Gupta, G.; Luo, J.; Sarma, R. H. *J. Am. Chem. Soc.* **1990**, *112*, 4539.

major groove (MM-groove; M2 groove according to Petitt's nomenclature). The phosphate−phosphate distance across the MM groove (17.1 Å) indicates a groove width of 11.3 Å, which is comparable with the total width of the major groove of a B-DNA duplex (11.5 Å following the same measurement criterion). Inspection of the MM groove shows the existence of three possible H-bond donors and acceptors in the bottom of the groove: N6(A), O4(T-WC), and O4(T-Hoogsten). However all three are buried behind the two methyl groups of thymine at C5, which may impede (but does not preclude) their accesibility to interacting molecules. It is worth noting that the MM groove suggested by our MD model is wide enough to interact with suitable proteins and peptides and has very different requirements for sequence-specific recognition than the major groove of duplex B-DNA.

The minor half of the major groove (see Figure 2 and Chart 1) is of special interest. This groove is very narrow, and its existence is difficult to predict *a priori*. The bottom of this groove is defined by the O2 of the Hoogsteen-binding thymine, which gives hydrogen-bonding properties to this narrow groove. The average minimum phosphate−phosphate distance across this groove is 9.2 Å, which leads to a groove width of 3.4 Å. This mM-groove is then around 2.2 Å narrower than the minor groove, but wide enough to accomodate a small interacting molecule. Inspection of Figure 1b shows that the mM-groove expands slowly from values in the starting structure during the first 250 ps of the trajectory, reaching the final value which is maintained with negligible variations throughout the rest of the simulation. Inspection of the trajectory suggests that the improvement in the hydration of the mM-groove is probably responsible for this slight expansion of the mM-groove at the beginning of the trajectory (see below).

It is also interesting to analyze the grooves of triplex DNA in terms of molecular flexibility (see Figure 2). As noted before, both the m- and the mM-grooves are quite rigid throughout the simulation, with a standard fluctuation near 0.4 Å. We believe that the existence of highly structured water molecules (see below) in both grooves is responsible for this rigidity. It is suggested that removal of the spine of hydration from the m- and mM-grooves should increase groove flexibility, making it possible to adapt groove shape to specific groove binding drugs. Figure 2 shows that in contrast to the smaller grooves, the MM-groove is quite flexible (standard fluctuation 1.3 Å), making possible expansions and contractions of the groove greater than 2 Å from the average values. This large flexibility could be used by the DNA triplex to adapt its shape to possible interacting proteins. The transition time for these MM-groove breathing movements seems to be on the scale of a few hundreds picoseconds. Inspection of other helical parameters show that these groove movements are not clearly connected to any change in particular helical parameters, which implies that MM groove flexibility does not lead to major alterations in the the general shape of the triplex.

A picture of the reactive characteristics of triplex DNA can be obtained by using molecular interaction potentials (MIP, see Methods) for the average structure of the DNA triplex. The MIP contours in selected values are displayed in Figure 3b. It is worth noting the existence of a large concentration of negative potential along sterically permited regions within the three grooves. There are not big differences in MIP between the three grooves, which suggests that all three of them can be targets for interactions with positively charged groups, but MIP calculations suggest that the minor grooves are better targets than the MM groove (see Figure 3b).

Qualitative analysis of the water atmosphere around the DNA triple helix reveals that the DNA is very well hydrated, specially in the grooves, as expected from MIP values noted before. Furthermore, a quantitative analysis of water distributions around the DNA has allowed us to determine new features previously undescribed for triplex DNA, which may be considered as predictors of possible interactions of the DNA with small molecules and proteins. Thus, integrating water populations for the B-type trajectory in 1 $Å^3$ volume elements around the DNA, regions of preferential hydration can be detected. These regions correspond to areas where the apparent water density (defined as number of waters/unit of time × volume) is clearly larger than that of pure water, which indicates a negative free energy of transfer for these water molecules from pure water to these regions. The water-density contour levels shown in Figure 3c reveal the existence of extensive regions of high water density in the grooves, and smaller regions near the phosphates. The most extensive region of high water density is located in the minor groove, where long strands of highly stable waters are H-bonded to the N3(A) and O2(T-WC) atoms. The calculated water density in this region is 8−9 times the density of pure water, which represents a preferential solvation of 1.2−1.3 kcal/mol from pure water. These results suggest that a large spine of highly ordered water molecules is present in the minor groove of the d(T·A·T) triplex, which agree very well with the qualitative description of hydration in high-resolution X-ray studies of crystals of DNA duplexes.

Inspection of Figure 3c reveals an additional strand of highly stable water molecules in the mM-groove. These waters are trapped inside this narrow groove, in a similar manner to the waters in the minor groove and interact via hydrogen bonds with the O2 of the Hoogsteen strand thymine. The apparent density of water in this groove is six times the density of pure water (preferential free energy of hydration around 1 kcal/mol), which is slightly smaller than that of the minor groove, but still significantly larger than the density of pure water. This suggests that the minor part of the major groove in triplex DNA is wide and polar enough to interact with small polar molecules. Since most of the duplex DNA minor groove binders recognize regions where the spine of hydration is specially stable, we suggest that the mM groove can be a triplex-specific target for specially designed minor groove binders. We note here that the possible specific hydration of the mM-groove of the DNA(t) was previously hypothesized by Patel and co-workers on the basis of the analysis of a DNA(t) model recently derived from NMR data on a different sequence.[28] Finally, similar strands of water have been previously found in MD simulations of purine-motif triplexes.[12a,b]

The MM-groove is also very well hydrated, especially surrounding N6 of adenine and O4 of thymines. However, no regions of very high density of water are found; the greatest apparent density values are around 3 times the density of pure water, which represents a preferential hydration of ∼0.5 kcal/mol from pure water. The reasons for this smaller density of water in the major part of the major groove are probably related to the larger size of this groove and to the existence of the apolar methyl groups of the thymines inside the groove, which hamper the interaction of water with the N6 of adenines and the O4 of thymines. In fact, the MM-groove of the TAT triplex is an intersting amphipatic structure, with an outer hydrophobic part formed by the two methyls of the thymines, and an inner hydrophilic part formed by the polar groups N6 and O4 of adenine and thymine.

In summary, our MD simulations suggest well-defined spines of hydration in the m- and mM-groove, which should be detected by high-resolution X-ray experiments, but such a spine will be probably undetectable in the MM-groove. However,

(28) Radhakrishnan, I.; Patel, D. J. *Structure* **1994** *2*, 395.

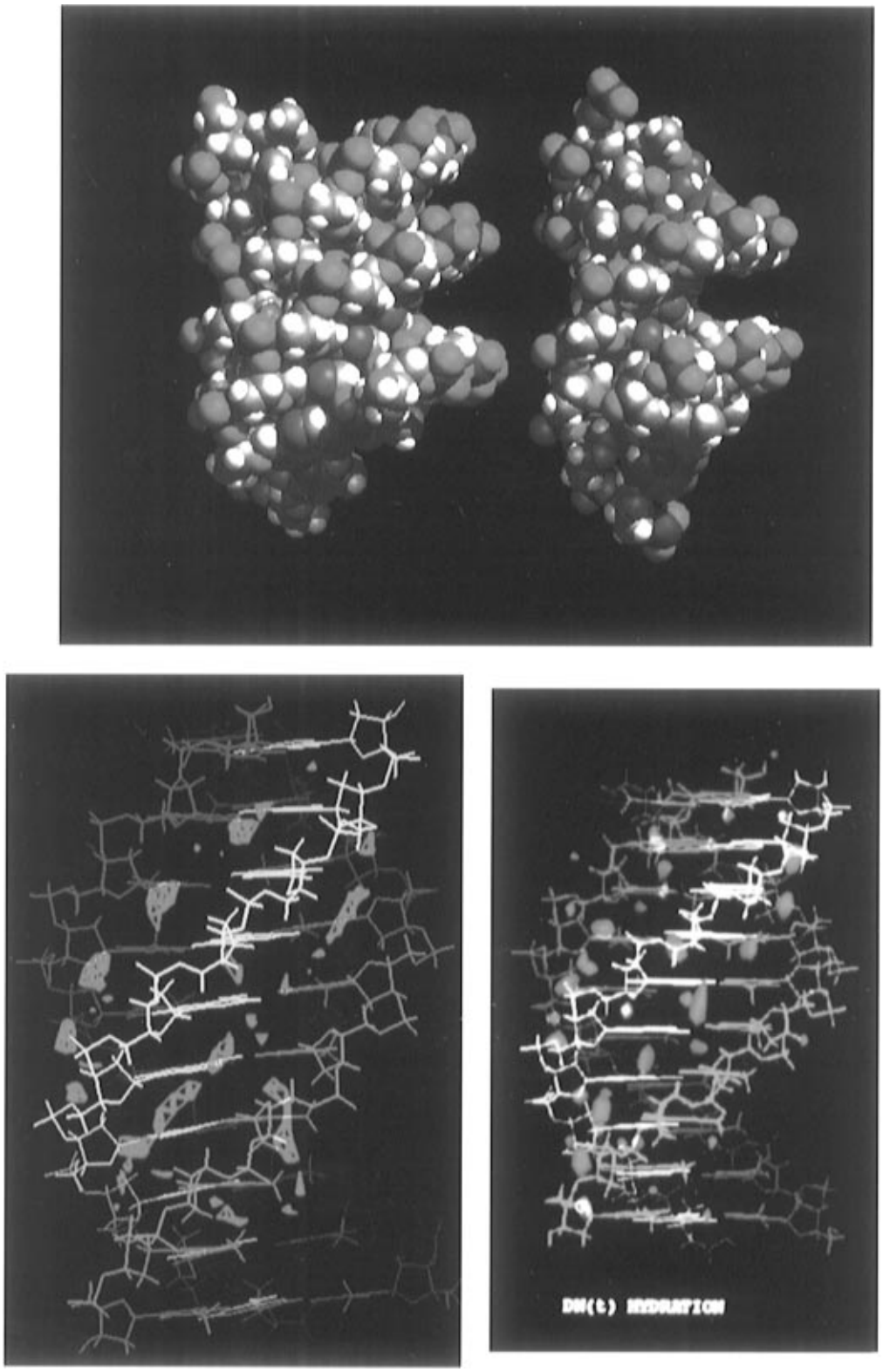


Figure 3. (a, top) Minor groove of a B-type DNA duplex and of the Bav triplex structure; (b, lower left) molecular interaction potential (MIP) for the Bav structure (see text); the −5.5 kcal/mol contour surface is plotted; and (c, lower right) water density around the DNA computed during the B-type trajectory; the contour surface corresponds to around 4 times the density of pure water.

the m- and mM-grooves are not really much better solvated than the MM-groove. In fact, the integration of water density in the grooves demonstrate that the relative amount of water is similar in all the grooves, and actually not too different to that expected for the same volumes of pure water. This means that all the grooves are equally well solvated, and that the differences lie in the mobility of the water molecules. In the m- and mM-grooves waters are stable only in small well-defined regions of the groove, resulting in lower mobility and longer residence times. The waters in the MM-groove are not located in fixed positions, but show high mobility within the groove.

MIP and hydration results noted above confirm the possibility that the two minor grooves of the $Y^+ \cdot R^- \cdot Y^-$ type DNA triplex may be targets for specific groove binders able to recognize and stabilize DNA, and that the MM groove of the triplex can be a good target for specific proteins. These interactions could be triplex specific, leading to an overall stabilization of the triplex with respect to its "parent" duplex. These findings will

contribute to the development of a new series of triplex-specific ligands, which may contribute to either antisense or antigene therapies.

In this paper, we have shown that long, stable trajectories starting from very different yet reasonable conformations for the triple helix have converged spontaneously to the same average structure. This, and the proven ability of AMBER-95 to model the dynamics of nucleic acids gives confidence in the correctness of the suggested model. However, we must consider that a classical force field is used and that the MD have been performed for a limited period of time. This warns against a too quantitative reading of the results and suggests that a validation of the results is necessary. Such a validation must be performed by comparison with the original experimental data used to construct the various triplex DNA models. In the absence of a high-resolution X-ray structure, not all the helical parameters of triplex DNA have been directly measured experimentally. This explains why experimental data have led to the suggestion of different models, even when the same technique is used (compare Table 1 and refs 5, 27, and 28). The parameters which are most clearly defined are the base pair-base pair parameters, twist and rise, since they can be accurately measured from either fiber difraction patern, or from interresidue NOEs. Both X-ray diffraction and NMR techniques suggest an average axial rise of 3.3−3.4 Å. Averaging over 1 ns of the B-type trajectory leads to an average rise of 3.4 ± 0.2 Å, matching both experimental estimates. Average helical twist is found from fiber difraction data to be around 30°, and the NMR estimates range from 27° to 32° with an average value of 31°. Our MD averages suggest a twist of 29 ± 1°, in good agreement with all the available experimental information. All experimental techniques suggest a periodicity of 12 residues per turn, which is exactly that determined in our MD simulations. It is clear then that MD simulations give to structures which agree very well with the major helical properties of the DNA triplex which can be exactly determined from experimental techniques.

The experimental measures of base pair-axis parameters (displacement, tip, and inclination) are probably less precise, since in most cases there is no direct experimental information on these parameters, and the parameters are inferred by modeling from base pair−base pair values. However, Patel and coworkers have reported values for the axis displacement, which can be determined from interresidue and long-range NOEs. They found the *Y* displacement to be near zero (as found in our MD simulations), and the *X* displacement is between −1.8 and −2.1 Å, which compares reasonably well with our average MD results (around −3 Å). No tip or inclination values are reported in NMR and fiber difraction studies, but in all cases it seems that these parameters are small, as found in our simulations.

Finally, base−base parameters are very difficult to determine from available experimental data. Thus X-ray fiber diffraction studies assume the rigidity of hydrogen bonds and cannot detect subtle changes in base pair planarity due to the resolution of the technique. NMR could in theory give some information, but the small number of interbase NOEs, and the typical use of restrained MD (bases are restrained to a fixed H-bond distance) at high temperature precludes the estimation of reliable base−base parameters. Our results suggest that hydrogen bonds are very well preserved, and that while out-of the plane movements of base pairs exist, they are moderate in extent.

Sugar puckering in triplex DNA has been a source of debate. X-ray fiber difraction was used to suggest a north (C3′-endo) for the riboses, consistent with an A-type conformation. Early NMR studies[28] were in agreement. However, more recent NMR[5,28,29] and IR[7] data have demonstrated that the deoxyriboses of triplex DNA exhibit south puckerings. Thus Feigon and coworkers, and Patel's group,[5,28,29] found that the average phase angle for the riboses ranged from 120° to 180°, covering the O4′-endo, C1′-exo and C2′-endo conformers, the average phase angle being around 144°, i.e., near the C1′-exo conformation, in excellent agreement with our MD results.

One of the most surprising findings in this paper is the size and the shape of the grooves. Particularly, we have found that the MM groove is similar in size to the major groove of a B-DNA, which suggests that it can be a target for peptides and proteins. This hypothesis cannot be directly verified by comparison with experimental data, but it agrees very well with recent experimental results (not available until the revised version of this manuscript was prepared) which demonstrate the existence of specific triplex-DNA binding protein.[30]

In summary, results presented here have shown that MD simulations, performed without the introduction of experimental restraints, are able to provide a reasonable picture of triplex DNA, which agrees to best of our knowledge with all the available experimental data. In particular, our results show that MD simulations on the time scale of a nanosecond can be used to drive the system from incorrect conformations to the "correct" conformation, opening up the possibility of using this technique to obtain estimates of nucleic acid structures in the absence of experimental information. Finally, a detailed structure for the poly(T·A·T) triple helix has been suggested and has been used to obtain new information on the structural and chemical characteristics of $Y^+ \cdot R^- \cdot Y^-$ type triplex DNA, which may play a major role in the design of new strategies for antisense and antigene therapy.

Acknowledgment. We are indebted to Dr. F. J. Luque and Dr. M. Pons for many helpful discussions. We thank Dr. Steven Jordan for making the X-ray coordinates of the PNA·DNA·-PNA triple helix available. This work has been suported by the Centre de Supercomputació de Catalunya, the Spanish DGICYT (PB93-0779), and the British Council (British−Spanish Joint Research Program: HB-118 B). G. Shields thanks the Spanish Ministry of Science for supporting him (SAB95-0053) during his sabbatical year in Barcelona.

JA970601Z

(29) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry* **1991**, *30*, 9022.

(30) Guiysse, A. L.; Praseuth, D.; Helene, C. *J. Mol. Biol.* **1997**, *267*, 289.