Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes

Thomas E. Cheatham, III,[†] and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94143-0446

Received October 18, 1996[⊗]

Abstract: Nanosecond length simulations applying the particle mesh Ewald method within AMBER 4.1 on canonical A-form and B-form geometries of d[CCAACGTTGG]$_2$, r[CCAACGUUGG]$_2$, and d[CCAACGTTGG]−r[CCAACGUUGG] duplexes in aqueous solution are reported. DNA duplexes only adopt a stable B-DNA geometry, in contrast to RNA duplexes which adopt both a stable A-RNA and "B-RNA" geometry. The observation of a stable "B-RNA" structure is somewhat surprising and suggests significant kinetic barriers to structural conversion in RNA structures on a nanosecond time scale. The "B-RNA" can be converted to A-RNA by forcing a concerted flip in the sugar puckers from C2′-endo to C3′-endo. The A-RNA structure displays features similar to A-form crystal structures, specifically interstrand purine stacking at the central pyrimidine−purine step is observed. When started in a canonical A-form geometry, DNA:RNA hybrid duplexes converge to a structure that is characteristic of experimental solution structures; specifically, a minor groove width intermediate between A-form and B-form geometries, the RNA strand in an A-form geometry, a mixture of C2′-endo and C3′-endo sugar puckers in the DNA strand, expected distribution of backbone angles and reasonable agreement with the helicoidal parameters are observed. In all of the simulations reported, A-form geometries appear to be less flexible than B-form geometries. There are also significant differences in the patterns of hydration and counterion association between A-form and B-form duplexes. In A-RNA, sodium counterions tend to associate into "pockets" in the major groove whereas these counterions tend to associate into the minor groove in B-form structures.

Introduction

In order to better understand biological information transfer, molecular interactions of nucleic acids, and the polymorphic character of nucleic acid conformation, it is important to understand the structure, dynamics, and relative flexibility of DNA:DNA, RNA:RNA, and DNA:RNA duplexes. A better understanding of the differences in sequence specific structure and dynamics can provide insight into protein−nucleic acid interactions, such as why the HIV-1 virus-encoded reverse transcriptase RNase H domain degrades the RNA strand of DNA:RNA hybrids faster than RNA:RNA duplexes[1] and what structural change in DNA:RNA hybrids, compared to duplex DNA, leads to the affinity change of the RNA polymerase core enzyme for the σ subunit.[2] Flexibility is clearly important in protein−nucleic acid recognition; rigidifying critical residues in the unbound protein can reduce the entropic cost of induced fit, as shown with the interaction of methionyl tRNA synthetase and $tRNA^{Met}$.[3] DNA:RNA complementary hybridization is important in a variety of biological processes including DNA replication,[4] normal and reverse transcription,[5] and recombination.[6] In addition, a better understanding of DNA:RNA hybrid structure is important for antisense drug development, since the potential drug−mRNA complex needs to be recognized by RNase H to allow the RNA to be degraded and the drug to have potent inhibitory activity.[7]

To date, most of our understanding of nucleic acid structure has come from X-ray crystallographic and NMR, CD, and Raman spectroscopic studies. Theoretical calculations have been of some use; however, earlier simulations employing molecular dynamics methods with an explicit representation of solvent and counterions [see reviews by Beveridge *et al.*[8,9]] were limited to a short time scale (∼100 ps) and during the simulation typically displayed anomalous structure (such as base pair fraying). More recent simulations of nucleic acids with explicit water on a longer time scale (∼1 ns) suggest the importance of properly treating the long-ranged electrostatic interactions.[10−14] In addition, there is a dependence of the results on the molecular mechanical force field applied. For example, Yang and Pettitt observed a B-DNA to A-DNA transition[15] when the CHARMM-

* To whom correspondence should be addressed.

[†] Present address: LSB, DCRT, 12A2041, National Institutes of Health, Bethesda, MD 20892-5626.

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.

(1) Gotte, M.; Fackler, S.; Hermann, T.; Perola, E.; Cellai, L.; Gross, H. J.; Le Grice, S. F.; Heumann, H. *EMBO J.* **1995**, *14*, 833−41.

(2) Hansen, U. M.; McClure, W. R. *J. Biol. Chem.* **1980**, *255*, 9556−63.

(3) Ribas de Pouplana, L.; Auld, D. S.; Kim, S.; Schimmerl, P. *Biochemistry* **1996**, *35*, 8095−8102.

(4) Ogawa, T.; Okazaki, T. *Annu. Rev. Biochem.* **1980**, *49*, 421−57.

(5) Varmus, H. *Science* **1988**, *240*, 1427−35.

(6) Daniels, G. A.; Lieber, M. R. *Nucleic Acids Res.* **1995**, *23*, 5006−11.

(7) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 543−584.

(8) Beveridge, D. L.; Swaminathan, S.; Ravishanker, G.; Withka, J. M.; Srinivasan, J.; Prevost, C.; Louise-May, S.; Langley, D. R.; DiCapua, F. M.; Bolton, P. H. *Molecular dynamics simulations on the hydration, structure and motions of DNA oligomers*; Beveridge, D. L., Swaminathan, S., Ravishanker, G., Withka, J. M., Srinivasan, J., Prevost, C., Louise-May, S., Langley, D. R., DiCapua, F. M., Bolton, P. H., Eds.; Macmillan Press: Riverside, NJ, 1993; pp 165−225.

(9) Beveridge, D. L.; Ravishanker, G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 246−255.

(10) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194.

(11) Louise-May, S.; Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1996**, *6*, 289−298.

(12) Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y. K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1995**, *117*, 2147−2158.

(13) York, D. M.; Yang, W.; Lee, H.; Darden, T. A.; Pedersen, L. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.

(14) Zichi, D. A. *J. Am. Chem. Soc.* **1995**, *117*, 2957−2969.

S0002-7863(96)03641-4 CCC: $14.00

23[16] all hydrogen parameter set[17] was applied with an Ewald treatment on the dodecamer d[CGCGAATTCGCG]$_2$, which suggests that the A form of this structure is more stable. In contrast, B-DNA is more stable than A-DNA when the force field described by Cornell et al.[18] is applied in molecular dynamics simulations with the particle mesh Ewald method[19] within AMBER 4.1[20] to a variety of DNA sequences,[21−23] including the above dodecamer.[24]

In this study, comparable simulations with RNA:RNA (r[CCAACGUUGG]$_2$) and DNA:RNA (d[CCAACGTTGG]−r[CCAACGTTGG]) duplexes were performed to determine if we might be able to properly represent the various differences in structure and dynamics among these models. In solution, DNA is expected to be within the larger B-type domain of right-handed duplex conformations. Crystallographic and NMR studies clearly demonstrate the heterogeneity in the B-DNA "family" of structures, most notably from sequence specific structure (bending, twisting), various accessible backbone conformations (B_I and B_{II}), and inherent flexibility resulting from sugar repuckering. This flexibility is manifest not only by noting how easily the DNA can be deformed by crystal packing forces[25−27] but also since slightly different structures appear when DNA is crystallized into different space groups[28,29] in contrast to RNA.[30] Moreover, B-form structures tend to diffract to lower resolution than A-form structures and B-form fibers have lower crystallinity and lesser order than A-form fibers.[31] The flexibility of B-DNA is further confirmed in NMR experiments which suggest a large range of possible conformations,[32,33] sugar repuckering,[34] and (α,γ) backbone "crankshaft" transitions.[35−37] A comparison of *J* coupling constants measured by NMR shows the enhanced flexibility of DNA duplexes compared to the more rigid RNA duplexes. Theoretical calculations also suggest an inherent flexibility in B-DNA, such as the "substates" of B-DNA conformations suggested by Lavery[38,39] to the frequent repuckering and (ϵ, ζ: t, g− to g−, t) backbone transitions observed during molecular dynamics simulations.[24] RNA duplexes, which are known to adopt a fairly small range of conformations within the A family (A, A′), are generally more rigid than corresponding DNA duplexes, as can be seen in the ^{31}P NMR experiments[32,33] and indirectly via crystallography.[30] On the other hand, an analysis of crystal structures suggests that double stranded DNA and RNA have a similar level of vibrational motion and sampling of conformational substates.[40] However, the simulations reported herein support the idea that A-form structures are more rigid than B-form structures.

Hybrid duplexes with one strand RNA and the complementary strand DNA tend to crystallize in the A-form,[41,42] but in solution are found in a conformation intermediate between an A- and B-form geometry. This is based on fiber diffraction data which suggest a different conformation than true A- or B-form geometry for hybrids at high relative humidity[43] with the DNA strands adopting C2′-endo and the RNA strands adopting C3′-endo sugar puckers. The CD data confirm this picture and further suggest that the overall helix is more A-like than B-like, with positive base pair inclination to the helical axis, small positive roll into the major groove, small positive buckle, negative propeller, and negative *x*-displacement from the helical axis.[44−47] Further support comes from the NMR data which clearly show that the RNA strand remains in an A-form geometry with C3′-endo puckers throughout, while the DNA strand is in a near B-form geometry with some controversy as to whether the pucker is O4′-endo or a mixture of C2′-endo and C3′-endo.[32,48−54] The latter is more consistent with the *J* coupling and dynamics data.[55] The NMR data also suggest differences in the expected distribution of backbone angles between the DNA and RNA strands in DNA:RNA hybrid duplexes. Beyond the angles directly correlated with the sugar pucker, specifically δ and χ which should be lower in the RNA strand, it is generally observed that α is typically lower, and ϵ, ζ, and γ are slightly higher in the RNA strand than in the DNA strand. All of the experimental data also suggest that the minor groove width in DNA:RNA hybrids is intermediate between

(15) Yang, L. Q.; Pettitt, B. M. *J. Phys. Chem.* **1996**, *100*, 2564−2566.
(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187−217.
(17) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(19) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(20) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comp. Phys. Commun.* **1995**, *91*, 1−41.
(21) Cieplak, P.; Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* In press.
(22) Duan, Y.; Wilkosz, P.; Crowley, M.; Rosenberg, J. M. *J. Mol. Biol*. Submitted for publication.
(23) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59−69.
(24) Cheatham, T. E., III; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434−44.
(25) Dickerson, R. E.; Goodsell, D. S.; Kopka, M. L.; Pjura, P. E. *J. Biomol. Struct. Dyn.* **1987**, *5*, 557−79.
(26) Dickerson, R. E.; Goodsell, D. S.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3579−83.
(27) Ramakrishnan, B.; Sundaralingam, M. *J. Biomol. Struct. Dyn.* **1993**, *11*, 11−26.
(28) Lipanov, A.; Kopka, M. L.; Kaczor-Grzeskowiak, M.; Quintana, J.; Dickerson, R. E. *Biochemistry* **1993**, *32*, 1373−89.
(29) Shakked, Z.; Guerstein-Guzikevich, G.; Eisenstein, M.; Frolow, F.; Rabinovich, D. *Nature* **1989**, *342*, 456−60.
(30) Portmann, S.; Usman, N.; Egli, M. *Biochemistry* **1995**, *34*, 7569−75.
(31) Thomas, G. J., Jr.; Benevides, J. M.; Overman, S. A.; Ueda, T.; Ushizawa, K.; Saitoh, M.; Tsuboi, M. *Biophys. J.* **1995**, *68*, 1073−88.
(32) Fujiwara, T.; Shindo, H. *Biochemistry* **1985**, *24*, 896−902.
(33) Shindo, H.; Fujiwara, T.; Akutsu, H.; Matsumoto, U.; Kyogoku, Y. *Biochemistry* **1985**, *24*, 887−95.
(34) Ulyanov, N. B.; Schmitz, U.; Kumar, A.; James, T. L. *Biophys. J.* **1995**, *68*, 13−24.
(35) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669−99.
(36) Schmitz, U.; Ulyanov, N. B.; Kumar, A.; James, T. L. *J. Mol. Biol.* **1993**, *234*, 373−89.
(37) Weisz, K.; Shafer, R. H.; Egan, W.; James, T. L. *Biochemistry* **1994**, *33*, 354−66.
(38) Poncin, M.; Hartmann, B.; Lavery, R. *J. Mol. Biol.* **1992**, *226*, 775−94.
(39) Lavery, R.; Hartmann, B. *Biophys. Chem.* **1994**, *50*, 33−45.
(40) Holbrook, S. R.; Kim, S.-H. *J. Mol. Biol.* **1984**, *173*, 361−388.
(41) Wang, A. H.; Fujii, S.; van Boom, J. H.; van der Marel, G. A.; van Boeckel, S. A.; Rich, A. *Nature* **1982**, *299*, 601−4.
(42) Egli, M.; Usman, N.; Zhang, S. G.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 534−8.
(43) Zimmerman, S. B.; Pheiffer, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 78−82.
(44) Gray, D. M.; Ratliff, R. L. *Biopolymers* **1975**, *14*, 487−98.
(45) Hall, K. B.; McLaughlin, L. W. *Biochemistry* **1991**, *30*, 10606−13.
(46) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463−6.
(47) Steely, H. T., Jr.; Gray, D. M.; Ratliff, R. L. *Nucleic Acids Res.* **1986**, *14*, 10071−90.
(48) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2435−43.
(49) Fedoroff, O.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1993**, *233*, 509−23.
(50) Gao, X.; Jeffs, P. W. *J. Biomol. NMR* **1994**, *4*, 367−84.
(51) Gonzalez, C.; Stec, W.; Kobylanska, A.; Hogrefe, R. I.; Reynolds, M.; James, T. L. *Biochemistry* **1994**, *33*, 11062−72.
(52) Katahira, M.; Lee, S. J.; Kobayashi, Y.; Sugeta, H.; Kyogoku, Y.; Iwai, S.; Ohtsuka, E.; Benevides, J. M.; Thomas, G. J. *J. Am. Chem. Soc.* **1990**, *112*, 4508−4512.
(53) Lane, A. N.; Ebel, S.; Brown, T. *Eur. J. Biochem.* **1993**, *215*, 297−306.
(54) Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R. *Biochemistry* **1993**, *32*, 4207−15.
(55) Gonzalez, C.; Stec, W.; Reynolds, M. A.; James, T. L. *Biochemistry* **1995**, *34*, 4969−82.

A- and B-form duplexes. JUNMA[56] minimizations[57] and *in vacuo* molecular dynamics simulations[58] support these observations.

In our simulations, we see the expected structural and dynamic trends. Specifically, we observe that A-RNA duplex structures are stable and within the canonical A family of structures, and moreover display sequence specific features that are consistent with the crystal data, specifically at the central CpG step which has a large rise and low helical twist value. DNA:RNA hybrids also show the expected structural trends with a DNA strand that repuckers between C2′-endo and C3′-endo sugar puckers, groove widths intermediate to A-form RNA:RNA and B-form DNA:DNA duplexes, expected distributions of backbone angles, and reasonable agreement with the helicoidal parameters.

Methods

The creation of the initial structures, equilibration, and dynamics were performed as described in our previous paper.[24] The starting canonical A- and B-form duplex structures[59] of d[CCAACGTTGG]$_2$, r[CCAACGUUGG]$_2$, and d[CCAACGTTGG]−r[CCAACGUUGG] were generated using the NUCGEN module of AMBER 4.1.[20] Hydrogens were added with the EDIT module with AMBER 4.1 and the initial hydrogen positions were minimized (*in vacuo*) while holding all non-hydrogen atoms fixed. Care was taken to ensure the hydrogens were added with the proper stereochemistry. Explicit net-neutralizing sodium counterions were placed at the phosphates of these models by the EDIT module of AMBER 4.1 and the nucleic acid and 18 counterions were surrounded by a periodic box of TIP3P waters which extended approximately 10 Å (in each direction) from the nucleic acid atoms. This leads to a periodic box size of ∼55 Å by ∼42 Å by ∼42 Å for the B-form structures and ∼59 Å by ∼40 Å by ∼40 Å for the A-form structures. The parameters described by Cornell et al.[18] [see also http://www.amber.ucsf.edu] were used in all of the simulations. All simulations were run using the sander module of AMBER 4.1 with SHAKE[60] (tolerance = 0.0005 Å) on the hydrogens, a 2-fs time step, a temperature of 300 K with Berendsen temperature coupling[61] and a time constant of 0.2 ps, a 9-Å cutoff applied to the Lennard-Jones interactions, and constant pressure with isotropic molecule based scaling[61] with a time constant of 0.2 ps. The nonbonded list was updated every 10 steps. Equilibration was performed by first holding the positions of the DNA fixed and running 1000 steps of minimization followed by dynamics for 25 ps with a cutoff of 9 Å on all interactions. In order to avoid shifting of the two DNA strand molecules during constant pressure equilibration (when the DNA was held fixed), both strands were treated as if they were a single molecule. After this initial equilibration, all subsequent simulations were run using the particle mesh Ewald method (PME)[19] within AMBER 4.1 using a cubic B-spline interpolation order and a 10^{-5} Å tolerance for the direct space sum cutoff. To speed up the fast Fourier transform in the calculation of the reciprocal sum, the size of the PME charge grid is chosen to be a product of powers of 2, 3, and 5 and to be slightly larger than the size of the periodic box. This leads to a grid spacing of ∼1 Å or less. Equilibration was continued with 25 kcal/(mol·Å^2) restraints placed on all solute atoms, minimization for 1000 steps, followed by 3 ps of MD which allowed the water to relax around the solute. This equilibration was followed by 5 rounds of 600-step minimization where the solute restraints were reduced by 5 kcal/mol during each round. Finally, the system was heated from 100 K to 300 K over 2 ps and then production runs were initiated. It should be noted that the main goal of the equilibration protocol outlined above is to first let the counterions and water equilibrate, then secondarily let the DNA slowly relax away from the starting geometry to avoid bad contacts, relieve poor bond, angle, and dihedral deviations in the model structure, yet help it remain "close" to the initial structure. The most important step in this regard is the initial water and counterion equilibration. To determine if 25 ps is enough time to relax the solvent, the pressure, volume, and density are typically monitored. Although not shown, these indicators are easily equilibrated within the 25 ps of water/counterion equilibration. For more discussion about the equilibration protocol used herein, see the presentation available on the AMBER world wide web page at "http://www.amber.ucsf.edu/amber/tutorial/polyA-polyT/".

Since the pairlist is not updated every step, the SHAKE tolerance used is rather modest (0.0005 Å), and constant pressure is utilized, some small energy drain during the simulations can occur. Since uniform scaling of velocities by Berendsen coupling was utilized to bring the very slowly dropping temperature back up to 300 K, the center of mass velocity can slowly grow. Therefore, periodically in the simulation (at every restart or every ∼80−100 ps) this center of mass velocity was removed during the production dynamics.

Simulations were run on RNA:RNA duplexes starting from canonical A (2030 ps, referred to as A-RNA) and canonical B (2370 ps, referred to as B-RNA) duplexes. After ∼1.5 ns with the canonical B start, it was realized that the trajectory had converged to a "B-RNA" conformation that is remarkably close to the average B-DNA structure observed in the corresponding DNA:DNA simulations of d[CCAACGTTGG]$_2$. Therefore, in an attempt to push the "B-RNA" structure away from the B-DNA average structure and perhaps initiate a B- to A-RNA transition, a simulation was run for 540 ps restarting the trajectory from 1565 ps with the temperature increased to 400 K. An additional simulation was also started from 1565 ps and run for 1070 ps where a concerted flip in the sugar puckers was forced by applying restraints on the C1′−C2′−C3′−C4′ torsion of each nucleotide for a limited time as described below. To determine how to best restrain the C1′−C2′−C3′−C4′ torsion to give a particular sugar pucker pseudorotation value, calculations on adenine nucleotides *in vacuo* and in solution were run with various force constants on the restraints. In order to maintain reasonable distributions of the sugar pucker about the mean pucker pseudorotation value, flatwell restraints were applied. To constrain the pucker to C3′-endo, flatwell restraints are applied with no penalty between 30° and 40°, parabolic penalties from 30° to 20° and 40° to 50°, and linear penalties outside this range. To constraint the pucker to C2′-endo the flatwell restraint is applied between −40° and −38°, parabolic penalties from −40° to −44° and −38° to −34°, and linear penalties outside this range. The typical force constant necessary to "restrain" the pucker to the appropriate range is on the order of 30 kcal/(mol·rad^2). However, to force a concerted flip larger restraints were necessary; the goal was to allow a quick concerted flip in the pucker such that the biasing restraints could then be turned off. In this simulation, a concerted flip from the "B-RNA" C2′-endo puckers to C3′-endo was forced by gradually increasing the restraint penalty force constant from 0 to 300.0 kcal/(mol·rad^2) over 5 ps, then gradually reducing this restraint from 300.0 to 30.0 kcal/(mol·rad^2) over the next 45 ps. After this time, all the restraints were turned off, and free dynamics were continued. Note that a simulation where the force constant on the restraints was held at 30.0 kcal/(mol·rad^2) for 25 ps initially, then removed, was not sufficient to completely force a concerted flip in the puckers.

Simulations were also run both on canonical A (2005 ps, referred to as A-hybrid) and canonical B (2045 ps, referred to as B-hybrid) forms of DNA:RNA hybrid d[CCAACGTTGG]−r[CCAACGUUGG] duplexes. The canonical B-form DNA:RNA hybrid simulation was continued from 2045 ps for ∼400 ps at 400 K. Simulations were also run, and some of the results previously reported[24] on canonical A and canonical B (1400 ps, referred to as A-DNA and B-DNA, respectively) models of d[CCAACGTTGG]$_2$ duplexes.

All of the results were analyzed using the carnal, anal, nmode, and mdanal modules of AMBER 4.1, the Dials and Windows[35] interface to Curves,[62] a more recent version of Curves, version 5.1 dated June 1996, or some adapted trajectory analysis software (rdparm). Standard

(56) Lavery, R.; Zakrzewska, K.; Sklenar, H. *Comp. Phys. Commun.* **1995**, *91*, 135−158.
(57) Sanghani, S. R.; Lavery, R. *Nucleic Acids Res.* **1994**, *22*, 1444−1449.
(58) Fritsch, V.; Wolf, R. M. *J. Biomol. Struct. Dyn.* **1994**, *11*, 1161−1174.
(59) Arnott, S.; Hukins, D. W. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504−9.
(60) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327−341.
(61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comp. Phys.* **1984**, *81*, 3684−3690.
(62) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.

angle (α, β, γ, δ, ϵ, ζ, χ)[63] and helicoidal parameter[64] names and definitions are presented in the analysis. Sugar pucker pseudorotation values and sugar pucker amplitudes were calculated based on the Altona and Sundaralingam conventions;[65] in the text "sugar pucker" or "pucker" will be used synonymously with "sugar pucker pseudorotation phase". Nucleic acid residue names are referred to in the text as one letter codes. Where necessary, a subscript for the residue number is also presented; the residue number is in the 5′ to 3′ direction with the first strand numbered 1−10 and the second strand 11−20. To avoid confusion between base pairs and base pair steps in the text, base pair steps are denoted with a "p", *i.e.* TpG steps in contrast to TG base pairs.

Average structures from the trajectories were calculated using the carnal module of AMBER to coordinate average the root-mean-square (RMS) coordinate fit frames (over all DNA atoms) taken at 1-ps intervals. No extra processing of these average coordinates (*i.e.* minimization) was performed. Since these structures were not minimized post averaging, they may contain some anomalous structural features, such as is exemplified with DNA thymine methyls which will on average display each hydrogen co-linear with the C−C bond. This may lead to higher calculated RMSd values and slight differences in the calculated helicoidal parameters. However, minimization to fix up the structure is tricky since it is impractical to include water in the calculation. Without water to balance the interactions, *in vacuo* minimization will distort the structure significantly from what is observed during the dynamics with explicit water. In order to investigate the effect of minimization on the average structures, two short minimizations of the B-form RNA:RNA duplex average structure over 1370−2370 ps, one with a constant dielectric constant and the other using a distance dependent dielectric function with a dielectric constant of 4, were run where the energy was minimized until the RSMd in energy between steps changed by less than 1.0 kcal/mol. These short minimizations led to structures that were only 0.37 and 0.22 Å away from the average structure, respectively. Despite the rather small root-mean-square deviation between the structures, there are small, but significant, differences in the backbone angles (less then 5°) between the average and minimized structures and a higher average sugar pucker amplitude (43.0°) in the minimized structure. This is not surprising since during the minimization process, the dihedral angle values will tend to move toward minimizing their deviation from the equilibrium force field values. Although the helicoidal parameters are very sensitive to the base atom positions, the differences in the helicoidal parameters between the minimized and non-minimized structures are significantly smaller than the standard deviations of the values over the trajectory. Given these small differences between the minimized and average structures, and considering the fairly low all atom self symmetric RMSd values (ssRMSd, A-RNA is ∼0.34 Å and B-RNA is ∼1.0 Å) of the average structures, the non-minimized average structures calculated will be used in the analysis presented herein. The self symmetric RMSd (ssRMSd) values are defined herein for duplexes in which both strands have the same sequence; the ssRMSd is the RMSd of the structure to the symmetric structure obtained by rotating the duplex to match the second strand to the first.

Diffusion constants were calculated using the Einstein relationship (or the slope of the mean square displacements in angstroms versus time in picoseconds multiplied by 10.0/6.0 which leads to units of 10^{-5} cm^2/s) using the software rdparm. The average self diffusion constants of TIP3P water and Na^+ counterions in these simulations (A-RNA, B-RNA, A-hybrid, B-hybrid, and B-DNA) over the course of a nanosecond are $4.7-4.9 \times 10^{-5}$ and $1.2-2.1 \times 10^{-5}$ cm^2/s, respectively. This is slightly lower than the calculated values for pure TIP3P as is expected due to some condensation of the ions and water with the nucleic acids. These data are presented here to show that diffusion is not seriously inhibited by the imposition of true periodicity [see Essman *et al.*[19] for a more thorough analysis of water diffusion with and without Ewald]. Atomic positional fluctuations were determined over nanosecond portions of the trajectory using the mdanal module from AMBER 4.1. Normal mode calculations were performed using the nmode module from AMBER 4.1.

Solvent and counterion distributions were calculated by binning atom positions from RMS coordinate fit frames over all DNA atoms at 1-ps intervals into 0.5 $Å^3$ grids over 1-ns portions from the trajectories. In other words, the value of each grid element represents the number of times the coordinates of the center of a particular atom of interest (*i.e.* water oxygen) were within the 0.5 $Å^3$ represented by that particular grid element. These grids can then be contoured using the density delegate of UCSF MidasPlus.[66] For 1000 frames, the expected number of waters per grid element, assuming bulk water density, is 4.18. In the graphics of the water and counterion hydration presented, the contouring of the water/counterion density was typically performed at 12.0 or 15.0 hits per 0.5 $Å^3$, or approximately three times expected bulk water density. In the text, the contouring level will be referred to as "*x* hits per 0.5 $Å^3$", which represents *x* visits to each 0.5-$Å^3$ grid from 1000 frames of the trajectory taken at 1-ps intervals.

All the molecular graphics images herein were produced using the MidasPlus software available from the Computer Graphics Laboratory, University of California, San Francisco. All the molecular dynamics calculations were either run on an SGI R8000 at UCSF or 16 processors of the Cray T3D at the Pittsburgh Supercomputing Center using a modified version of the sander module of AMBER 4.1. The Cray T3D parallel version was adapted from the MPI version of sander originally developed by Vincent and Merz[67] and incorporated into AMBER 4.1. Parallelization of the particle mesh Ewald code specifically for the Cray T3D and also more generally under MPI was performed by Michael Crowley of the Pittsburgh Supercomputing Center. Approximately 1 week on 16 processors of the Cray T3D or 17.5 days on 1 processor of the 75-MHz SGI R8000 is required to simulate each system for 1 ns.

Results and Discussion

RNA Maintains a Stable A-RNA Structure with Features Similar to A-Form Nucleic Acid Crystal Structures. In the simulations we find that when the RNA duplex is started in a canonical A geometry it remains in a canonical A geometry. The A-form RNA:RNA duplex (A-RNA) adopts a "stable" average structure over the last nanosecond of a ∼2-ns simulation that is within ∼2 Å of canonical A-RNA. The convergence to this average structure is rather good, as judged by comparing the "self" symmetric root mean square deviation (ssRMSd) which is only 0.34 Å. The data in Table 1 show that the A-RNA remains very close to canonical A. Moreover, the data in Table 2 show that the structure is characteristic of the canonical A family of RNA structures.[63] This average A-RNA structure has an average helical inclination of 15.0°, *x*-displacement from the helical axis of −5.2 Å, C3′-endo sugar puckers, and the rise between the base pairs of 2.7 Å. Interestingly, the A-RNA average structure does not have the sequence specific bending patterns, such as the notable roll into the major groove at the TpG and CpG steps, seen in comparable B-DNA simulations.[24] Instead, as shown by the dark black lines in Figure 1, we see a generalized base pair roll into the major groove and relatively more uniform base pair propeller and buckle. Although the average structure appears to be close to canonical A, there are some notable deviations. In particular, a recent analysis of individual dimer steps from DNA crystal structures clearly demonstrates that A-form and B-form structures show little overlap in the respective helicoidal parameters, particularly in slide and roll.[68] In this analysis, Gorin *et al.* show that A-form structures typically display a more negative average slide and higher average roll. Our slide versus roll values for the A-RNA simulation tend to fall into the "B-DNA" region, or on the boundary of the B-DNA region, depending on exactly how the values are calculated (discussed in more detail below). Overall, our slide values are of a slightly lower magnitude than has been

(63) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(64) Dickerson, R. E. *Nucleic Acids Res.* **1989**, *17*, 1797−803.
(65) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.
(66) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13−27.
(67) Vincent, J. J.; Merz, K. M. *J. Comp. Chem.* **1995**, *16*, 1420−1427.
(68) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1995**, *247*, 34−48.

Table 1. Root Mean Square Deviations (RMSd) of All Atoms (not mass weighted) in Various Structures (in angstroms)[a]

	A	B	B-DNA	A-RNA	B-RNA	A-hybrd	B-hybrd
A		3.81	3.23	1.74	3.14	2.15	3.41
		3.79	3.10	1.88	3.50	1.72	3.35
		5.80	3.86	2.09	3.82		
B	2.68		1.74	3.39	2.03	2.65	1.71
	2.70		1.94	3.44	2.18	3.32	1.93
	3.29		3.02	5.41	3.47		
B-DNA	2.28	1.13		2.36	0.81	1.46	0.80
	2.25	1.29		2.35	0.85	2.34	0.90
	2.61	1.69		2.97	0.95		
A-RNA	1.19	2.49	2.00		2.23		
	1.14	2.58	1.96		2.45	1.03	2.44
	1.64	3.18	2.17		2.72		
B-RNA	2.37	1.23	0.61	2.00			
	2.46	1.51	0.70	2.01		2.64	1.12
	2.80	2.04	0.76	2.13			
A-hybrd	1.55	1.85	1.17				1.83
	0.96	2.37	1.84	0.79	2.03		2.39
							2.56
B-hybrd	2.20	1.19	0.36			1.09	
	2.43	1.17	0.75	2.08	0.86	1.99	
						1.75	

[a] The upper triangle is over all residues, the lower triangle is the internal 6 residues from each strand. For each comparison, three numbers are presented, where applicable. The top number is the RMSd of the first strand (residues 1−10 in the upper triangle or residues 3−8 in the lower triangle), the middle number is the second strand (residues 11−20 in the upper triangle and residues 13−18 in the lower triangle), and the bottom number is the RMSd of both strands. In the case of the DNA:RNA hybrids, the DNA strand is always the first strand. The canonical A and canonical B models were built into both DNA:DNA and RNA:RNA duplexes as discussed in the methods. The average structures are generated by a straight average of the root-mean-square fit coordinate streams taken at 1-ps intervals as discussed in the methods; no minimization of these structures was performed. The average DNA model, B-DNA, is described in our previous paper[24] and represents the B2 trajectory; the average is over 0−1400 ps. A-RNA and B-RNA represent the canonical A and canonical B starts of the RNA:RNA duplex over 1030−2030 and 1370−2370 ps, respectively. The A-hybrd and B-hybrd average structures are from the canonical A and canonical B DNA:RNA hybrid duplex simulations over 1005−2005 and 1045−2045 ps, respectively.

observed in various A-DNA[27,69] or A-RNA[30] crystal structures. The significance of these small deviations is difficult to judge, since we are comparing a model of a "solution" structure of A-RNA to data derived from the analysis of A-DNA crystal structures.

Beyond the general observations, the most notable sequence specific deviations from canonical A-RNA can be seen by examining the dark black lines in Figure 1. In particular, the central CpG step shows an anomalous rise of 3.77 Å and a low helical twist of 25.0°, with correspondingly lower inclination values of ∼8° and relatively less propeller twisting of ∼−8°, at the central CG and GC base pairs, and a negative cup (or difference between the CG and GC base pair buckle) equal to −11.7°. Neglecting for a moment the lower than expected inclination and the absence of a more positive roll at this step (instead a lower roll of −0.5° is observed), the "anomalies" noted above are characteristic of a "low twist profile" base pair step which is expected for CpG steps.[70] Moreover, an analysis of canonical A-forms does clearly show the negative correlation between rise and inclination.[71] This is expected since in A-form structures the inclination of the base pairs leads to a lower axial rise.[70] The only unexpected observation in these values is the low roll (into the minor groove) at this step, since roll is inversely correlated with twist.

Figure 2a displays a stereoview picture of the A-RNA average structure and the calculated global helical axis. Traversing the helical axis down the sequence from top to bottom, it appears to curve one direction, until the CpG step is reached (in the middle), where the direction changes, then it curves off in another direction at the end. Based on the Curves analysis, the largest deviation from true helicity occurs at this central CpG step. Looking at the base pair stacking, it appears that the cytosines of the CG base pairs are relatively unstacked and the two guanines from opposite strands have positively shifted and negatively slid so as to partially stack on top of each other. This result is very similar to what has been seen in the A-DNA crystal structures of d[CCCCGGGG]$_2$,[69] d[GGGCGCCC]$_2$,[29] and other octamers[27,72] where interstrand stacking of the central guanines from opposite strands is observed and low twist and more slide (both of which improve the interstrand stacking), low propeller twist (which reduces interstrand steric clashes[73]), and low roll angle (which compensates for the slide) are observed. The helicoidal structure at the CpG step is locally perturbed in order to more favorably stack at the intrinsically weak 5′ to 3′ pyrimidine on purine step. It is also very interesting that we observed a spontaneous crankshaft transition in the α and γ backbone angles from the more common *gauche*−, *gauche*+ state to *trans*, *trans* values at ∼872 ps between the CpG step of the first strand which serves to increase the base pair separation[74] and may serve to improve the cross-strand overlap of the guanine bases.[69] The all-trans conformation of the P−O5′−C5′−C4′ bonds, low twist, negative slide, and interstrand guanine stacking at the central CpG steps have been observed in a variety of A-DNA tetragonal crystal structures[69,72] and in one strand of the rhombohedral crystal structure of r(CCCCGGGG)$_2$.[30]

It is significant that the interstrand guanine stacking and a single crankshaft (α,γ) transition are observed spontaneously at the central CpG step in solution phase simulations of a decamer. All of the tetragonal octamer A-DNA crystal structures analyzed by Ramakrishnan and Sundaralingam[27] have a pyrimidine−purine, and most commonly a CpG, base pair step in the center of the duplex; sequences with other types of step at the center do not crystallize into the tetragonal form. Moreover, each of these tetragonal structures has a low twist at the central pyrimidine−purine step. In addition to the above-mentioned features, there is also postulated a particular hydration pattern in the tetragonal crystal structures; specifically a chain of water molecules across the CpG step that links the minor groove to the backbone of other duplexes in the unit cell is observed which may contribute to the stability of the tetragonal geometry.[72] It is not clear if other sequences avoid crystallizing in the tetragonal group because they do not have the deformable weak stacking interaction found in pyrimidine−purine steps and the water stabilizing crystal contacts which allow for facile crystallization or alternatively if the interstrand stacking of the CpG step is an inherent contextual sequence specific structure. Support for the hypothesis that it is the deformability of this step, induced by crystal packing, that leads to the low twist and interstrand purine stacking comes from analysis of A-DNA crystal structures, which demonstrate that the low twist occurs only at pyrimidine−purine steps at the center of the helix and not at other pyrimidine−purine steps.[27] However, the spontaneous observation of low twist and interstrand guanine stacking at the CpG steps without the characteristic crystal packing and hydration patterns (discussed later) in our simulations demonstrate that the crystal packing forces alone do not induce the

(69) Haran, T. E.; Shakked, Z.; Wang, A. H.; Rich, A. *J. Biomol. Struct. Dyn.* **1987**, *5*, 199−217.
(70) Yanagi, K.; Prive, G. G.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 201−14.
(71) Babcock, M. S.; Olson, W. K. *J. Mol. Biol.* **1994**, *237*, 98−124.
(72) Eisenstein, M.; Shakked, Z. *J. Mol. Biol.* **1995**, *248*, 662−678.
(73) Calladine, C. R. *J. Mol. Biol.* **1982**, *161*, 343−52.
(74) Olson, W. K. *Nucleic Acids Res.* **1982**, *10*, 777−87.
(75) Jain, S.; Sundaralingam, M. *J. Biol. Chem.* **1989**, *264*, 12780−4.

Table 2. Standard Angle and Helicoidal Values and Standard Deviations (in parentheses) Averaged Over All the Residues, Base Pairs, or Base Pair Steps (where appropriate) for the Various Duplex Structures Specified[a]

	A−RNA 1030−2030 ps	B−RNA 1370−2370 ps	A−hybrid 1005−2005 ps	B−hybrid 1045−2045 ps	B−DNA 400−1400 ps
α (deg)	277.0 (10.5)	286.3 (12.5)	279.9 (10.6)	288.1 (15.0)	290.4 (11.6)
β (deg)	175.6 (9.5)	173.3 (12.8)	173.8 (9.9)	171.6 (12.6)	168.4 (12.6)
γ (deg)	69.8 (8.9)	52.7 (11.9)	66.0 (9.4)	55.4 (13.9)	54.3 (10.6)
δ (deg)	79.3 (8.2)	127.7 (15.1)	95.2 (13.2)	120.2 (16.9)	116.6 (18.0)
ϵ (deg)	201.6 (10.1)	200.3 (18.4)	194.5 (9.7)	194.7 (19.9)	197.0 (18.5)
ζ (deg)	291.4 (8.6)	247.1 (26.0)	281.8 (11.3)	257.9 (26.2)	258.0 (25.5)
χ (deg)	201.7 (9.1)	248.7 (17.5)	214.9 (12.5)	241.4 (18.6)	234.2 (16.7)
pucker (deg)	22.6 (16.5)	137.0 (24.4)	66.8 (22.2)	127.6 (27.4)	122.8 (28.1)
amplitude (deg)	38.9 (5.0)	40.2 (5.2)	38.0 (5.4)	37.8 (5.8)	37.1 (6.1)
propeller (deg)	−12.6 (12.0)	−10.4 (13.6)	−13.2 (11.4)	−13.0 (13.5)	−10.4 (12.3)
buckle (deg)	−0.8 (10.6)	1.2 (13.5)	4.3 (10.7)	−1.7 (13.7)	0.4 (11.4)
opening (deg)	3.3 (6.0)	4.5 (7.5)	1.9 (5.9)	4.0 (8.2)	2.0 (5.6)
rise (Å)	2.7 (0.6)	3.2 (0.6)	3.0 (0.5)	3.2 (0.6)	3.3 (0.5)
tilt (deg)	−0.5 (5.4)	−0.7 (6.7)	0.0 (5.4)	−1.1 (6.8)	0.0 (5.4)
roll (deg)	2.4 (7.7)	2.2 (9.5)	1.6 (7.9)	2.6 (8.9)	1.3 (8.7)
twist (deg)	30.9 (4.3)	30.6 (5.4)	30.5 (4.2)	30.6 (6.5)	30.9 (5.1)
x-disp (Å)	−5.2 (0.8)	−3.1 (0.9)	−4.5 (0.8)	−2.7 (1.2)	−3.0 (0.7)
y-disp (Å)	0.1 (0.7)	0.5 (0.7)	−0.3 (0.6)	−0.1 (1.3)	0.0 (0.5)
inc (deg)	15.0 (9.3)	9.8 (10.8)	11.1 (9.2)	6.1 (13.5)	4.9 (7.3)
tip (deg)	−1.9 (6.7)	−2.4 (7.9)	4.6 (6.1)	−1.8 (13.5)	0.4 (5.8)
shear (Å)	0.0 (0.5)	0.1 (0.5)	0.0 (0.4)	−0.1 (0.5)	0.0 (0.4)
stretch (Å)	0.4 (0.4)	0.3 (0.3)	0.2 (0.3)	0.2 (0.3)	0.1 (0.2)
stagger (Å)	−0.2 (0.5)	−0.2 (0.6)	−0.2 (0.5)	−0.2 (0.5)	−0.2 (0.5)
shift (Å)	0.0 (0.6)	0.0 (0.7)	0.0 (0.6)	0.1 (0.7)	0.0 (0.6)
slide (Å)	−0.1 (0.4)	−0.1 (0.5)	−0.1 (0.4)	−0.1 (0.5)	−0.1 (0.4)

[a] The average values were calculated by determining the values for each average structure. The average structure is calculated by best fitting the RMSd over all atoms (mass weighted) at 1-ps intervals and performing a straight coordinate average. Each average structure represents a nanosecond portion of the trajectory (as specified). The values in parentheses represent the standard deviation of the values calculated for each 1-ps frame during the nanosecond portion of the trajectory specified. The "A" or "B" above denotes the starting conformation, which is canonical A or canonical B, respectively. All the values were calculated using the dials and windows[35] interface to Curves[62] or a more recent version of Curves, version 5.1, dated June 1996.

central CpG step behavior. Instead, the interstrand stacking is a real contextual sequence dependent structural effect.

The central pyrimidine−3′-5′-purine step is clearly more deformable, as results of crystallizing the r(CCCCGGGG)$_2$[30] and d(GTGTACAC)[75] sequences in two distinct lattices attest. The CpG steps in our simulations show somewhat enhanced flexibility which suggests a greater deformability. In the first strand (where the crankshaft transition to *trans*, *trans* is observed in the α and *γ* angles), the standard deviations in α (15.2°), *β* (10.5°), and *γ* (13.1°) are significantly larger than the average fluctuations at these angles (Table 2) and the fluctuations in the helical twist and rise at this step are more than 10% above the average. Given the above, the question still remains as to why this behavior is only observed at the central, and not flanking, pyrimidine−purine steps.

Analysis of Average Structures vs Averages of the Analysis of Snapshots from the Trajectory. Interpreting and judging the validity of correlations found in the helicoidal parameters is often difficult. While some correlations in the helicoidal parameters are expected and structurally significant, some may result from the methods used to calculate the helicoidal parameters,[71] some may be an artifact of insufficient sampling in the simulation leading to a misrepresentative average structure, and some may be an artifact of the force field representation. One means of checking the consistency of the average structure is by comparing the average of all the values calculated from analysis of each individual snapshot in the trajectory to those values calculated from the one average structure. To avoid confusion in the description, the former will be referred to as the "mean of the snapshots" and the latter as the "average". When comparing the mean of the snapshots to the average, all the backbone angles and sugar puckers are all respectively within ∼1° of each other. The agreement seen here is even better than the agreement between the average structure and the average structure after minimization, as discussed in the methods section. However, the sugar pucker amplitude appears to be slightly lower in the average than is expected based on averaging the amplitudes from each snapshot; this observation is fairly easy to rationalize based on the flattening of the ring during coordinate averaging of the various sugar puckers.

A more significant difference between the mean of the snapshots and the average is found for the helicoidal parameters. Specifically, the average *x*-displacement is 1.52 Å higher, the average inclination ∼10.6° lower, the average roll ∼5.6° larger, and the average rise between base pairs 0.3 Å higher when calculated using the mean of the snapshots from the A-RNA simulation over 1030−2030 ps. These differences are actually quite large and in the case of the *x*-displacement and inclination greater than the standard deviations. The differences in rise and inclination alone correspond to the difference between A-RNA and A′-RNA.[63] These differences are not restricted to the A-RNA simulation, but are seen in all of the simulations run. It is not clear what the significance of this is; however, it is worth pointing out since analysis of the snapshots (*i.e.* as with dials and windows) is actually quite common. These differences may suggest that we have not sampled long enough for the time average to converge. Alternatively, there could be a systematic difference in the two types of analysis (such as with the averaging of the sugar pucker amplitudes). Either way, the differences we observe demonstrate the sensitivity in the calculation of the helicoidal parameters to the structure. It should also be noted that although the values are shifted, depending on the way the data is analyzed, the trends or relative values for various base pairs or base pair steps are still maintained. In other words, the anomalous rise and lower inclination at the CpG step in the A-RNA simulation is still apparent whether the mean of the snapshots or the average structure is analyzed, it is just the actual value that is shifted. This implies that either the average or the mean of the snapshots

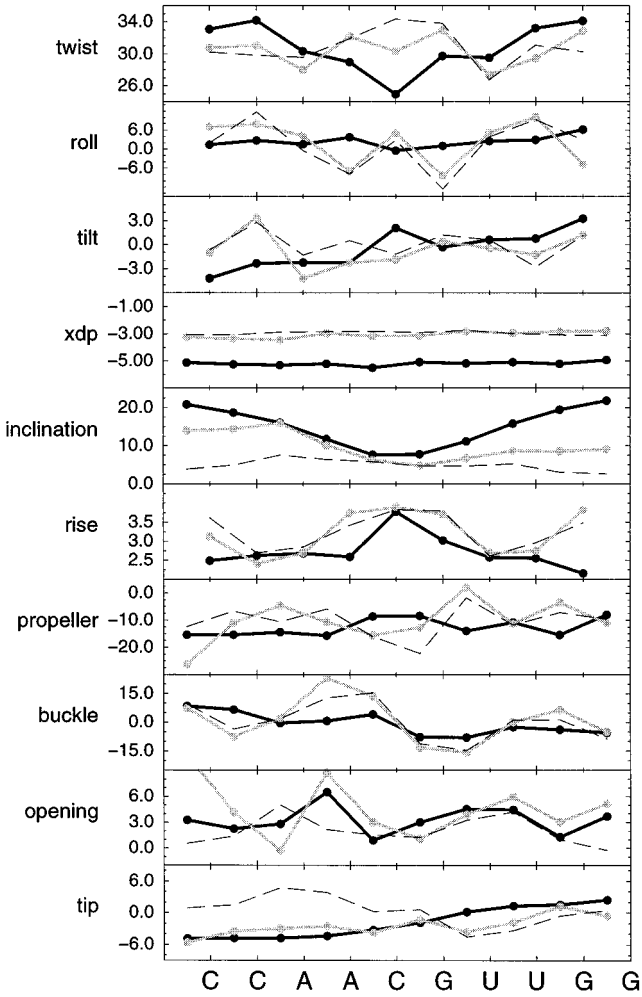

Figure 1. Helicoidal parameters calculated with the dials and windows[35] interface to Curves[62] for average structures from the trajectory. The canonical A-start RNA:RNA duplex average structure from 1030 to 2030 ps (A-RNA) is represented in black, the canonical B-start RNA:RNA duplex average structure from 1370 to 2370 ps (B-RNA) in gray, and the canonical B start DNA:DNA duplex average structure from 400 to 1400 ps (B-DNA) is shown as a dotted black line. The average structures were created by averaging all the coordinates of the nucleic acid from RMS fit frames taken at 1-ps intervals. The twist, roll, tilt, inclination, propeller twist, buckle, opening, and tip are all represented in degrees and the *x*-displacement from the helical axis and rise are represented in angstroms. The values are presented traversing the helix from left to right representing the 5′ to 3′ direction. Note that although the *x*-axis legend specifies a "U" for uracil, thymine nucleotides were used in the DNA strands.

is appropriate for analyzing sequence dependencies as long as the analysis is internally consistent. In Table 2 we present the analysis of the average structure since this nanosecond time averaged structure is perhaps more representative of what is observed experimentally by NMR or crystallography. The standard deviations, on the other hand, can only be estimated by analyzing a series of individual frames.

B-RNA Is Also Stable on the 1−2 Nanosecond Time Scale. When the simulation is started in a canonical B form, we do not observe a spontaneous B to A transition in the RNA:RNA duplex simulations. Instead, snapshots from the trajectory remain in the B family for over 2 ns and move toward an average structure that is exceptionally close to the "average DNA" structure calculated for DNA:DNA duplexes of the same sequence.[24] As shown in Table 1, the RMSd of the canonical B start RNA:RNA duplex over 1370−2370 ps to the canonical B start DNA:DNA duplex over 400−1400 ps is ∼1 Å and the "B-RNA" self agreement (ssRMSd) is 1.0 Å. Sugar repuckering from C2′- endo to C3′-endo (and back) does occur in "B-RNA", but at a much lower rate than is observed in the DNA simulations. Helicoidal analysis of the average B-RNA structure displayed in Figure 1 in gray compared to the DNA:DNA average B duplex structure (dotted black line) shows remarkable agreement between the two structures. Except for the base pair inclination at the terminal base pairs, all of the helicoidal parameters for the B-RNA and B-DNA structures (Figure 1, Table 2) are in the expected range. Although the *x*-displacement is slightly lower (∼−3 Å) than is expected for B-form structures (0 to −2 Å), if the mean of the snapshots over the same nanosecond portion is used to calculate the value, the *x*-displacement moves into the expected range (∼−1.3 Å).[24]

The largest discrepancies between the B-RNA and B-DNA structures are in the base pair inclination and tip. Since the inclination at the terminal base pairs has risen above the average B-DNA values and moved closer to A-RNA values, the tip values more closely resemble the A-RNA values (black line), and since some of the puckers do display some transient C3′-endo puckers, perhaps it is simply necessary to continue the simulation for a longer time in order to see the B to A transition. The fact that we can stably simulate a B-RNA structure for longer than 2 ns suggests that the B-RNA structure is certainly a minimum energy conformation with the Cornell et al. force field. JUMNA[56] minimizations by Lavery (personal communication), with both the Cornell *et al.* and the standard JUNMA force fields, support this observation. If, as suggested by Olson *et al.*,[76] the barrier to sugar repuckering is on the order of ∼4 kcal/mol for RNA and ∼2 kcal/mol for DNA, the time for an A to B or B to A transition should be $\sim e^{-\Delta\Delta E^{\ddagger}/RT}$ or approximately 20 times longer for RNA. This implies a time scale on the order of ∼10 ns based on the observation of a A-DNA to B-DNA transition in ∼500 ps.[24]

Can We Force a B-RNA to A-RNA Transition? The B-RNA simulation displays more frequent C2′-endo to C3′-endo sugar repuckering with longer lasting C3′-endo puckers compared to the infrequent and short-lived C2′-endo puckers observed during C3′-endo to C2′-endo repuckering in the A-RNA simulation. This suggests that the transition to C3′-endo from C2′-endo in B-RNA is easier than the C3′-endo to C2′-endo repuckering in A-RNA. Moreover, looking at the black line in Figure 3 which displays some of the common indicators of A vs B form geometry as a function of time, transient spikes in the inclination and dips in the rise and *x*-displacement, such as just after 500 ps, are observed which indicate the structure becomes more "A-form" like during the B-RNA simulation. These observations indirectly suggest that the A-RNA structure is more stable, but that the barrier allowing the concerted repuckering necessary for a B-RNA to A-RNA transition is too high and cannot be surmounted in 1−2-ns simulations. To investigate this, simulations were run at a higher temperature, 400 K, which should allow for more frequent repuckering. Interestingly, although the increase in temperature to 400 K does lead to a slight increase in the rate of sugar repuckering, the repuckering is not nearly as pronounced as is observed in the DNA simulations. Although the temperature does not significantly increase the rate of sugar repuckering, it does lead to transient breaking of the Watson−Crick base pairs and a correspondingly large increase in the fluctuations within the helicoidal parameters. The largest disruption in the structure comes from terminal base pair fraying on one side of the helix where approximately halfway through the simulation the terminal CG base pair breaks and the base pairs from across the strand pack on top of each other. Given the disruption in the structure, it appears that 400 K is too "hot" for these

(76) Olson, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 270−278.

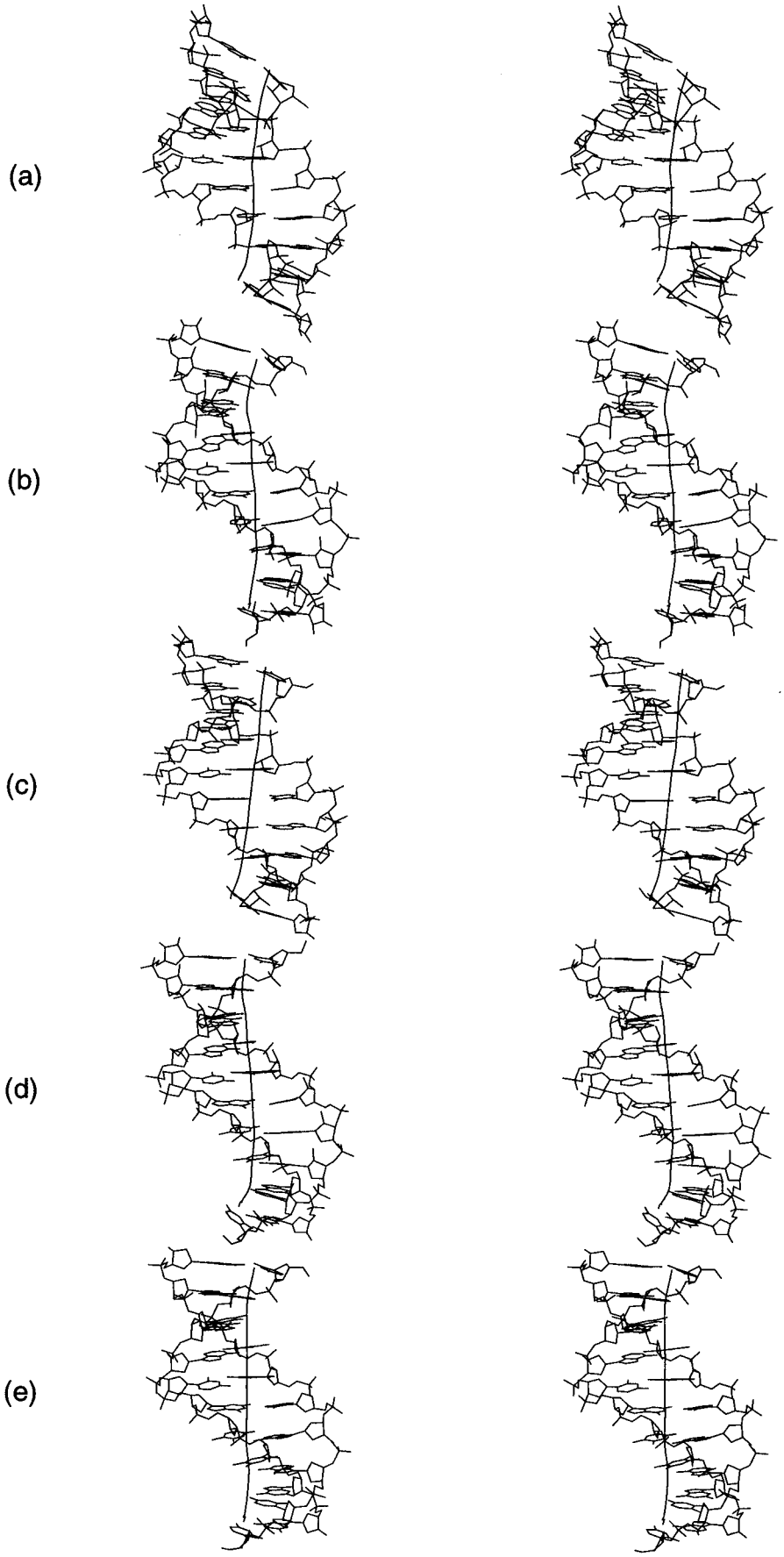

Figure 2. Average structures and the global helical axis (calculated from Curves 5.1[62]) for all of the average structures represented in the calculations described herein are plotted in stereo. Each structure represents the final nanosecond from their respective trajectories and is calculated from a straight coordinate average over all nucleic acid atoms from RMS fit coordinate frames taken at 1-ps intervals. All of the nucleic acid atoms, except the hydrogens, are displayed. All the plots were created using MidasPlus.[66] (a) **A-RNA**: canonical A start of the RNA:RNA duplex over 1030−2030 ps. (b) **B-RNA**: canonical B start of the RNA:RNA duplex over 1370−2370 ps. (c) **A-hybrid**: canonical A start of the DNA:RNA duplex over 1005−2005 ps. (d) **B-hybrid**: canonical B start of the DNA:RNA duplex over 1045−2045 ps. (e) **B-DNA**: canonical B start of the DNA:DNA duplex over 400−1400 ps.

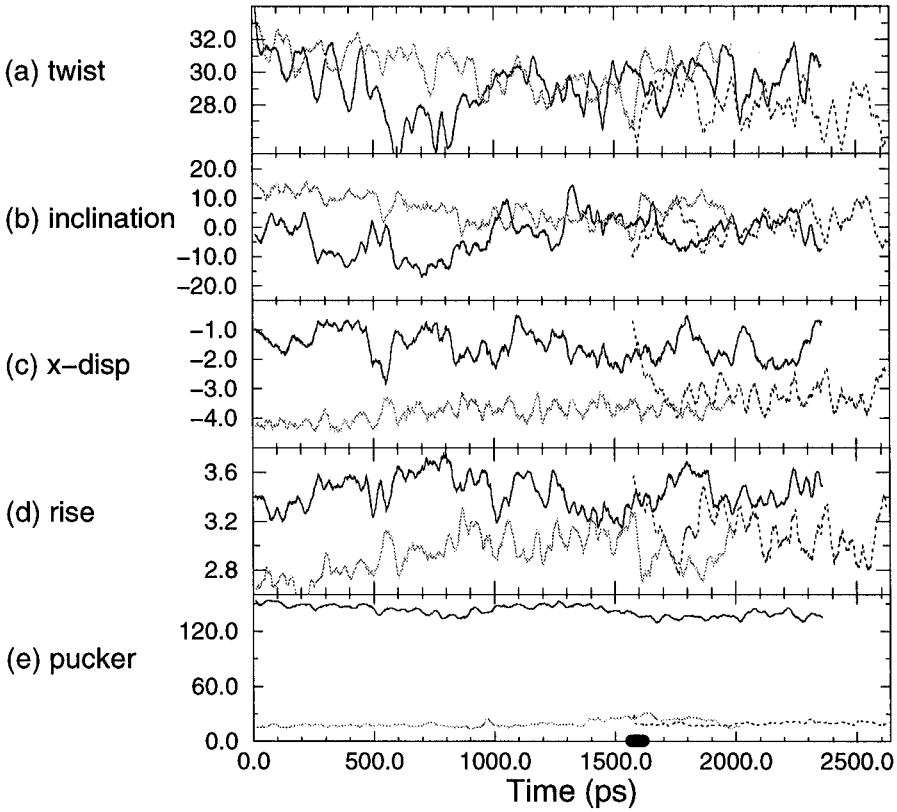

Figure 3. Plot of selected helicoidal parameters versus time for the B-RNA (black), A-RNA (gray), and B-RNA simulation continued from 1565 ps where a concerted flip in the puckers was forced over 50 ps (dashed black). The helical twist, base pair inclination, and sugar pucker pseudorotation phase (pucker) are in degrees and the rise between base pairs and *x*-displacement from the helical axis is in angstroms. The data represent an average over all nucleotides, base pairs, or base pair steps, as appropriate, and have been smoothed by performing a running average over 25 ps. The black oval along the *x*-axis of the pucker graph represents the time over which the restraints on the C1′−C2′−C3′−C4′ torsions were applied to induce a concerted flip in the sugar puckers.

simulations. While it may be possible to run at an elevated temperature somewhere between 300 and 400 K that still maintains the base pairing, it is not anticipated that this will provide for a significant enough rate enhancement in the sugar puckering to allow a concerted change to C3′-endo on a nanosecond time scale.

To see if we could observe a B-RNA to A-RNA transition more directly, a simulation was run where a concerted change in the pucker to C3′-endo was forced. As discussed in the methods section, this was accomplished by forcing the C1′−C2′−C3′−C4′ torsion over a short period of time (∼50 ps), to values which lead to C3′-endo puckers. As Figure 3 shows, when the C3′-endo restraints are applied, the rise and *x*-displacement (dotted lines) move rather quickly (over approximately 250 ps) from the B-RNA values (black) to the A-RNA values (gray). The restraints were only applied for 50 ps (a time corresponding to the back oval on the *x*-axis of the pucker graph) and after the restraints were removed, all the puckers remained C3′-endo for the ∼1-ns simulation (except for one short C3′-endo to C2′-endo repuckering event at one of the terminal guanines). An average structure calculated from the trajectory, starting after the first 70 ps and representing 1 ns of simulation, converged to within 1.07 Å of the average structure calculated over the last nanosecond of the A-RNA trajectory. Convergence to this average structure was not as close (ssRMSd ∼0.64 Å) as what was seen in the A-RNA simulation (which had a ssRMSd ∼0.34 Å). A major difference in these average structures is the backbone conformations.

As discussed previously, one of the strands in the A-RNA simulation displayed a crankshaft (α, γ: g−,g+ to t,t) transition between the CpG step; it was claimed that this allows for better cross-strand overlap[69] and increased separation of the base pairs which led to better interstrand guanine overlap. In the simulation with the concerted pucker flip, instead of the (α, γ) crankshaft transition at the CpG step, it is observed at the adjacent ApC step in both strands. Moreover, the second strand additionally displays a crankshaft transition at the UpG step and a *trans* γ at the first cytosine; the latter was also observed in the second strand of the A-RNA average structure. The crankshaft transition at the ApC step in both strands occurs within the first 3−4 ps as the sugar repuckers to C3′-endo under the influence of the applied restraints. The sudden transition in the sugar pucker to C3′-endo also effects the α, ϵ, and ζ backbone angles at this step. The most uncharacteristic backbone angle changes occur at the ApC step of the first strand. At this step, β is *gauche*+ rather than *trans*, and remains *gauche*+ throughout the 1-ns simulation. The characteristic B_I (ϵ,ζ:t,g−) backbone conformation is never observed at this step. The likely reason for the anomalous behavior at this step is that the backbone was in a B_{II} or (ϵ:ζ;g−,t) backbone conformation at the start of the flip in pucker to C3′-endo. Since these B_{II} backbone conformations are not observed in the A-RNA simulation, this suggests that they are unfavorable in A-form structures, which may explain the unexpected behavior. It is somewhat surprising that during 1 ns of dynamics, the structure does not transition to the more characteristic B_I backbone conformation at this step. Moreover, it appears that "crankshaft" backbone transitions are rare events in RNA simulations.

Note that with the exception of the anomalies mentioned above, all of the other backbone angles are in the expected range and we also still observe the distinctive rise and low inclination at the CpG step. However, instead of a low twist at the CpG step, the twist is above average at 36.4°. Interstrand stacking of the guanines is still allowed by compensating lower twists

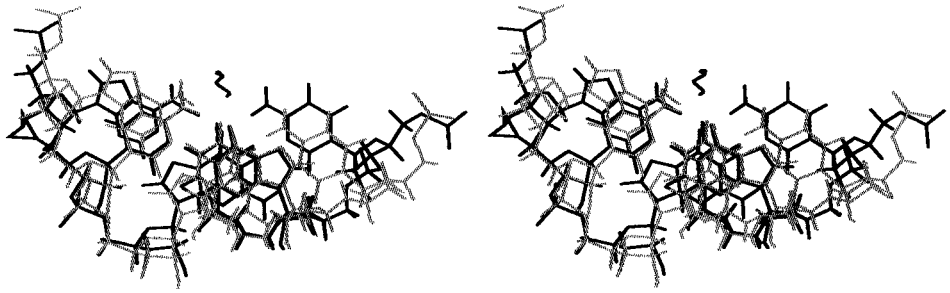

Figure 4. Stereoview plots, generated with MidasPlus,[66] of three stacked base pairs from average structures calculated from the B-RNA simulation where a concerted flip in puckers was forced (70−1070 ps, in black) and the A-RNA simulation (1030−2030 ps, in gray). The structures were RMS fit to these three base pairs and the view is looking down the helical axis (shown as the line visible in the top center of the figure in black) with the C_5-G_{16} base pair on top of G_6−C_{15} and the T_7-A_{14} on the bottom.

and slide values at the adjacent ApC (16.5°, −1.4 Å) and GpT (20.1°, −1.2 Å) steps. A slightly higher than average *x*-displacement (which is −5.0 Å) is also observed at the CG (−4.7 Å) and GC (−4.8 Å) base pairs. This backbone and helicoidal arrangement apparently leads to slightly better overlap of the guanines at the central step. This can be seen by examining the stereoview plot shown in Figure 4. This shows a view down the helical axis (shown in the upper center of the picture in black) with the C_5-G_{16} base pair stacking on top of the G_6-C_{15} and T_7-A_{15} base pairs. In black is shown the nanosecond time averaged structure from the simulation with the concerted flip in the puckers. The displacement of the top guanine from the helical axis is less and it appears to better stack on the guanine below it than is seen in the average structure from the A-RNA simulation (gray). Analysis of the interaction energy of the two guanines *in vacuo*, using the coordinates of the average structures after they have been briefly minimized to an RMS energy gradient of 1.0 kcal/mol with a distance dependent dielectric function and dielectric constant of 4, applying the program ANAL from AMBER 4.1,[20] suggests that the concerted pucker flip guanine stacking energy is indeed more favorable than that of the A-RNA. However, the better stacking of the guanine leaves the paired cytosines even less favorably stacked and more solvent exposed as can be seen in Figure 4.

Looking back at Figure 1 and the data for the B-RNA simulation (in gray), the low twist at the CpG and adjacent steps is not seen. However, with the change in puckers to C3′-endo B-RNA converts to A-RNA and the low twist (albeit at the steps adjacent to the CpG and not the CpG step itself) and interstrand stacking of the guanines appears. This is interesting since the backbone compensates immediately to the pucker change at the steps adjacent to the CpG step. This observation also provides further evidence that the intrastrand stacking is a real contextual sequence dependent structural effect. The helicoidal parameters do take some time to convert to A-RNA values. In Figure 5, snapshots from the simulation where the concerted flip in pucker to C3′-endo are displayed. From these, it is clear that application of the restraints causes massive structural perturbation. When the pucker is converted to C3′-endo, this immediately decreases the intrastand phosphate distance leading to significant base pair buckling. The conversion to C3′-endo was not done smoothly, but abruptly as can be seen by looking at the snapshots in Figure 5. By 40 ps into the simulation, the terminal base pair is almost broken due to significant propeller twisting. However, by 70 ps into the simulation (or 20 ps after the termination of the restraints), the structure begins to settle down and display more reasonable helicoidal values. Figure 5 shows that the structure can react to the fairly drastic and quick transition from C2′-endo to C3′-endo puckers without completely breaking up. Short term effects on the helicoidal parameters are clearly evident. More importantly, backbone angles do get caught in specific conformations, such as the *gauche*+ ApC β angle from the first strand which persists for longer than 1 ns after the restraints are removed.

The A-RNA Structure Is More Rigid Than B-RNA or B-DNA. Despite the observation of a crankshaft (α, γ: g−,g+ to t,t) transition at the CpG step in one of the strands of the A-RNA average structure, the self convergence of the A-RNA average structure (ssRMSd = 0.34 Å) is much better than is observed in the B-RNA structure (1.0 Å). This high level of self convergence is not unexpected, since as mentioned in the introduction, the A family of nucleic acid structures are expected to be more rigid than the B family. The rigidity of the A-RNA structure is readily apparent in the simulation. Little sugar repuckering and no correlated (ϵ, ζ: t,g− to g−,t) B_I to B_{II} backbone transitions, both of which are seen frequently in DNA simulations, are observed. The sugar repuckering that is seen in the A-RNA simulation is limited; the only events that are observed occur at the terminal guanine residues where G_{10} repuckers twice, once for ∼10 ps and a second time for ∼500 ps, and G_{20} repuckers once for ∼100 ps. The overall fluctuations in the backbone torsion angles (see Table 2) are also reduced in the range of 10−50% compared to simulations of the RNA (or DNA) started in the B family, with the angles related to the pucker displaying the largest reduction in relative fluctuations (δ,γ,χ). Although the fluctuations in the backbone angles are reduced, we still observe higher than expected fluctuations in roll, tilt, and twist as seen in previous simulations.[24]

Interestingly, the B-RNA structure is not only more flexible than A-RNA but also more flexible than the B-DNA simulations, as judged by looking at the standard deviations in backbone angles and helicoidal values presented in Table 2. All the values show enhanced fluctuation, except for the sugar pucker (and therefore the sugar pucker amplitude and δ backbone angle) and the ϵ and ζ backbone angles. Although the sugar repuckering is more frequent in the B-RNA simulation, where greater than 10 events longer than ∼100 ps are observed, than in the A-RNA simulation (where as previously discussed only the terminal guanines repuckered), the repuckering is considerably less frequent than is observed in the B-DNA simulations (*i.e.* roughly 10% of what is seen in the B-DNA simulations). This is presumably due to the larger barrier to repuckering in RNA.[76] The fluctuations in ϵ and ζ backbone angles are also less since the B_I to B_{II} backbone transitions are less frequent; however, these B_I to B_{II} backbone transitions do occur in the B-RNA structure, which suggests that the lack of these transitions in A-RNA is not due to the presence of the O2′ hydroxyl group.

DNA:RNA Hybrids−Flexible DNA and Rigid RNA. Hybrid duplexes which have one of the strands RNA and the other DNA, as previously discussed, are known to adopt a "mixed" form between a canonical A and B geometry. Simulations of DNA:RNA hybrid duplex structures (d[CCAACGT-

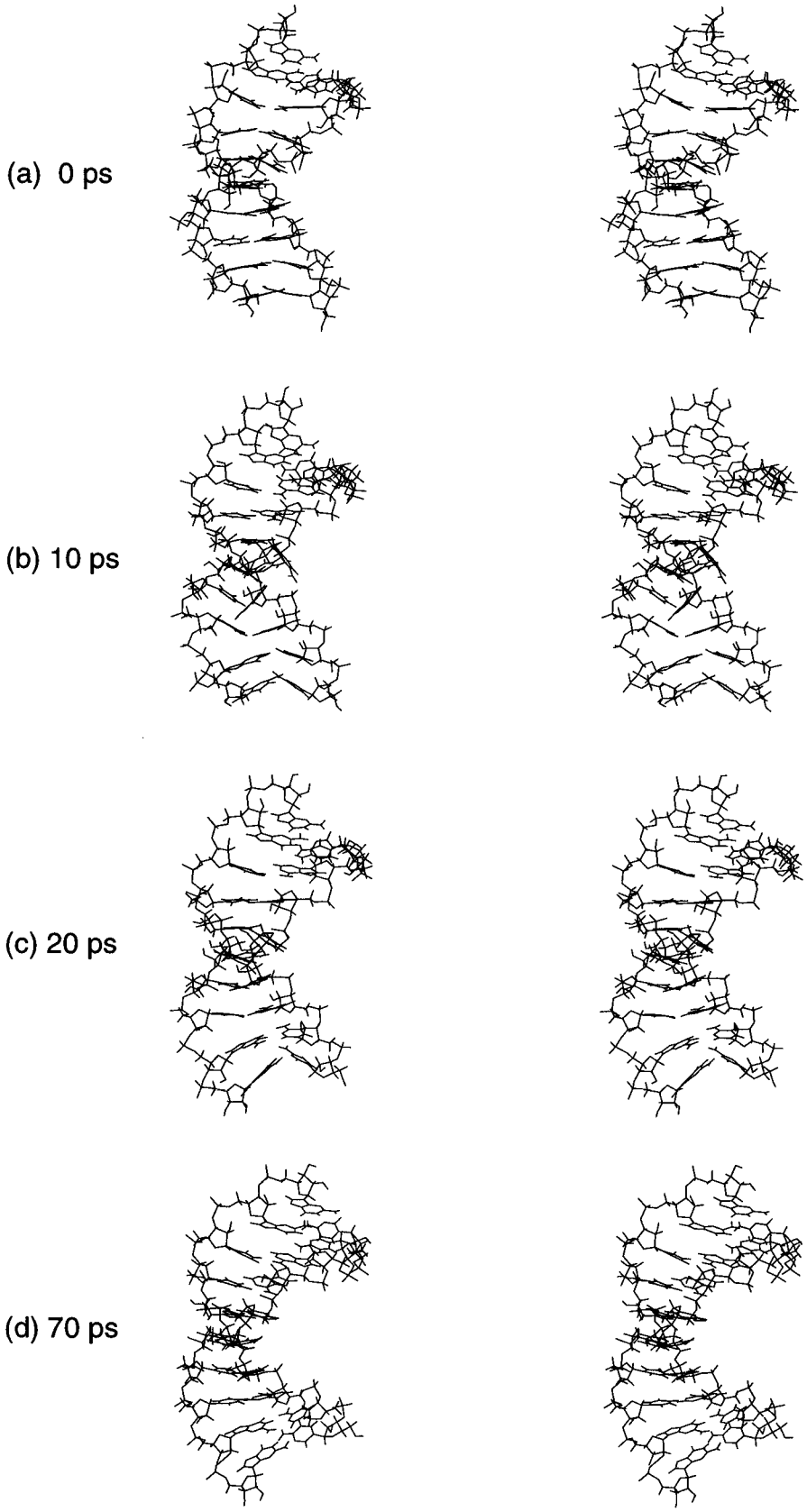

Figure 5. Snapshots from the simulation of B-RNA where a concerted flip in the puckers was forced, in stereo, are displayed in plots (a) through (d), representing the structure at various intervals. At 0 ps (a), no restraints have yet been applied, and by 50 ps, all restraints have been removed, as is discussed in the Methods section. The snapshots were all atom RMS fit to a common reference frame prior to display.

TGG]−r[CCAACGUUGG]) started in both canonical A and canonical B structures for ∼2 ns each were performed to determine if molecular dynamics simulations with the Cornell et al.[18] force field could accurately represent the structure of hybrid duplexes. In these simulations, as seen in the RNA:RNA simulations, the RNA strand remains in either a canonical A or canonical B geometry depending on the initial RNA conformation. The DNA strand on the other hand, as in the DNA:DNA simulations, undergoes an A-DNA to B-DNA transition regardless of its, or the RNA strands, starting structure.

As shown in Table 1, the DNA strand is closer to a canonical B geometry when the hybrid is started from a canonical B

geometry (RMSd ~1.71 Å to canonical B-DNA and ~0.80 Å to the average B-DNA) than when it is started from a canonical A geometry (RMSd ~2.65 Å to canonical B-DNA and ~1.46 Å to average B-DNA). This is not surprising since, as seen in the RNA:RNA duplex B-RNA simulation, the RNA strand does not convert to a A geometry during 2+ ns of simulation. When the hybrid is started in a canonical A geometry, the DNA strand moves away from an A-form geometry but does not go quite all the way to a B-form geometry which demonstrates that it is clearly influenced by the RNA strand's A-like conformation. This is of particular interest since it is this structure that is most consistent with and relevant to what has been seen experimentally.

In Figure 2c and 2d, average structures over the last nanosecond from the two trajectories are displayed to allow comparison to the RNA:RNA and DNA:DNA simulations. The average structure from the DNA:RNA hybrid duplex simulation started in a A geometry (A-hybrid) shown in Figure 2c has a surprisingly straight and regular helical axis compared to the other structures displayed. There is no kink at the TpG steps as is seen in the B-RNA and B-DNA average structures since there is not any significant bending into the major groove at this step. There is no change in direction of the helical axis at the central CpG step as is seen in the A-RNA structure, since this average structure does not have the low twist, high rise, and interstrand guanine stacking seen in the A-RNA structure at the central CpG step. As will be discussed in more detail later, the A-hybrid structure is less inclined and less bent overall, with a narrower minor groove than is seen in the corresponding A-RNA structure.

The average structure calculated from the final nanosecond of the DNA:RNA hybrid duplex simulation started in a B geometry (B-hybrid) shown in Figure 2d is very similar, as is expected based on the low RMSd values reported in Table 1, to the B-DNA (Figure 2e) and B-RNA (Figure 2b) structures. Noteworthy is the kink in the helical axis and inclination of the cytosine at the bottom of the structure (Figure 2d). Looking at the RNA strand (on the right) it can be seen that the O2′ hydroxyl of the first cytosine residue (bottom) is pointing away from the helical axis rather than up and along the helical axis as is seen in the other RNA nucleotides. This is because this cytosine repuckered to C3′-endo and remained C3′-endo during the final nanosecond of the simulation from which the average structure was created. The 3′-terminal guanine ribonucleotide also repuckers frequently during the simulation and some of the interior steps also have some persistent C3′-endo puckers during the final nanosecond of the simulation (C_{12} for ~50 ps and U_{18} for ~500 ps). These repuckering events, which are more frequent and more persistent than was observed in the simulation of the B-RNA duplex, suggest that perhaps a longer simulation may allow the RNA strand to convert to A geometry. However, simulations run at 400 K were not sufficient to drive the transition; instead, as was observed in the simulation of the B-RNA at 400 K, the structure became distorted and the terminal base pairs frayed and stacked upon each other.

The A-hybrid average structure displays many properties that are consistent with experimental results. In particular, the DNA strand has sugar puckers that are primarily C2′-endo whereas the RNA strand has nearly all C3′-endo sugar puckers throughout. As shown in Table 2, the helicoidal parameters are also consistent with what is known from experiment about DNA:RNA hybrid structure. Specifically, the A-hybrid average structure is positively inclined (11.1°), has a small positive roll in the major groove (1.6°), small positive buckle (4.3°), negative propeller twist (−13.2°), and a negative *x*-displacement from the helical axis (−4.5 Å). In Figure 6, the individual backbone

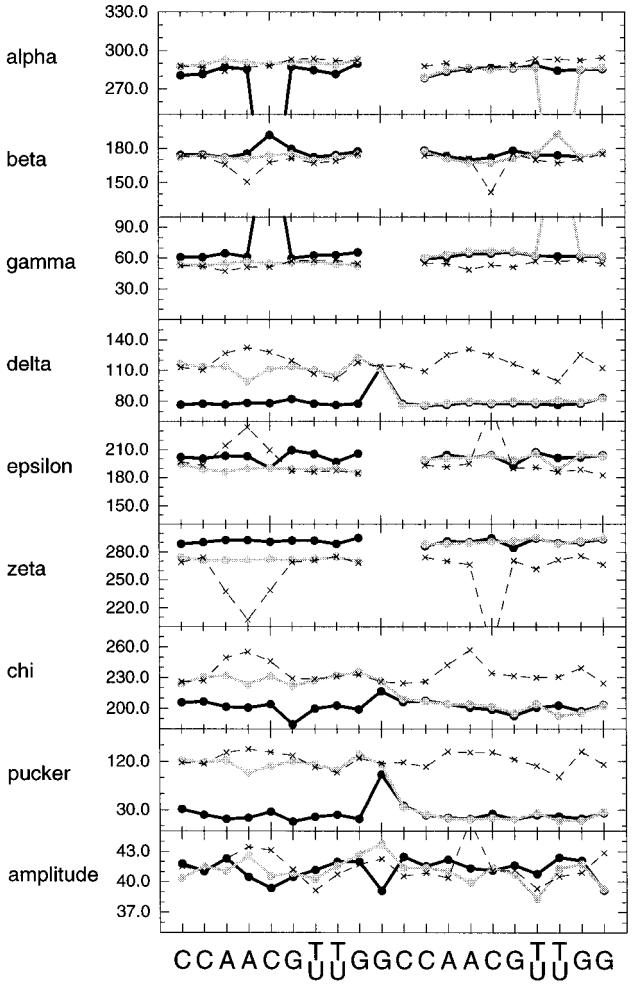

Figure 6. Average backbone angles along the sequence from 5′ to 3′ for the first strand followed by the second strand (from left to right) for average structures from the A-RNA (black, 1030−2030 ps), A-hybrid (gray, 1005−2005 ps), and B-DNA (dashed black, 400−1400 ps) trajectories. All the angles are listed in degrees.

angles for all the nucleotides are represented for the A-hybrid structure (gray), A-RNA (black), and B-DNA (dashed black). From this figure the (α, γ) crankshaft transition, discussed previously, at the CpG step in the first strand of the A-RNA is readily apparent. A similar crankshaft transition, at the UpU step, can be seen in the RNA strand of the A-hybrid structure (top right of Figure 6). In general the A-hybrid angles of the DNA strand (gray, left side) tend to match the B-DNA angles (black dashed) and the angles of the RNA strand (gray, right side) tend to match the A-RNA angles (black). We see the expected trend in the A-hybrid structure, as discussed in the introduction, that the α angle is lower and the ϵ, ζ, and γ angles slightly higher in the RNA strand (gray, right) than in the DNA strand (gray, left). The average ϵ and ζ angles at each step for the B-RNA (not shown), A-RNA, and A-hybrid are all similar, in contrast to the B-DNA simulations where peaks and troughs in ϵ and ζ are evident. These peaks and troughs result from relatively more frequent B_I to B_{II} backbone transitions which push the average ϵ closer to *gauche*− and the average ζ closer to *trans* in the B-DNA and B-hybrid (data not shown) simulations. Despite the similarity of the average ϵ and ζ angles in the B-RNA and A-hybrid simulations, B_I to B_{II} backbone transitions are still observed in both the DNA and RNA strands, albeit at a lower rate than is seen in the B-DNA simulations. These transitions are more frequent and correlated in the DNA strand than in the RNA strand. During more than 2 ns of simulations of A-RNA, B_I to B_{II} backbone transitions were never observed. This provides further evidence that B_I to B_{II}

Table 3. Minor Groove Widths in the Various Models (as denoted by the row headings) Represented by Selected Inter-Strand Phosphate Distances (as specified in column 1) (in angstroms)[a]

	A-RNA 1030−2030 ps	B-RNA 1370−2370 ps	A-hybrid 1005−2005 ps	B-hybrid 1045−2045 ps	B-DNA 400−1400 ps
P_5-P_{20}	17.32 (0.69)	15.52 (1.58)	15.28 (0.90)	13.53 (1.46)	14.19 (1.13)
P_6-P_{19}	17.12 (0.68)	12.72 (1.31)	15.72 (0.78)	12.15 (1.18)	12.66 (1.53)
P_7-P_{18}	16.60 (0.54)	11.17 (0.97)	15.07 (0.81)	10.59 (0.87)	11.10 (1.07)
P_8-P_{17}	16.14 (0.53)	10.72 (1.00)	14.82 (0.80)	11.17 (1.31)	10.98 (1.21)
P_9-P_{16}	16.93 (0.58)	11.87 (1.67)	15.47 (0.88)	12.72 (1.61)	11.65 (1.48)
$P_{10}-P_{15}$	17.69 (0.71)	15.10 (1.90)	15.94 (0.98)	13.51 (1.58)	14.05 (1.14)

[a] Distances were selected to match those chosen in analysis of the B-DNA crystal structure[77] and previously reported calculations.[24] The distances are averages over 1 ns, in angstroms, and the standard deviations are given in parentheses.

transitions are easier in B-form structures and that the B_I state is preferred in A-form structures.

The groove widths are also consistent with the experimental data on DNA:RNA hybrid duplex structures. Shown in Table 3 are average interstrand phosphate distances (and standard deviations in parentheses) across the minor groove for each of the simulations. The A-RNA (left) has the widest minor groove and little sequence specific narrowing. The B-form structures of DNA, RNA, and the B-hybrid have narrow minor grooves and display sequence specific narrowing in the center of the helix similar to that observed in the crystal structure.[77] The A-hybrid structure has a minor groove width intermediate between the A-RNA and B-DNA structures. There is also no significant sequence specific narrowing at the center of this duplex. Overall the A-hybrid structure is closer to an A-form geometry than a B-form geometry, although the DNA strand has primarily C2′-endo sugar puckers.

C2′-endo sugar puckers tend to increase the intrastrand phosphate separation and therefore B-DNA structures have a larger rise between base pairs and are longer overall. This brings up an interesting question with respect to the structure of DNA:RNA hybrids of longer sequence. As the helix becomes longer, how can the structure compensate for the larger discrepancy between the longer end to end length of the DNA strand and the shorter RNA strand; will the DNA strand "shrink" or perhaps present more C3′-endo puckers, in order to maintain good structure? Clearly hybrid structures can adopt a canonical A geometry (which would alleviate this problem) as has been seen in crystal structures.[41,42] The current simulations do not address this question. However, these simulations do further support the observation that the B structure is more stable for DNA, as A to B transitions have been seen for the DNA strand in DNA:DNA and DNA:RNA hybrid duplexes. Moreover, it is clear that molecular dynamics simulations can reasonably represent the difference in structure between duplex A-RNA, DNA:RNA hybrids, and B-DNA.

DNA:RNA Hybrids and Nucleic Acid Sugar Repuckering. The time course of the individual sugar puckers at 1-ps intervals and the histograms of the individual sugar puckers are shown in Figure 6. The DNA strand is represented in Figures 7a (pucker versus time) and 7b (histogram) with a graph for each nucleotide from the 5′ (top) to the 3′ end (bottom). The corresponding RNA base pair is shown alongside in Figure 7c (pucker versus time) and Figure 7d (histogram) for the RNA strand from the 3′ (top) to the 5′ end (bottom). The one letter code for each nucleotide is specified in the upper right of the histograms (Figures 7b and 7d). From the data in Figure 7a, it is clear that the DNA strand is repuckering throughout the simulation. Also from this figure, the time course of the transition from C3′-endo to C2′-endo puckers, indicative of the A-DNA to B-DNA transition in the left strand, is also evident. Most of the DNA sugars have repuckered within the first 100 ps and all have repuckered by ∼300 ps. This is similar to the time course seen in the DNA:DNA duplex simulations. The histograms of the DNA strand (Figure 2b) show a mix of puckers with C2′-endo puckers favored. The RNA strand (shown in Figures 7c and 7d), on the other hand, is clearly not repuckering, except for one repuckering event at the terminal guanine, and C3′-endo puckers dominate (Figure 2d). Although the data are not shown, the number of sugar repuckering events seen in the RNA strand of the A-hybrid is consistent with the A-RNA duplex simulation. B-RNA, in both the B-hybrid and B-RNA simulations, does repucker more frequently than A-RNA, however at a significantly lower rate than is seen in the DNA simulations. In general, all of these simulations suggest that the RNA does not repucker too frequently on a nanosecond time scale. In the B-RNA simulations the repuckering from C2′-endo to C3′-endo is longer lived than is seen in the DNA simulations, which implies, as expected,[76] that the barrier to repuckering is higher in RNA than in DNA. The repuckering of the sugars in the DNA strand of the A-hybrid, despite the presence of the RNA strand, occurs at a similar rate and gives a similar distribution to that seen in the corresponding B-DNA duplex simulations.

Inherent Fluctuations during Molecular Dynamics. The standard deviations, or fluctuations, in the backbone angles and helicoidal parameters, calculated based on analysis of 1-ns portions of the trajectories at 1-ps intervals, are presented in parentheses in Table 2. All of the B-form structures have larger fluctuations than are seen in the A-form structures. The greater flexibility of the B-form structures is consistent with experiment since, as discussed in the introduction, A-form structures are generally found to be more rigid than B-form structures. What does this flexibility, on a picosecond to nanosecond time scale, mean and where does it come from? The fluctuations reported in Table 2 relate to differences both in the simple anharmonic atomic motions and short time scale (ps) collective motions such as base pair propeller twisting, sliding, and helical twisting and bending. Sugar repuckering may also influence the flexibility; RNA, with a larger barrier to sugar repuckering might be expected to be more rigid than DNA.[76] Moreover, backbone transitions (B_I to B_{II}, (α,γ) crankshaft, *etc.*) are expected to also influence the dynamics. However, the data in Table 2 suggest the enhanced flexibility of B-form over A-form structures is not entirely due to more frequent repuckering or the presence of backbone transitions. Considerably less repuckering is seen in the RNA simulations, yet both the B-RNA and the B-hybrid structures have fluctuations in the backbone angles and helicoidal parameters that are higher than are seen in A-RNA. DNA simulations where the puckers are held fixed at C3′-endo and C2′-endo also confirm that the observed flexibility is not dependent on repuckering; when the puckers are forced to remain C3′-endo (and the structure moves closer to a A-form geometry), the fluctuations are significantly damped. When the pucker is held fixed at C2′-endo, the fluctuations are slightly enhanced with respect to simulations where the pucker is not

(77) Prive, G. G.; Yanagi, K.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 177−99.

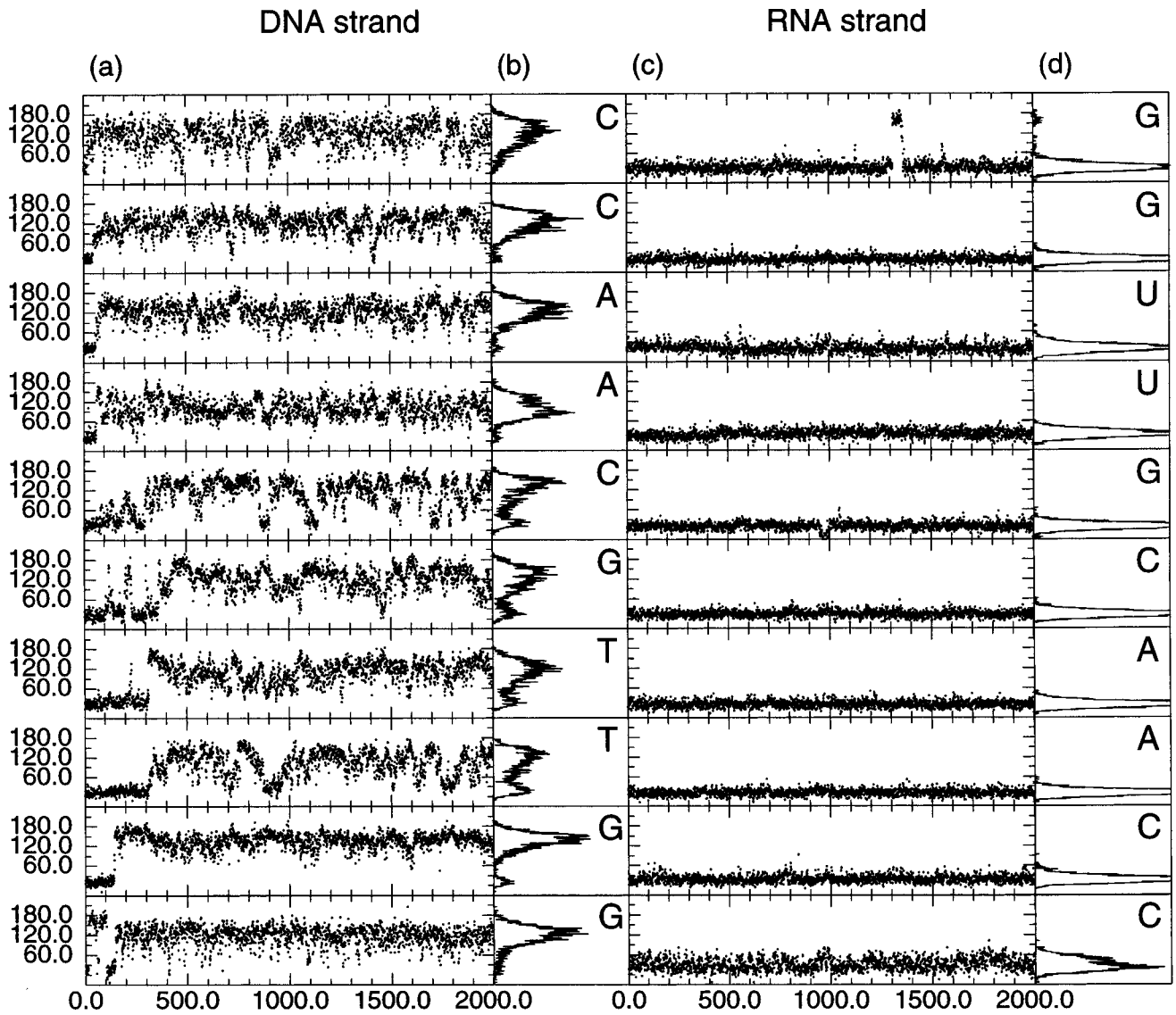

Figure 7. Sugar pucker pseudorotation phase (degrees) versus time (ps) and histograms for each individual nucleotide from the A-hybrid simulation: (a) the pucker versus time from top to bottom for the DNA strand from 5′ to 3′; (b) the histogram of the pucker from top to bottom for the DNA strand from 5′ to 3′ along with 1-letter code labels for the DNA residues; (c and d) the same as parts (a) and (b) except for the RNA strand from 3′ to 5′.

held fixed (data not shown). The flexibility also is not due solely to B_I to B_{II} transitions in the backbone since these do not happen as frequently in B-RNA as in the B-DNA, yet B-RNA shows a similar magnitude in the fluctuations. However, the absence of B_I to B_{II} transitions in the A-RNA simulation may partially explain the rigidity. Although the B-hybrid structure does appear to be the most flexible, as can be seen by examining the fluctuations in Table 2, the high relative fluctuations are mostly the result of large fluctuations in the individual angles, base pairs, and base pair steps at the 5′-end of the RNA strand. As mentioned previously, the 5′-terminal cytosine in the RNA strand of the B-hybrid repuckers to C3′-endo. This observation, coupled with the enhanced flexibility, indirectly suggests that the B-hybrid is in a higher energy state.

The high flexibility of all the B-form structures suggests that the free energy landscape around the minimum is flatter, and the B-form is a broader energy minimum allowing more picosecond time scale motions, such as base pair twisting, sliding, and propeller twisting, among other collective motions.[78] The A-form geometry of RNA on the other hand may represent a deeper and tighter minimum. This helps explain why it is hard to find the A-RNA state during the B-RNA dynamics; it is difficult to transition from the broad flat energy minimum of the B-form geometry representing a high configurational entropy state to the more ordered A-state, not to mention the difficulty in overcoming the concerted barriers to sugar repuckering. A comparison of the atomic positional fluctuations over nanosecond portions of all the trajectories (data not shown) clearly demonstrates that the individual atomic motions in the B-form structures (B-DNA, B-RNA, B-hybrid) are uniformly higher than those observed in the A-form structures (A-RNA, A-hybrid). The average atomic positional fluctuations are roughly 20−30% higher in the B-form structures (B-RNA = 1.35 ± 0.40 Å, B-hybrid = 1.37 ± 0.41 Å, B-DNA = 1.24 ± 0.36 Å) than the A-form structures (A-RNA = 1.09 ± 0.35 Å, A-hybrid = 1.03 ± 0.29 Å). The differences in relative flexibility in the two states may represent a mechanism for recognition and distinction between RNA, DNA, and hybrid duplexes beyond the obvious differences in structure.

Hydration: B-Form Structures and Flexibility Revisited. The relative flexibility of B-form and A-form structures, as mentioned above, is clearly not due entirely to backbone transitions, sugar repuckering, or a significant difference in the

(78) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231−252.

in vacuo low-frequency normal modes (not shown). Perhaps hydration or the associated salts play a role in the relative rigidities? It is clear that water is an integral part of nucleic acid structure. Since water, and likely also salt, is an integral part of the structure, it is likely that both the water and salt influence the dynamics. Perhaps the A-form geometry is partially rigidified by more tightly bound or specifically associated water and counterions? This is a difficult question to answer directly. However, our simulations suggest that the more flexible structures have "less" associated water and counterions. Additionally, the A-form structure seems to have more counterions specifically associated in the major groove, and more specific hydration in the grooves, than B-form structures which could lead to higher stability and less flexibility (discussed below).

Similar to what has been done in the analysis of water in nucleic acid crystal structures,[79,80] characterization of pharmacophores,[81] and analysis of counterion density around DNA,[82] atom density on a grid has been contoured (as discussed in the methods section). In Figure 8 water density around the average structures computed over 1-ns portions of the trajectories is displayed. The purpose of this current analysis is not to precisely map out the locations of the waters in the simulations, but to provide a general picture of the overall hydration and to characterize the distinction between the hydration of B-DNA (Figure 8b), B-RNA (Figure 8c) and the B-hybrid (Figure 8d) structures contoured at an equivalent level (15.0 hits per 0.5 Å^3 grid element, or ∼3.6 times the expected water density). It should be noted that the structures shown are the average structures from the trajectory; an average structure does not clearly show the relative range of motion of all the atoms in the duplex over the course of the simulation. In general, each individual 1-ps frame deviates from the average structure by on the order of ∼1.3 ± 0.3 Å, with B-DNA closest to the average (1.2 Å) and the B-hybrid furthest from the average (1.35 Å). Each individual frame never gets closer than 0.6 Å, nor further than 2.4 Å, from the average structure over the final nanosecond of each simulation.

Figure 8a displays the B-DNA average structure contoured at 12.0 hits per 0.5 Å^3 grid element or ∼2.9 times bulk water density. The condensation of the water around the DNA is clearly evident. The highest density appears in the minor groove where waters which directly hydrogen bond to the nucleic acid, and secondary waters directly on top of those waters, can be visualized. The former sit deep in the minor groove and interact with the bases and the sugar O4′ oxygens. At the lower contour levels, two to three waters per base pair step are visible, except in the narrow portion of the minor groove at the center of the duplex, where one water per base pair step deep in the groove is visible. The most obvious hydration pattern is this "spine of hydration"[83,84] which extends out of the center of the duplex in the minor groove. Density is also clearly visibly associated with the backbone and major groove as well, especially at the lower contour levels (Figure 8a). The beginning of a spine of hydration in the major groove is visible at both ends of the helix which twists the other direction around the helical axis compared to the minor groove and backbone spines of hydration. In Figure 8, very little preferential hydration of the phosphate groups is visible, however water density can be seen off the bisector of the phosphate oxygens. This is likely due to the diffuse nature of the "cone of hydration"[85,86] around the phosphates and the enhanced mobility of the phosphate atoms with respect to the other nucleic acid atoms. Based on this analysis, no specific hydration of the O2′ hydroxyls by water oxygen atoms in the B-RNA is seen. This is most likely since the O2′ hydroxyls spend a significant time hydrogen bonded to the O5′ and/or O1P atoms of the following residue. The 3′-terminal guanines have no following residue to interact with, hence the C1′−C2′−O2′−HO2′ torsion is essentially freely spinning.

Comparing the various B-form simulations (which all converged to the same structure) we generally see equivalent hydration patterns in the minor groove. However, at an equivalent water density contour, the water occupancy appears to be highest in the B-DNA and lowest in the B-hybrid structure. The trend here mimics the trends in relative flexibility, with the more flexible B-hybrid appearing to have lower water occupancy. The trend does not result from the solvent diffusing more rapidly, on average, in the B-hybrid than the other simulations since the average diffusion of water in all of these simulations is comparable.

The lower density observed in the more flexible structures could be an artifact of the coordinate fitting and water oxygen atom gridding and visualization procedure. An example of artifactual behavior is readily apparent in Figure 8 where it appears that the ends of the duplex are less hydrated. In order to create the grid of water density, each snapshot is RMS fit to a common reference frame, which in this case was all the nucleic acid atoms, and then the grid is constructed. Since the ends of nucleic acids are more flexible, the density appears lower. This is similar to what is seen in the crystal structures. The higher mobility leads to relatively less and more irregular solvent density. In fact, it was not until the low temperature crystal structures[87] (which effectively reduce the thermal fluctuations) that explicit water density around the phosphates was visualized. This is the basic point. If water is in more regular positions, such as is seen in both the grooves of A-RNA (discussed below), it will be easily visualized and moreover may tend to rigidify the structure. Alternatively, if the structure is more flexible, less water density anchoring both sides of the grooves (for example) may be present. Of course, a causal relation is not apparent; while it is clear that a more rigid structure will lead to more well-defined water positions, these simulations do not determine if water rigidifies the structure.

It should be noted that if the grid is built around a set of coordinates that are RMS fit to only the first base pair of the helix, specific water density is clearly visible at the end of the helix (data not shown). When the grid is instead built from snapshots RMS fit to the central two base pairs, the apparent water density is increased particularly at the central base pairs and along the backbone (and virtually absent at the terminii). Qualitatively, the location of the water hydration is not altered. This implies that in order to specifically analyze the hydration each individual base pair, or base pair step, should be RMS fit and the water density independently calculated to remove these dynamic effects. When this is done, the hydration results are in general agreement with the analysis of crystal structures presented to date.[79,80,88]

From the pictures presented in Figure 8, the minor groove in the B-form geometry is clearly preferentially hydrated and is

(79) Schneider, B.; Cohen, D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M. *Biophys. J.* **1993**, *65*, 2291−303.
(80) Umrania, Y.; Nikjoo, H.; Goodfellow, J. M. *Int. J. Radiat. Biol.* **1995**, *67*, 145−52.
(81) Rosenfield, R. E. J.; Swanson, S. M.; Meyer, E. F. J.; Carrell, H. L.; Murray-Rust, P. *J. Mol. Graphics* **1984**, *2*, 43−46.
(82) Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Phys. Chem.* **1995**, *99*, 11591−11599.
(83) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1983**, *163*, 129−46.
(84) Subramanian, P. S.; Ravishanker, G.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1836−40.
(85) Pullman, A.; Pullman, B. *Annu. Rev. Biophys.* **1975**, *7*, 505−566.
(86) Subramanian, P. S.; Beveridge, D. L. *J. Biomol. Struct. Dyn.* **1989**, *6*, 1093−1122.
(87) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535−56.
(88) Schneider, B.; Berman, H. M. *Biophys. J.* **1995**, *69*, 2661−9.

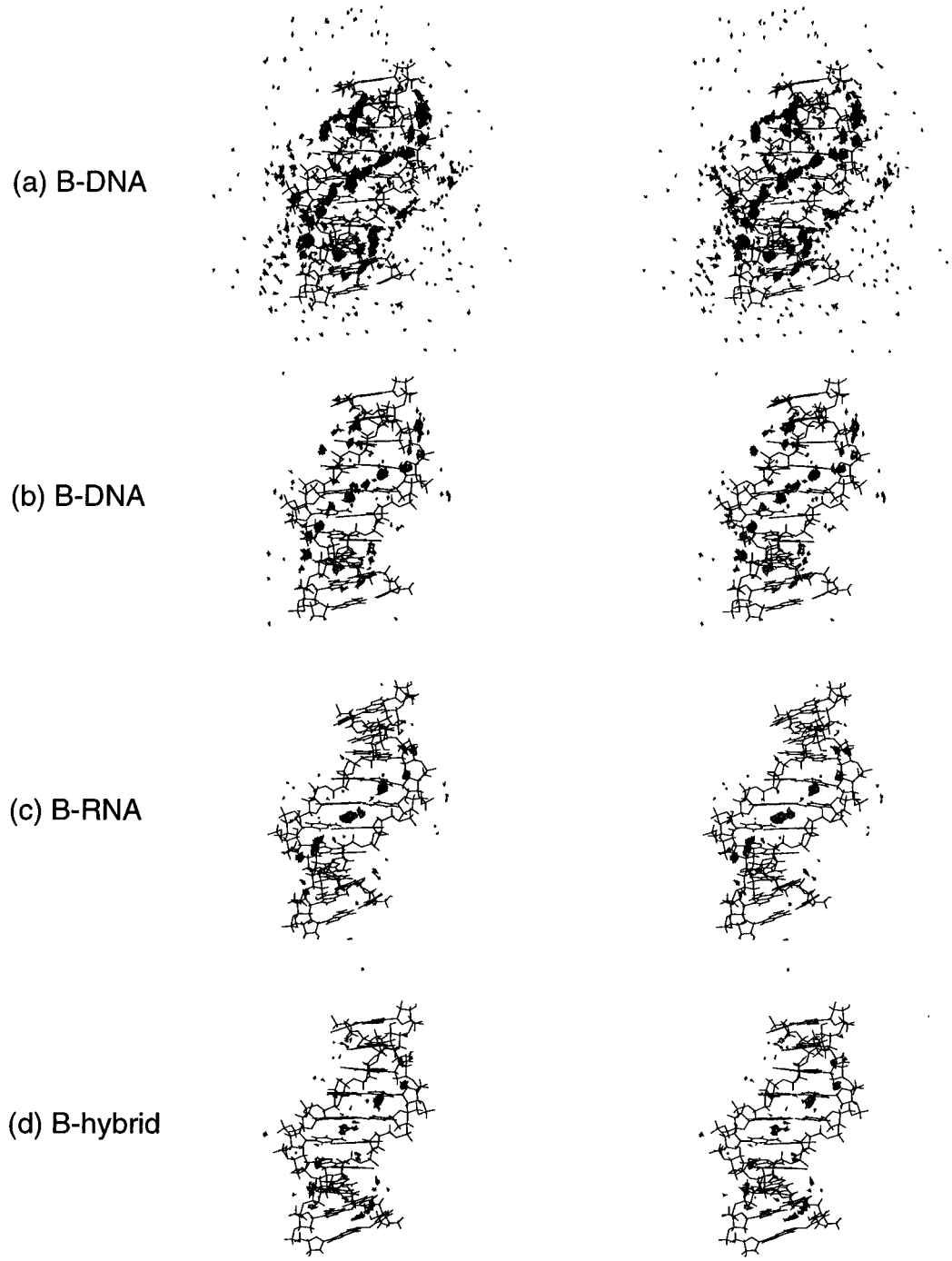

Figure 8. Hydration of the average structures. Stereoview picture of the average structures from various trajectories presented along with contoured water oxygen atom density. The contours of the water oxygen density over 1 ns from each trajectory, at 1-ps intervals, into 0.5 Å^3 grid elements over a 50 Å^3 cubed grid are displayed using the density delegate from MidasPlus. (a) B-DNA average structure over 400−1400 ps at a contour level of 12.0 hits per 0.5 Å^3. (b) B-DNA average structure at a contour level of 15.0 hits per 0.5 Å^3. (c) B-RNA average structure over 1370−2370 ps at a contour level of 15.0 hits per 0.5 Å^3. (d) B-hybrid average structure over 1045−2045 ps at a contour level of 15.0 hits per 0.5 Å^3.

probably directly involved in stabilizing the B-form geometry. This is in agreement with nucleic acid crystal structures of B-DNA which show preferential hydration in the minor groove[83] and, in some cases, little hydration of the major groove.[89] The observation of less hydration in the major groove is also in agreement with NMR studies on B-DNA, which suggest that the water in the major groove is highly mobile and characterized by residence times less than 500 ps.[90] Although the major groove is less hydrated, there is visible major groove hydration. This hydration is most visible in the B-DNA structure (Figure 8a,b) and resembles a "spine" that runs down the middle of the major groove between the bases. This hydration is also apparent in the B-hybrid structure where the major groove density appears darker at the bottom of the helix, where it interacts with 3′ terminal guanines of the DNA strand, than on the top, where it interacts with the 3′ terminal guanines of the RNA strand. Not only is the density lower on the top part of the B-hybrid, but density also appears lower in the B-RNA structure, which suggests that the minor groove of RNA in a B-form geometry is less preferentially hydrated than in B-DNA.

Hydration of A-RNA. The location of bound water is distinctly different in A-form structures than in B-form structures

(89) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S. *J. Mol. Biol.* **1992**, *226*, 1161−1173.
(90) Liepinsh, E.; Otting, G.; Wuthrich, K. *Nucleic Acid Res.* **1992**, *20*, 6549−6553.

as has been seen in the analysis of crystal structures[72,79] and via fiber diffraction.[91] The deeper major groove of A-form structures, and rotation of the phosphate group into the major groove, leads to more well-defined hydration in the major groove. The minor groove is also hydrated; however, the more open minor groove of A-DNA is not as extensively hydrated as B-DNA.[79] A-RNA, on the other hand, has an extensively hydrated minor groove, likely due to the addition of the O2′ hydroxyl groups which provide an anchor point for hydration traversing the minor groove, as is seen in a recent high-resolution A-RNA crystal structure of r[CCCCGGGG]$_2$.[92] The extensive hydration of A-RNA and differences between A-RNA and B-RNA/B-DNA are readily apparent in the simulations of A-RNA. In Figure 9a, a contour plot of the A-RNA hydration, contoured at 12.0 hits per 0.5 Å^3, is shown with a view into both grooves. The extensive hydration of the major and minor grooves is readily apparent, and the overall hydration patterns are very distinct from that seen in the B-form structures (Figure 8).

In contrast to the visible "spine of hydration" in the B-form structures, the minor groove in A-RNA lacks a clear "spine", appears more hydrated, and has 2−3 waters per base pair step interacting with the bases and backbone (except at the terminal residues which appear less hydrated as was seen and discussed in the analysis of the hydration of the B-form structures). The hydration of the A-hybrid structure is similar (data not shown). In the high-resolution A-RNA crystal structure of r[CCCCGGGG]$_2$, two common transversal hydration motifs are seen which link the O2′ hydroxyl's across the minor groove.[92] Most of the major groove in this crystal structure is characterized by two waters between the RNA strands linking the O2′ atoms from adjacent residues in base pair steps across the groove; this motif has the shortest distance between O2′ atoms in different strands. In the average A-RNA structure from the molecular dynamics simulation, the average distance between the O2′ atoms in adjacent base pair steps across the minor groove is slightly larger (8.74 Å) than is observed in the crystal structure (8.52 Å). The other transversal hydration motif was observed at two base pair steps in the crystal structure of r[CCCCGGGG]$_2$ where three waters linking the O2′ atoms across base-paired nucleotides are seen; the distance across the base pair linking the O2′ atoms is 11.30 Å in the crystal, compared to 10.93 Å, on average, in the A-RNA average structure. Figure 9b shows the minor groove hydration of the A-RNA average structure, contoured at a level of 15.0 hits per 0.5 Å^3 (or ∼3.6 times the expected water density). From this figure, both transversal minor groove hydration motifs are simultaneously apparent, such as can be seen at the central CpG step. The behavior at the AU base pairs is slightly different, as might be expected since the AU base pair only has two hydrogen bond acceptors in the minor groove compared to the three present in GC base pairs. At the UpU steps, two waters deep in the groove are seen interacting with the adenine O2 and uracil N3 atoms. This leads to the observation of one and two waters, respectively, in the two O2′ atom transversal hydration motifs discussed above. A third water interacts with the uracil O2′ atom and the water bound to the uracil N3 atom. In addition to the transversal water linking the O2′ atoms within the minor groove, hydration of the O2′ atoms out of minor groove and above the backbone can be seen (as displayed in Figure 9b). This water hydrating the O2′ atom likely interacts with water directly solvating the backbone. Overall, the O2′ atoms appear to be solvated more in A-RNA than was seen in the B-RNA simulation.

The backbone is also fairly extensively hydrated. The relatively short distance between intrastrand phosphates in A-form structures commonly leads to single water bridges between the adjacent O1P atoms.[93,94] In the A′-RNA structure of r[CCCCGGGG]$_2$,[92] this motif is seen despite the slightly longer intrastrand phosphate distance in A′-RNA. In this crystal structure, the bridged water tends to be closer to the O5′ side of the bridge. In the A-RNA average structure, although water oxygen density bridging the O1P atoms is generally observed, the tendency of the water to be closer to the O5′ atoms is not reproduced. In some cases this water density is closer to the O5′ atoms; however, more often the water oxygen density is closer to the O3′ side of the O1P−water−O1P bridge where the water is closer to other donors, such as the O4′ and purine N7 atoms (where present). It is not clear if the simulation is incorrect here. The differences in hydration could relate to differences in crystal versus solution phase structures, differences between A′-RNA and A-RNA structures, and sequence specific hydration patterns or may relate to subtle deficiencies in the simple partial charge model's representation of hydrogen bond directionality. In addition to O1P−water−O1P bridges, water can be seen bridging the O2P atoms as well.

Although the major groove is not as well hydrated as the minor groove, judged by comparing the water oxygen density presented in Figure 9, there is still specific hydration and apparently more hydration, particularly in the center of the duplex, than is seen in the corresponding B-form structures (displayed in Figure 8). In Figure 8c, at the CG base pair (or the fifth base pair down from the top) water can be seen traversing the major groove from the O1P of one strand, across the bases to the O1P of the other strand. Hydration is also present at the GpG step at both ends of the duplex where considerable water oxygen density is present. As is discussed in the next section, this water is hydrating counterions in the GpG "pocket" interacting between the N7 atoms. At the water oxygen densities displayed, we do not observe the regular major groove hydration pattern seen in the crystal structure.[92] The differences in hydration are currently being investigated in our lab in simulations on the r[CCCCGGGG]$_2$ duplex structure.

Overall, the A-RNA average structure seems more specifically hydrated than the B-form structures studied; this is probably since A-RNA is more rigid. The observation of different hydration patterns, not only between A-form and B-form structures but between A-RNA and A-DNA,[79] and sequence dependently within A-RNA, suggests that specific hydration patterns may be important for recognition and distinction among various nucleic acids.

Counterions in the Groove. The phosphates pointing into the major groove in A-RNA lead to a more negative electrostatic potential in the major groove than is seen in B-form structures. Therefore, it is expected that counterions will preferentially move into the major groove rather than the more hydrophobic minor groove of A-RNA. Chemical acylation experiments suggest that the narrow and deep major groove of A-RNA may be inaccessible for specific recognition.[95] Despite this, NMR and crystallography experiments suggest that ions (Ba^{2+} and $[Co(NH_3)_6]^{3+}$) can specifically interact in the major groove, most notably interacting with guanine N7 atoms.[96] This suggests that Na^+ ions may penetrate the deep and narrow major groove of A-RNA; this is seen in our simulations where sodium counter-

(91) Langan, P.; Forsyth, V. T.; Mahendrasingam, A.; Pigram, W. J.; Mason, S. A.; Fuller, W. *J. Biomol. Struct. Dyn.* **1992**, *10*, 489−503.
(92) Egli, M.; Portmann, S.; Usman, N. *Biochemistry* **1996**, *35*, 8489−94.
(93) Saenger, W.; Hunter, W. N.; Kennard, O. *Nature* **1986**, *324*, 385−388.
(94) Westhof, E. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 125−144.
(95) Weeks, K. M.; Crothers, D. M. *Science* **1993**, *261*, 1574−1577.
(96) Gao, Y.-G.; Robinson, H.; van Boom, J. H.; Wang, A. H.-J. *Biophys. J.* **1995**, *69*, 559−568.

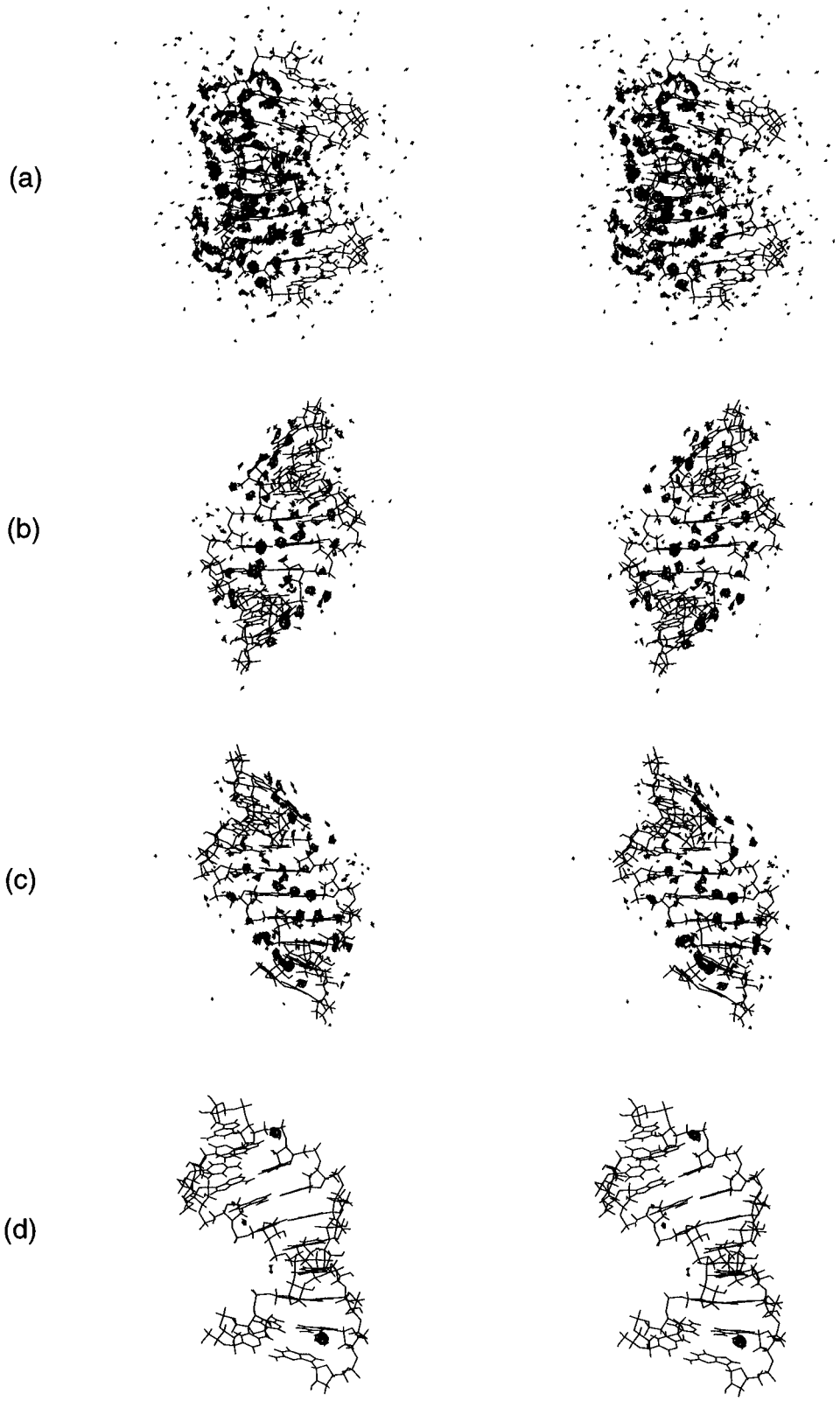

Figure 9. Average structure from the A-RNA trajectory (1030−2030 ps) in various views with contoured water oxygen density (a−c) and counterion density (d) at 1-ps intervals. In part (a), water oxygen density at a 12.0 hits per 0.5 Å^3 contour level is displayed with a view into both grooves. Water oxygen density at a 15.0 hits per 0.5 Å^3 contour level is shown with a view into the minor groove (b) and major groove (c). Sodium counterion density is displayed in part (d) at a contour level of 12.0 hits per 0.5 Å^3.

ions move to distinct locations in the major groove of A-RNA. Symmetrically, at both ends of the duplex, high density is seen for counterions at the GpG steps, closely associated with the N7 atoms of the guanines, as can be seen in Figure 9d which shows the sodium atom density at 12.0 hits per 0.5 Å^3. Although the view is not exactly the same, in Figure 9c as mentioned previously, water is visible at the GpG steps which forms part of the coordination shell around the sodium atoms. Also visible in Figure 9d is more faint density more toward the center of the duplex associated with the adenines at the ApA step and also at the ApC step. This counterion density, at a contouring level equivalent to Figures 8a and 9a, shows that counterions specifically, and preferentially, associate with the major groove of A-RNA. It is not until the contour level is

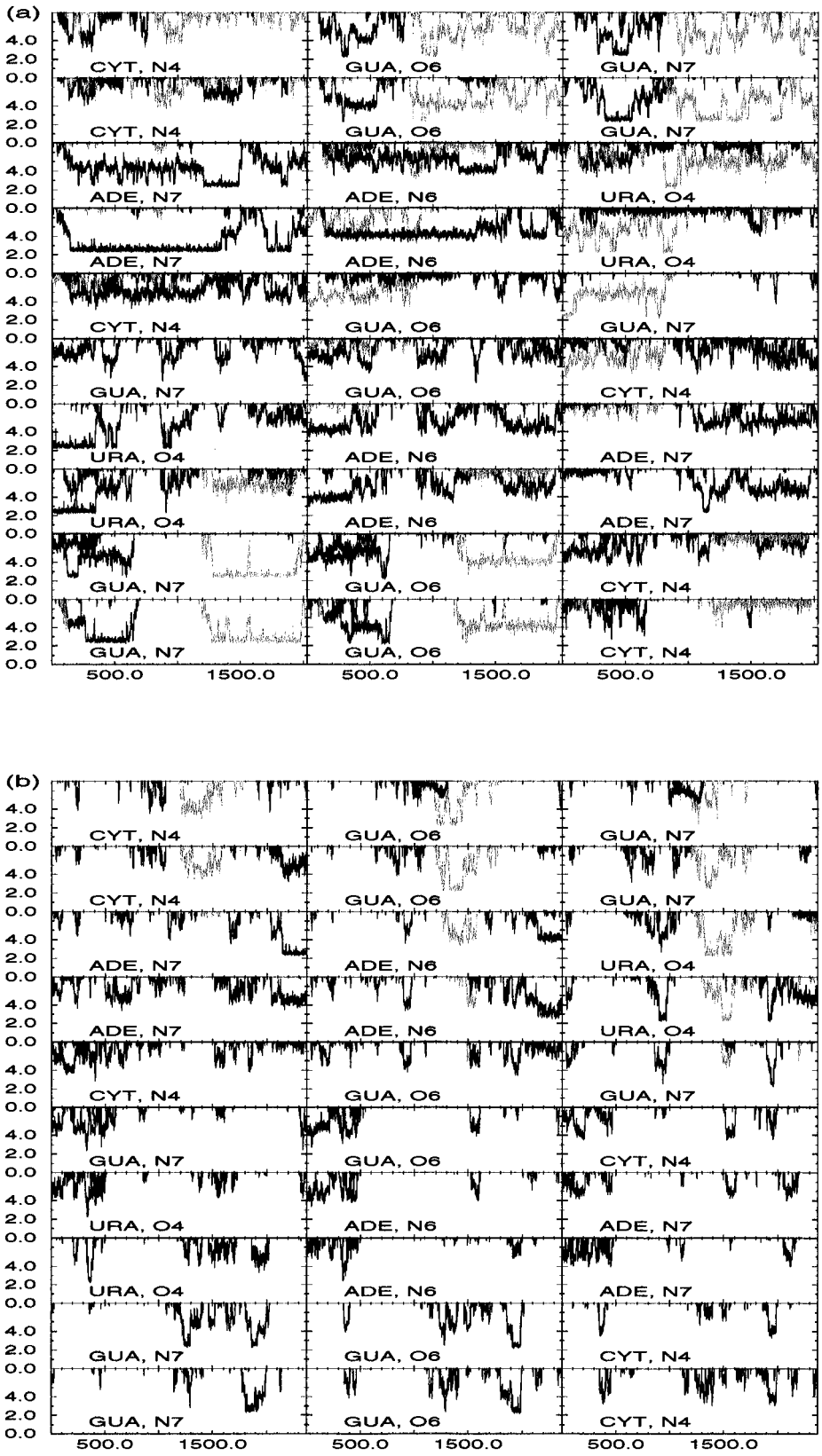

Figure 10. Distances from individual counterions to atoms along the duplex representing the major groove. Each individual graph represents the distances of each counterion to an atom (as labeled in each subgraph) as a function of time. Each base pair is represented from the 1st (top) to the 10th (bottom). The distances, along the *y*-axis, are in angstroms and the time in ps. (a, top) Distances for the A-RNA simulation. (b, bottom) Distances for the B-RNA simulation. To highlight the time course of individual ions, two ions are represented in gray in the top and bottom of part a and one ion is in gray in part b.

lowered to 4.0 hits per 0.5 Å^3 that counterion density is seen in the minor groove of A-RNA and this density is not deep in the groove, but relatively close to the backbones interacting with the minor groove waters and the backbone phosphates and O2′ hydroxyls. The density seen in these simulations for the counterions is not large since only 18 counterions were added to neutralize the system and these counterions diffuse throughout the simulation box during the trajectory. However, it may be claimed that, even though greater than 2 ns of simulation were run for each of the models with RNA and only the final

nanosecond was used in the analysis, the counterion positions could be biased by the initial starting coordinates. In other words, perhaps counterions appear in the major groove of A-RNA since they were started near the major groove. However, we think that this is not the case, in part because not only do the counterions diffuse at near the expected rate ($1.2-2.1 \times 10^{-5}$ cm^2) but the counterions in the major groove exchange during the simulation. In Figure 10 distances between key atoms which represent the major groove and the individual counterions versus time for all the base pairs are shown. Comparing the A-RNA (Figure 10a) to the B-RNA (Figure 10b), counterions are clearly more often closer to atoms in the major groove in the A-RNA simulation than in the B-RNA simulation. The guanine N7 atom distances from the A-RNA simulation (the top right and bottom left of Figure 10a highlight two ions in gray), show that not just one ion interacts with the GpG step at both ends of the duplex, but at least two distinct ions interact at different times. A persistent interaction is also seen with the N7 atoms of the first strand adenines. By following the gray in the middle to top of Figure 10a, a particular ion can be followed moving from interactions with the center of the duplex up through ~1 ns where it moves to interact with the GpG step N7 atoms (at the top of the duplex) over the latter part of the trajectory. The B-RNA simulation clearly shows less specific, close, and persistent interaction of sodium ions and groove atoms. Moreover, tracking an individual ion, such as the gray ion seen in the top of Figure 10b, the interaction is less specific to the GpG step; it moves from interactions with the guanine O6 atoms to interactions with the uracil O4 atoms.

The density of counterions associated with the A-RNA structure at the 12.0 contour level is greater than is seen in any of the other trajectories (B-RNA, B-DNA, B-hybrid, and A-hybrid). In fact, density in the B-RNA simulation does not appear at all until the contour level is dropped down to 8.0 hits per 0.5 Å^3, where a little density appears in the minor groove. As the contour level is lowered in the B-RNA, counterions appear associated in the minor groove, major groove, and along the backbone. The major groove counterion positions are not in the distinct "pockets" seen in the A-RNA simulation where the counterions tend to interact with the N7 atoms at purine−purine steps, but is more diffuse and resembles the inverse spine of hydration seen in the major groove of the B-DNA simulation. Simulations on the dodecamer d[CGCGAATTCGCG]$_2$ with various initial counterion positions[23] suggest that the most favorable "pocket" in the major groove of B-DNA will be intrastrand GpG "pockets" with ions interacting with the guanine O6, followed by ApA "pockets", where the ion would interact with the thymine O4 atoms. In Figure 10b, representing the B-RNA simulation, this general trend is seen. The ions are closest to the guanine O6 or uracil O4 atoms when in GpG or ApA "pockets", respectively. However, since the occupancy of the ions in the major groove of B-RNA is lower than is seen in the A-RNA, not enough occupancy of the major groove by ions was observed by us to unequivocally support the conclusions of Young *et al.*[23] on ions in grooves. However, our B-RNA simulations do support their observation that there is an overall lesser propensity for fractional occupation by mobile counterions in the major groove than the minor groove in B-form structures. The current simulations add to this observation of Young *et al.*[23] by pointing out that the propensity for ions in the grooves of A-form structures is distinct from B-form structures. Although intrastrand GpG and ApA pockets are still favored in the A-RNA major groove (see Figure 10b and Figure 9c), the ions tend to interact more strongly with the guanine or adenine N7 atoms rather than the guanine O6 or thymine/uracil O4 atoms. The general trend seen in these simulations is that the B-form structures tend to favor counterions in the minor groove, whereas the A-form structures favor having counterions in the major groove. Given that the counterions in the major groove of A-RNA tend toward specific locations, rather than the diffuse density seen in the major groove of B-form structures, suggests that the counterions may in part stabilize and rigidify the A-RNA structure. A good test of these hypotheses regarding counterion association would be to run simulations of both A- and B-form structures where a significant number of counterions were initially placed ~10 Å from the solute to see if the same localization of the counterions is observed, similar to the experiments performed by Young *et al.*[23]

Conclusion

These calculations show that molecular dynamics simulations—with a reasonable force field,[18] proper treatment of the long-ranged electrostatics,[10,13,19] and representation of the solvent and counterions—can accurately represent the differences in structure between A-RNA, B-DNA, and DNA:RNA hybrid duplexes. Spontaneously in the simulation of A-RNA, a single (α,γ) crankshaft transition in one strand, along with the observation of low twist and interstrand guanine stacking, occurred at the central CpG step. This feature, seen in a variety of A-RNA[30] and A-DNA[27,69,72] duplex crystal structures, has been thought to result from crystal packing. These results suggest that this feature is a context-dependent, sequence-specific structure that can appear even in the absence of crystal packing or specific hydration patterns. It is clearly a context-dependent effect since this behavior is not seen at other, non-central, pyrimidine−3′-5′-purine steps in these simulations or in crystal structures. The results also suggest that the A-RNA structure does not have the sequence specific narrowing at the center of the duplex as seen in the B-DNA crystal structure,[77] nor the sequence specific bending patterns seen in the crystal (where TpG bend into the minor groove) or in molecular dynamics simulations (where TpG and CpG bends into the major groove).[24] Instead, a generalized roll into the major groove is observed.

The A-hybrid simulations demonstrate that the DNA strand can undergo an A-DNA to B-DNA transition, despite the presence of the A-RNA strand, to converge to a structure that has features very similar to what has been seen by NMR. In particular, the DNA strand has sugar puckers that interconvert between C2′-endo and C3′-endo, while the RNA strand sugars remain in a C3′-endo conformation. Additionally, a minor groove width intermediate between A-RNA and B-DNA is observed along with positive base pair inclination to the helical axis, negative propeller twist, and negative *x*-displacement from the helical axis. Similar to the A-RNA, the A-hybrid structure displays a small positive roll into the major groove; however, the low twist and interstrand guanine stacking at the central CpG step is not observed. Clearly the DNA strand is more deformable, as the properties of the overall duplex are more similar to an A-form geometry than to a B-form geometry, yet the DNA strand still adopts primarily C2′-endo sugar puckers.

Each of the structures, A-RNA, B-DNA, and A-hybrid, have distinct structural features which allow for discrimination and help explain how enzymes can distinguish between the structures.[49,53] In addition to the structural differences, there are also differences in the relative flexibility between A-form and B-form structures, the latter being considerably more flexible. Although the data are not conclusive, it appears the solvent and counterions influence the flexibility of the structures. The rather specific association of water and counterions into the major groove and transversal water bridging the O2′ atoms across the minor groove of A-RNA may stabilize the structure. The

B-form structures on the other hand are characterized by more diffuse counterion association to the backbone and in the grooves and a less specifically stabilizing "spine" of hydration in the minor groove. Overall the more flexible structures show less specific hydration (which may be an artifact of the analysis) and the fractional occupancy of counterions in the major groove of B-RNA is considerably less than that seen in A-RNA.

The convergence to the "same" B-form geometry when simulations are started from canonical A-DNA or B-DNA, canonical B-RNA, or a canonical B-form hybrid duplex is somewhat surprising. It was previously thought that the unacceptable stereochemistry of the O2′ hydroxyl "bumping" into the following phosphate group, sugar ring, and base in RNA would destabilize the B-form geometry and make B-RNA unfavorable. While the hydroxyl does point up toward the following nucleotide (as can be seen in Figure 2b), the interaction is not unfavorable; in fact, the hydroxyl group hydrogen bonds with one of the phosphate oxygens, the O5′ of the backbone, or both from the following residue. This stable interaction of the O2′ hydroxyl, coupled with the increased barrier to sugar repuckering,[76] helps explain why B-RNA is a stable conformation during more than 2 ns of simulation and why it is difficult to force the B-RNA to A-RNA transition. Based on the current data and since we have not observed a spontaneous B- to A-RNA transition, it is impossible to determine which structure is more stable. Simulations are currently underway in an attempt to directly calculate the relative free energies of the A-RNA and B-RNA models and the free energy barrier to interconversion. However, the observation that C3′-endo to C2′-endo repuckering occurs less frequently than the reverse during 2 ns of simulation on the RNA duplexes, coupled with the observation of the sugar pucker transition to C3′-endo in the terminal cytosine of the RNA strand from the B-hybrid and the large fluctuations, indirectly suggests that A-RNA may be the more stable, consistent with observation.

The observation of stable B-RNA in the RNA and hybrid duplexes suggests that conformational sampling of RNA is clearly an issue. Not only do the structures get locked in B-RNA or A-RNA conformations, but the backbone can get locked into conformations that are persistent for longer than a nanosecond time scale (*i.e.* the ApC step in the simulation with the concerted flip in puckers). Little repuckering occurs, no B_I to B_{II} backbone transitions are observed in A-RNA simulations and few in B-DNA, and few (α,γ) crankshaft transitions are observed in over 4 ns + simulation of RNA duplexes. Moreover, short simulations may not be sufficient to observe backbone transitions, such as the (α,γ) crankshaft seen after ~1 ns of simulation in only one strand of the A-RNA simulation. As one of the anonymous reviewer's of this manuscript remarked, "the observation of 'B-RNA' is a 'shock'; the results suggest that molecular dynamics simulations are totally unable to distinguish a 'real' structure (*i.e.* coming from X-ray or NMR data) from an 'unreal' non-structure."

The rigidity of the RNA clearly presents difficulties to modellers of RNA. Limits in computer power and complexity of the calculations currently restrict simulations to a nanosecond time scale which is not long enough to allow sampling between the various conformational states. Clearly simulations on RNA could benefit from application of some reasonable means to enhance conformational sampling, such as locally enhanced sampling.[97] With the current methods, it is likely that a given RNA model, such as a tRNA crystal or model structure, will remain close to the initial model during a simulation, whatever the validity of the starting structure, for many nanoseconds. This apparent stability of the structure does not validate the force field, per se, since the RNA may be caught in a metastable state and therefore does not give any meaningful information if the structure is "unreal". This has been observed in nanosecond time scale simulations of an RNA hairpin.[98] These results suggest that fairly long simulations, *i.e.* tens of nanoseconds, may be necessary to investigate RNA structures via unrestrained molecular dynamics. Since the barriers to conformational transition in DNA are clearly smaller and transitions such as the A-DNA to B-DNA transition can readily be observed,[24] simulations in the 1−2 ns time scale may be sufficient to properly represent right-handed DNA duplexes. Despite these caveats with respect to conformational sampling, nanosecond time scale simulations seem to be able to provide useful insights into the overall sequence specific structure and dynamics of nucleic acids.

Acknowledgment. P.A.K. is grateful to acknowledge research support from the NIH through grant CA-25644. T.E.C. would like to acknowledge research support as a NIH Biotechnology Training Grant (GM08388) and UCSF Chancellor's Graduate Research Fellow. We would also like to acknowledge the Pittsburgh Supercomputing Center (PSC) (MCA93S017P) and Silicon Graphics Incorporated for significant computational resources; the UCSF Computer Graphics Laboratory (RR-1081); Michael Crowley of PSC for parallelizing the PME code on the Cray T3D; Tom Darden for helpful discussion and for making the PME code available; Jim Vincent and Ken Merz for the original MPI version of AMBER; Matt Young and David Beveridge as well as Yong Duan and John Rosenberg for making preprints of their work available; Jed Pitera and Randy Radmer for helpful discussions about the density grid generation and contouring; and Jennifer Miller for helpful discussions.

JA963641W

(97) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277−9286.
(98) Miller, J. L.; Kollman, P. A. *J. Mol. Biol.* In press.