JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Molecular Dynamics of RNA with the OPLS Force Field. Aqueous Simulation of a Hairpin Containing a Tetranucleotide Loop

Dominic A. Zichi

Contribution from NeXagen, Inc., 2860 Wilderness Pl., Boulder, Colorado 80301

Received September 19, 1994[®]

Abstract: The results are reported for a 200 ps, unconstrained simulation of the RNA hairpin molecule GGGCGCAAGCCU in aqueous solution with charge-neutralizing counterions. The OPLS force field is used to compute the RNA energies and forces. Periodic, truncated-octahedral boundary conditions coupled with Ewald summations for computing long-range electrostatic interactions are employed. The accuracy of the computational methodology is assessed by a detailed comparison of the structural features derived from the simulation to those obtained from NMR data. The analysis of the simulation shows overall good agreement with experiment, including NMR-derived interproton distances and ribose—phosphate backbone torsion angles. The secondary and overall three-dimensional structure is well preserved throughout the 200 ps trajectory. An interesting structural feature seen in the simulation is the formation of a water-mediated hydrogen bond between the unusual G—A base pair in the loop. Such an interaction may add to the observed stability of this hairpin molecule. Analysis of the cross-correlation of atomic displacements indicates that the loop and stem behave dynamically as distinct structural units.

I. Introduction

Our understanding of the functional roles of ribonucleic acid (RNA) molecules has been changing dramatically over the past several years. We now know that RNA molecules can perform a diverse set of biological functions, including transmission of genetic information, catalysis of phosphodiester bond formation and cleavage, and the regulation of transcription, splicing, and translation.[1] Recently, a procedure for *in vitro* selection called SELEX (Systematic Evolution of Ligands by EXponential enrichment) has been used to isolate RNA molecules from vast repertoire libraries for specific activities, most notably high-affinity binding to protein targets.[2] These diverse *in vivo* and

in vitro functions are undoubtedly related to the structure of the RNA molecule. For example, the specificity and activity of enzymatic RNA catalysts result from the three-dimensional disposition of functional groups within the folded oligonucleotide. Similarly, the enormous diversity of structural shapes present in random sequence RNA libraries used in SELEX is key to finding high-affinity binding to a wide range of molecular targets, including small-molecule targets.[3]

As a result of these expanding discoveries of the roles played by RNA molecules, recent interest in RNA structure has increased dramatically. Many biologically important, single-stranded RNA macromolecules have previously been identified, such as transfer RNA (tRNA) and messenger RNA, as well as complex assemblies of RNA and protein molecules, such as spliceosomes and ribosomes. Single-stranded RNA molecules are known to form stable, compact structures over a wide range

® Abstract published in *Advance ACS Abstracts*, March 1, 1995.

(1) Chastain, M.; Tinoco, I., Jr. *Prog. Nucleic Acid Res. Mol. Biol.* **1991**, *41*, 131–177.

(2) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505–510. Schneider, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1992**, *228*, 862–869. Tuerk, C.; MacDougal, S.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6988–6992.

(3) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425–1429.
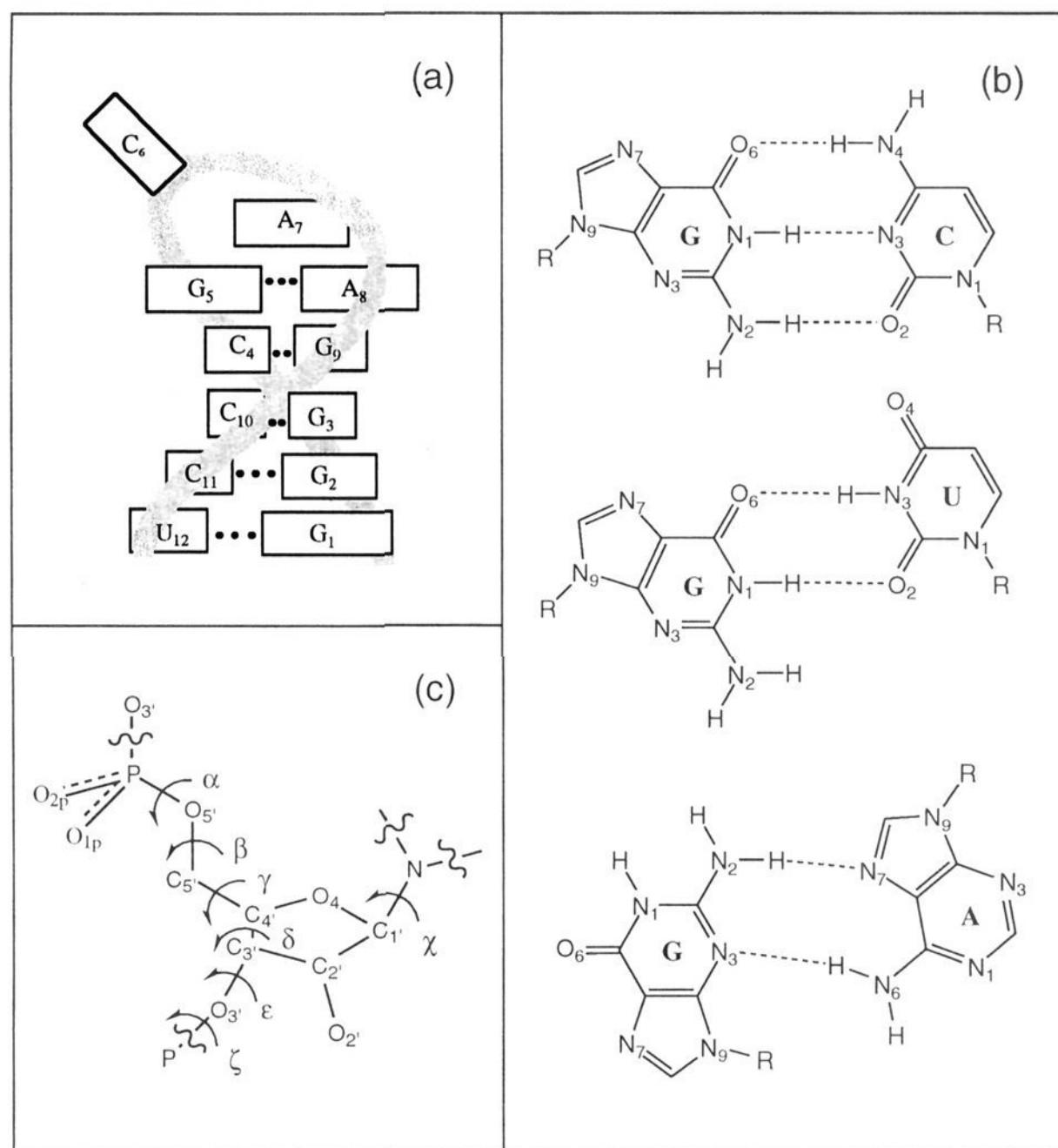
Figure 1. (a) Schematic representation of the GCAA hairpin structure determined by NMR. The full sequence of the hairpin is GGGCGCAAGCCU. The gray ribbon represents the phosphodiester backbone, and the boxes represent nucleotide bases. Hydrogen bond interactions are indicated by filled circles, and stacking interactions are implied from the schematic. (b) Schematic representation of hydrogen bonding between the base pairs formed in the hairpin, G—C, G—U, and G—A. (c) Torsion angle identities for a single base.

of sequence length. At one extreme are the ribosomal RNAs (rRNAs), among the largest known single-stranded RNA molecules, consisting of several thousand nucleotides. Group I and group II catalytic introns are molecules of intermediate size, comprising several hundred bases. Somewhat shorter in sequence length are tRNA molecules, typically containing 75 nucleotides. tRNA is the largest RNA molecule for which high-resolution structural data exists. Finally, at the other extreme, relatively short sequences of less than 25 nucleotides can form stable secondary and tertiary structures in solution. This allows us to study in isolation small key functional elements which compose larger tertiary structures. For example, a hairpin loop is a functional motif known to bind proteins and form tertiary interactions in larger RNA molecules.[1] The topic of this work is a computer simulation of a solvated RNA hairpin molecule.

RNA hairpins are one of the most common structural motifs in single-stranded RNA molecules, including rRNAs, ribozymes, and tRNAs. These hairpin structures consist of a double-stranded stem and a single-stranded loop region. The thermo-dynamic stability of hairpin structures depends strongly on the sequence and size of the loop, with certain tetranucleotide loops forming unusually stable molecules. Based on phylogenetic analysis, two distinct families of hairpins with tetranucleotide loops have been identified, namely, the GNRA and UNCG loop

sequences.[4] The interactions responsible for stabilizing these two distinct loop structures have been determined by NMR solution studies.[5,6] Common three-dimensional features between these two sequences include unusual interactions among loop bases and interactions between loop bases and the sugar—phosphate backbone.

In the present work, molecular dynamics (MD) simulation techniques are used to probe the structural and dynamical features of an RNA hairpin molecule. Computer simulations of biological molecules can augment experimental data by providing atomic level detail for the systems being studied.[7,8] Such calculations have the potential to refine and, in the most favorable case, aid in predicting the three-dimensional structure of molecules. It is also possible to extract certain thermody-namic energies from these calculations. Since the reliability of these types of computer experiments rests heavily on the

(4) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; D'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I., Jr.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1364—1368. Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467—8471.

(5) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680—682.

(6) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191—194.

(7) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(8) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; John Wiley and Sons: New York, 1988.

simulation models and methods employed, the primary purpose of this MD simulation is to investigate our ability to accurately simulate RNA molecules in solution. The fidelity of this computer experiment was assessed by comparing the simulation results to experimental NMR studies.

In order to compute accurate properties of aqueous solutions from MD simulations of biological molecules, it is essential that a detailed, realistic model be defined for the system of interest. This model comprises a number of features, including the constituent atoms, their interaction potentials or force fields, the nature of the boundary conditions, and the treatment of long-range electrostatic interactions. These aspects of the model are interrelated and can greatly influence the quality of the results obtained. Since the system of interest here is a polyelectrolyte molecule in a polar, ionic environment, careful account of solvent, counterions, and electrostatic interactions is essential. Therefore, an explicit water solvent model and neutralizing counterions are employed. The long-range electrostatic interactions are assessed via the well-known Ewald summation technique,[9] resulting in essentially no truncation of the electrostatic interactions. The Ewald method has been well-established for simulating many particle Coulombic systems, e.g., solid and molten salts,[10,11] ionic solutions,[12] molecular systems,[13] and, more recently, small peptides[14,15] and nucleic acids[16] in aqueous solution. In addition to allowing for a more faithful account of the long-range screening effects of the ionic environment, there are several other advantages for using the Ewald method here. It avoids artificial forces associated with switching functions and nonconservation of energy when straight truncation is used.[14] Also, it avoids the necessity of imposing spatially inconsistent residue- or group-based cutoffs in order to prevent "splitting dipoles" during energy evaluation.

The NMR solution conformation of an RNA hairpin containing a GCAA tetranucleotide loop serves as the main structural reference point for the computer simulation results presented below. The NMR study was done by Heus and Pardi and is the first structure to be determined for RNA hairpins containing the GNRA loop sequence.[6] The key features of the GCAA hairpin structure are illustrated in Figure 1 and are summarized below.

(1) All nucleotides in the structure have anti-glycosidic bond conformation (χ).

(2) There is an unusual $G_5 - A_8$ base pair between the first and last bases in the loop, see Figure 1b. Two interbase hydrogen bonds are reported, one from $G_5:N_2-H$ to $A_8:N_7$, and the other from $A_8:N_6-H$ to $G_5:N_3$. The second hydrogen bond is not actually observed in the ensemble of NMR-refined structures; most $H\cdot\cdot N_3$ distances are greater than 2.9 Å. An additional hydrogen bond from $A_8:N_6-H$ to $G_5:O_{2'}$ is observed.

(3) A hydrogen bond between $G_5:N_2-H$ and an A_7pA_8 phosphate oxygen is observed.

(4) Extensive stacking occurs in the loop; G_5 stacks on C_4, A_8 stacks on G_9, and A_7 stacks on A_8. C_6 is not stacked; rather, it is disordered.

(5) C_6 and A_7 have mostly 2′ endo sugar puckers; the remaining residues have 3′ endo ribose conformations.

(9) Ewald, P. *Ann. Phys. (Leipzig)* **1921**, *64*, 253–287.

(10) Cowley, E. R.; Jacucci, G.; Klein, M. L.; McDonald, I. R. *Phys. Rev. B* **1976**, *14*, 1758–1769.

(11) Adams, D. J.; McDonald, I. R. *J. Phys. C* **1974**, *7*, 2761–2775.
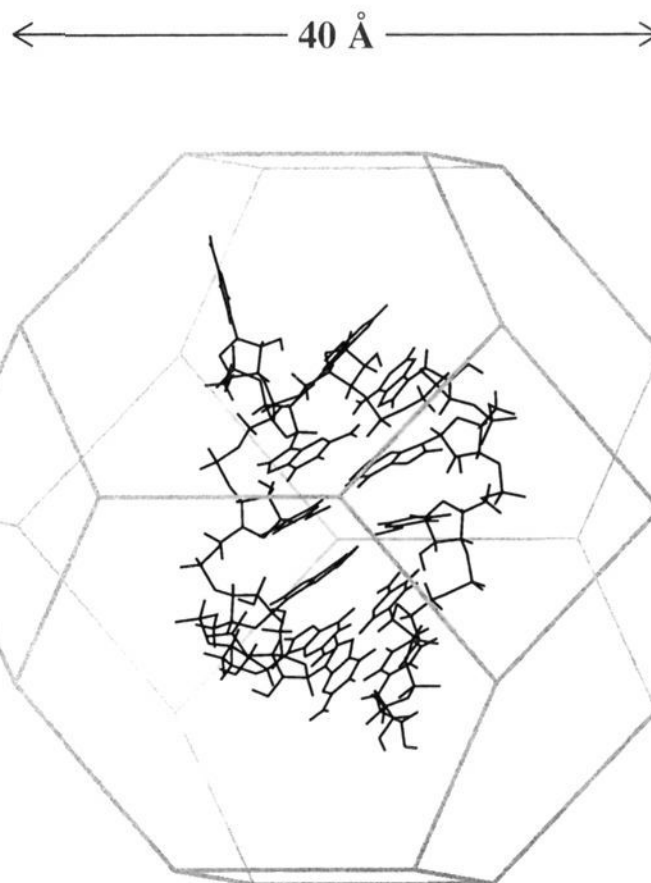
(12) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071–5083.

(13) Nosé, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055–1076.

(14) Smith, P. E.; Pettitt, B. M. *J. Chem. Phys.* **1991**, *95*, 8430–8441.

(15) Schreiber, H.; Steinhauser, O. *Biochemistry* **1992**, *31*, 5856–5860.

(16) Laaksonen, A.; Nilsson, L. G.; Jönsson, B.; Teleman, O. *Chem. Phys.* **1989**, *129*, 175–183. Forester, T. R.; McDonald, I. R. *Mol. Phys.* **1991**, *72*, 643–660.

Figure 2. Unit cell geometry with the initial configuration of the RNA molecule.

(6) The residues in the helical stem are of standard A-form geometry.

These features, in addition to NOE distances and torsion angles derived from NMR data, have been monitored during the course of the simulation and are reported below. Also included in the analyses are some static and dynamic features of the hairpin molecule obtained from the simulation.

The organization of this paper is as follows. In the next section a detailed description of the computational methodology used here is presented. The results are presented and discussed in section III. First, a detailed comparison of the simulated structural results with those obtained from the NMR study are presented. This is followed by a discussion of the static and dynamic features of the simulated molecule, including atomic motions and their cross-correlations. Conclusions are presented in section IV. A detailed account of the solvent and ion structure, energetics, and dynamics will be the subject of a future publication.

II. Methods

Model System. The primary simulation cell is composed of an all-atom model of the RNA molecule of sequence GGGCGCAAGCCU with 11 sodium counterions to neutralize the solute charge, immersed in 897 water molecules within a truncated octahedron, see Figure 2. Thus there were a total of 3091 atoms in the primary simulation cell. The unit cell containing the truncated octahedron is 40 Å in length, giving a system density of 1.0 g/cm³. The resulting ionic solution has a concentration of ~50 mM RNA and ~500 mM sodium ion giving a Debye length of ~4.0 Å, compared to the experimental system of 2 mM RNA and 120 mM NaCl with a Debye length of 8.0 Å. In order to reduce the ionic strength of the simulated system to compare more favorably with experiment, roughly an order of magnitude more water molecules would be needed in an 85 Å containing cube. Such a large simulation was deemed impractical for the present purposes. Care in properly accounting for the long-range electrostatic interactions is, however, crucial (see below).

The rigid water model SPC/E,[17] which has a computed dielectric constant of 71 at 298 K,[18] is used for evaluating solvent energies and forces. The sodium ion force field parameters are adopted from Åqvist.[19] OPLS atomic charges and Lennard-Jones parameters[20] together with Amber angle bending and torsion rotation parameters[21] were employed for evaluating the energies and forces of the RNA molecule. All covalent bonds in the solute are constrained to their equilibrium values specified in the Amber force field. The OPLS nonbonded parameters for nucleotide bases reproduce experimental and *ab initio* intermolecular hydrogen bond energies and geometries with no additional hydrogen bond functions.[20] Since one of the main interests of the present work is to evaluate the fidelity of the oligonucleotide force field, no restraints are placed on the Watson–Crick base pairs in order to preserve the observed secondary structure, although this is frequently done in DNA simulations.[22]

Computational Methodology. Periodic boundary conditions are applied to the primary simulation cell, a truncated octahedron.[23] This cell shape was chosen for the periodic system for several reasons: (i) it is most efficient for solvating nearly spherical solutes with a minimum number of solvent molecules, (ii) it reduces angular distortions associated with cubic boundary conditions with Ewald sums, and (iii) it allows for a more efficient calculation of the Ewald sum.[24] The images of the central cell are obtained from the primitive translations of a body-centered cubic lattice. No image atom coordinates, however, are actually generated during the simulation; rather, the method of Adams is used which computes the interatomic vector and distance implicitly for the nearest neighbor atoms of a pair.[24] The reciprocal-lattice vectors of this periodic system, required in the Ewald sum (see below), are obtained from the primitive translation vectors of the face-centered cubic lattice.[25]

The simulation is computed in the NVT ensemble at 298 K using the extended system method of Nosé.[26] The Verlet algorithm is used to integrate the equations of motion.[27] Since the Hamiltonian for the NVT ensemble contains a velocity-dependent term and the velocities at time t are not determined until the position at time $t + \delta t$ is known, the equations of motion are computed in a self-consistent manner, using an initial estimate for the velocities. Typically, three iterations lead to satisfactory convergence in both the positions and velocities. It is important to note that the interatomic forces are computed only once per time step. However, the constraint forces do need to be re-evaluated for each iteration within a time step. Therefore, an efficient bond constraint algorithm is desired. All bond constraints, both solvent and solute, are maintained by use of the SHAKE algorithm.[28] A fast matrix inversion method for imposing the water molecule bond constraints, which account for ~90% of the constraints here, was implemented and resulted in a 100-fold speedup compared to that of the standard iterative SHAKE procedure. Overall, evaluation of the constraint forces is computationally negligible and allowed for a time step of 2.0 fs to be used.

The electrostatic energies and forces are efficiently computed with the Ewald method as described by Adams and McDonald[11] and scale as the number of interaction sites, N, rather than as N^2. The Ewald method evaluates the electrostatic interactions of the periodic system by computing the sum of two rapidly converging series, one in Cartesian space (r-space) and the other in reciprocal-lattice space (k-space). The convergence behavior in both summations depends critically on the adjustable parameter α (in the notation of Adams and McDonald), whereas the net sum of these two independently convergent sums is independent of α. Preliminary calculations were performed to optimize the convergence of the Ewald sum. Satisfactory convergence is obtained for this system using an α of $9.5/L$, where L is the length of the unit cell, truncating the r-space sum at interatomic separations of 10.5 Å and including 688 reciprocal-lattice vectors in the k-space sum. It is important to realize that, even though a truncation is used for evaluating both the r-space and the k-space sums, the long-range electrostatic interactions are not being truncated; rather, the accuracy of the calculation is being affected. For the current simulation the relative error in the magnitude of the electrostatic energy is less than 0.01% using the prescription outlined above.

Since the Ewald method allows for the use of an r-space interaction cutoff, the same cutoff is used to truncate the Lennard-Jones interactions with no appreciable loss of accuracy. This also allows the use of an atom-based neighbor list to minimize the time spent in evaluating the interatomic interactions.[29] No artificial groupings of atoms are needed for cutoff evaluation in order to avoid "splitting dipoles", since no interactions are neglected. A separate neighbor-list cutoff greater than the r-space cutoff was used. The neighbor list was automatically updated whenever the cumulative sum of any two atoms' individual displacements exceeded the difference in the neighbor-list cutoff and the r-space interaction cutoff.[29] This conservative scheme ensures that no atom pair interaction is neglected between neighbor-list updates. Preliminary runs were used to optimize the neighbor-list cutoff. Longer cutoffs result in less frequent updating, but add overhead at each step. A neighbor-list cutoff of 12.5 Å, resulting in neighbor-list regeneration every 12 steps, on the average, was found to be most optimal, although a 0.5 Å increase or decrease does not seem to greatly influence the timing.

A computer program to perform the MD simulation outlined above was written in FORTRAN-77.[30] The CPU time required for one MD time step averaged 35.0 s on a single-processor SGI Iris 4D/340 VGX for this system of 3091 interaction sites. The force evaluation accounts for almost 88% of the total CPU time, with the r-space and the k-space sums accounting for 80% and 20% of this time, respectively. The remaining 12% is split between neighbor-list generation, 9%, and integration and constraint evaluation, 3%. Atomic coordinates from the simulation were saved at 0.01 ps intervals for computing the structural and dynamic properties that are presented and discussed in the following sections.

The starting coordinates for the RNA molecule were obtained from an NMR structure refined with NOE-restrained MD simulations.[33] The initial configuration was generated by centering the RNA molecule in the truncated octahedron and overlaying a 40 Å box of water obtained from an equilibrated configuration of the neat solvent. Water molecules outside the truncated octahedron or within 2.5 Å separation of any solute atom were removed. The sodium ions were introduced by changing 11 water molecules into counterions. That water molecule closest to

(17) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(18) Reddy, M. R.; Berkowitz, M. *Chem. Phys. Lett.* **1988**, *155*, 173–176.

(19) Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.

(20) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.

(21) Wiener, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(22) Swaminathan, S.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5027–5040.

(23) Adams, D. J. *Chem. Phys. Lett.* **1979**, *62*, 329–332.

(24) Adams, D. J. In *The Problem of Long-Range Forces in the Computer Simulation of Condensed Media*; Ceperley, D. M., Ed.; NRCC Workshop Proceedings: Berkeley, 1980; Vol. 9, p 13.

(25) Kittel, C. *Introduction to Solid State Physics*; John Wiley and Sons: New York, 1953; pp 312–313.

(26) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 298–307.

(27) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

(28) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(29) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987; pp 146–152.

(30) The entire simulation and analysis software was written by the author to assure internal consistency. Extensive testing and validation was performed on all aspects of the code during development. For example, key tests on the internal energy and force routines were done by comparison to both the CHARMM[31] and DISCOVER[32] simulation packages. The Ewald sum energies were tested by comparison to explicit lattice sums for some representative configurations while the forces were verified by numerical differentiation. The Amber force field implementation was verified with comparisons to DISCOVER. Total energy was conserved for a variety of systems during lengthy runs, including vacuum and solution. Finally, a number of structural and dynamical properties for liquid water were computed from a long simulation and compared to published results to further validate the code.

(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(32) BIOSYM Technologies, 10065 Barnes Canyon Road, Suite A, San Diego, CA 92121.

(33) Pardi, A. Personal communication, 1993.

time (ps)
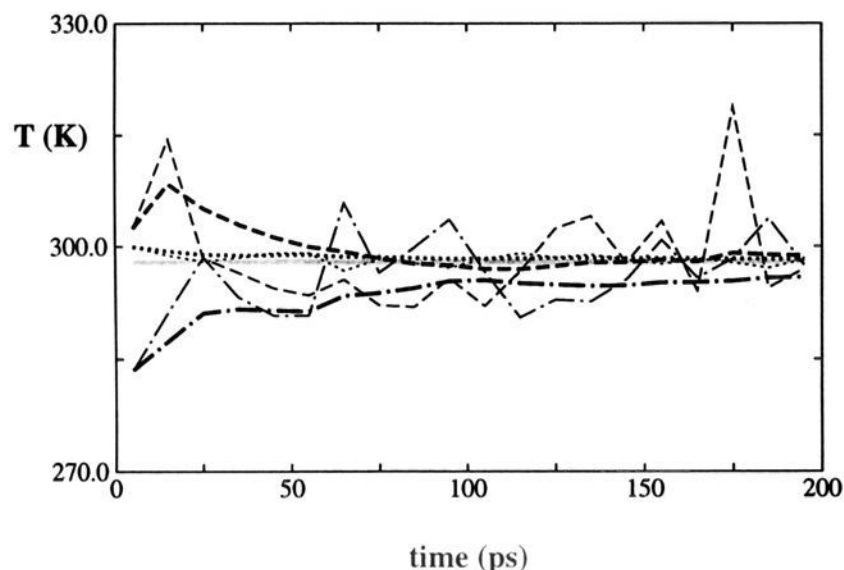
Figure 3. Temperature variation of the simulation components vs time. The thin lines are 10 ps averages, and the thick lines are running averages for the temperature of the total system (gray lines), RNA molecule (dot−dash lines), Na$^+$ ions (dashed lines), and water molecules (dotted lines).

the $O_{1p}-P-O_{2p}$ angle bisector and in the first solvent shell of the phosphate anion was selected. The oligonucleotide was held rigid, and the solvent and counterions were minimized with 100 steps of steepest descent, followed by 20 ps of dynamics for equilibration. During this initial equilibration stage, the velocities of the moving atoms were randomized every 5 ps by assignment from a Maxwell−Boltzmann distribution in velocities at 298 K. The resulting coordinates plus random initial velocities were used to initiate the trajectory of the subsequent 200 ps unrestrained MD run.

III. Results and Discussion

In the following, the structural features derived from the NMR study are first compared with those computed from the simulation. Further analysis is presented for some static and dynamic features of the molecule not directly observable in the NMR experiment, including atomic root mean square (rms) displacements and their cross-correlations and intramolecular hydrogen-bond geometry distributions.

Convergence Properties. The energy of the extended system Hamiltonian (H) is a rigorously conserved quantity in this simulation.[26] For the present calculation, the normalized rms energy fluctuations, $(\langle \delta H^2 \rangle / \langle H \rangle^2)^{1/2}$, $\delta H = H - \langle H \rangle$, are on the order of 0.008%, indicating very good energy conservation for the numerical integration. The overall system temperature is computed to be 298.0 K. Good thermal equilibration for each component of the solution was obtained, with average temperatures for the solute, ions, and solvent of 295.9, 298.8, and 298.3 K computed over the entire 200 ps trajectory. Temperature averages over 10 ps windows were computed for these components from their kinetic energies, which undergo fluctuations in the canonical ensemble. These results, along with their running averages, are displayed in Figure 3. The magnitude of the kinetic energy fluctuations for the ions are largest, followed by that of solute, and then the solvent. This trend is due primarily to sampling size.

A. Comparison to NMR Data. Interproton Distances and Torsion Angles. NMR studies on nucleic acid molecules can provide information on interproton distances and backbone torsion angles for the oligonucleotide. Interproton separations are obtained from 2-D nuclear Overhauser effect (NOE) spectra, resulting from a transfer of magnetization by a through-space dipolar coupling between nuclear spins.[34] The intensity of the observed NOE is proportional to r^{-6}, where r is the internuclear separation of the coupled nuclei. Torsion angle data are derived

from homo- and heteronuclear 2-D correlation spectroscopy, resulting from through-bond electron-coupled spin−spin interactions.[34] The magnitude of the spin−spin coupling is dependent on the dihedral angle ϕ between the two nuclei. Computed values for $\langle r^{-6} \rangle^{-1/6}$ and $\langle \phi \rangle$ have been obtained from the MD trajectory for comparison to the NMR experiments. The results were calculated as 20 ps block averages over the 200 ps trajectory and are displayed in Figures 4 and 5.

The interproton distances computed from the simulation generally lie within the ranges determined from the NOE data. The calculated distances do not change appreciably during the 200 ps trajectory, indicating that the simulation structure is quite stable in this time frame. The interproton distances computed from 0−20 ps are seen to be essentially the same as those derived from the 180−200 ps time period, see Figure 4. An outlier can be defined as a calculated distance range which lies outside of the NMR range. For this purpose the calculated range is defined to be ±1 standard deviation from the mean interproton distance. With this definition, less than 13% of the calculated distances are outside of the NMR measurements. The main outliers are the result of a fraying terminal G_1-U_{12} base pair and a ribose pucker for C_{11} which is inconsistent with an A-form helix. The fraying G_1-U_{12} pair is reflected by the two outliers in the 3.0−5.0 Å range and the two in the 2.0−4.0 Å range. These outlier distances are somewhat longer than those observed experimentally. The four calculated distances which lie above the 1.8−3.0 Å NMR range are due to a slight stacking rearrangement for residues C_{10}, C_{11}, and U_{12}. This is a combination of the terminal base pair fraying and of residue C_{11} having a $C_{2'}$ endo pucker, whereas the NMR data indicates a $C_{3'}$ endo form. The initial RNA structure had the $C_{2'}$ endo conformer, and no interconversion of sugar puckers was observed throughout the simulation, see below. This ribose conformation also leads to several computed distances being shorter than experiment in the 3.2−3.4 Å range.

Overall good agreement between the MD and NMR results is also seen in the torsion angle comparison displayed in Figure 5 (see Figure 1c for torsion angle definitions). The major torsion angle violations are seen for the G_9 backbone. Initially, the γ torsion angle is $+sc$ while β is $-sc$, both in accord with NMR measurements.[35] During the first 20 ps, β rotates to $-ap$, while γ rotates to $+ap$ and back to $+sc$. The preferred orientation of β is $+ap$, and it is difficult to rationalize the observed backbone angle of 284° in terms of favorable interactions. Indeed, the rearrangement in the simulation leaves both anionic phosphate oxygen atoms fully exposed to solvent, whereas the initial NMR conformation has one oxygen partially desolvated by its proximity to the G_9 base. The only other torsion angle which shows significant deviation from experiment is the terminal γ_1 torsion angle, which tends to extend $O_{5'}$ out into solution. It is important to note that these backbone rearrangements, detailed below, result in essentially no qualitative change in the overall folded state of the oligonucleotide.

RMS Deviations. Another means of comparing the simulated structure to that obtained from NMR is to examine the average rms deviation of the non-hydrogen atoms in the NMR structure to those along the MD trajectory. Although NMR refinement typically results in an ensemble of structures, that single representative NMR structure used here to generate the initial configuration will be the focus for comparison. In order to remove overall rotation and translation of the solute from the rms calculation, each MD structure was overlaid on the

(34) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(35) The Klyne−Prelog nomenclature[38] for torsion angle ranges is used throughout, with the following abbreviations: (0° ±30°) ± synperiplanar (±*sp*); (±30°, ±90°) ± synclinal (±*sc*); (±90°, ±150°) ± anticlinical (±*ac*); (±150°, ±180°) ± antiperiplanar (±*ap*).
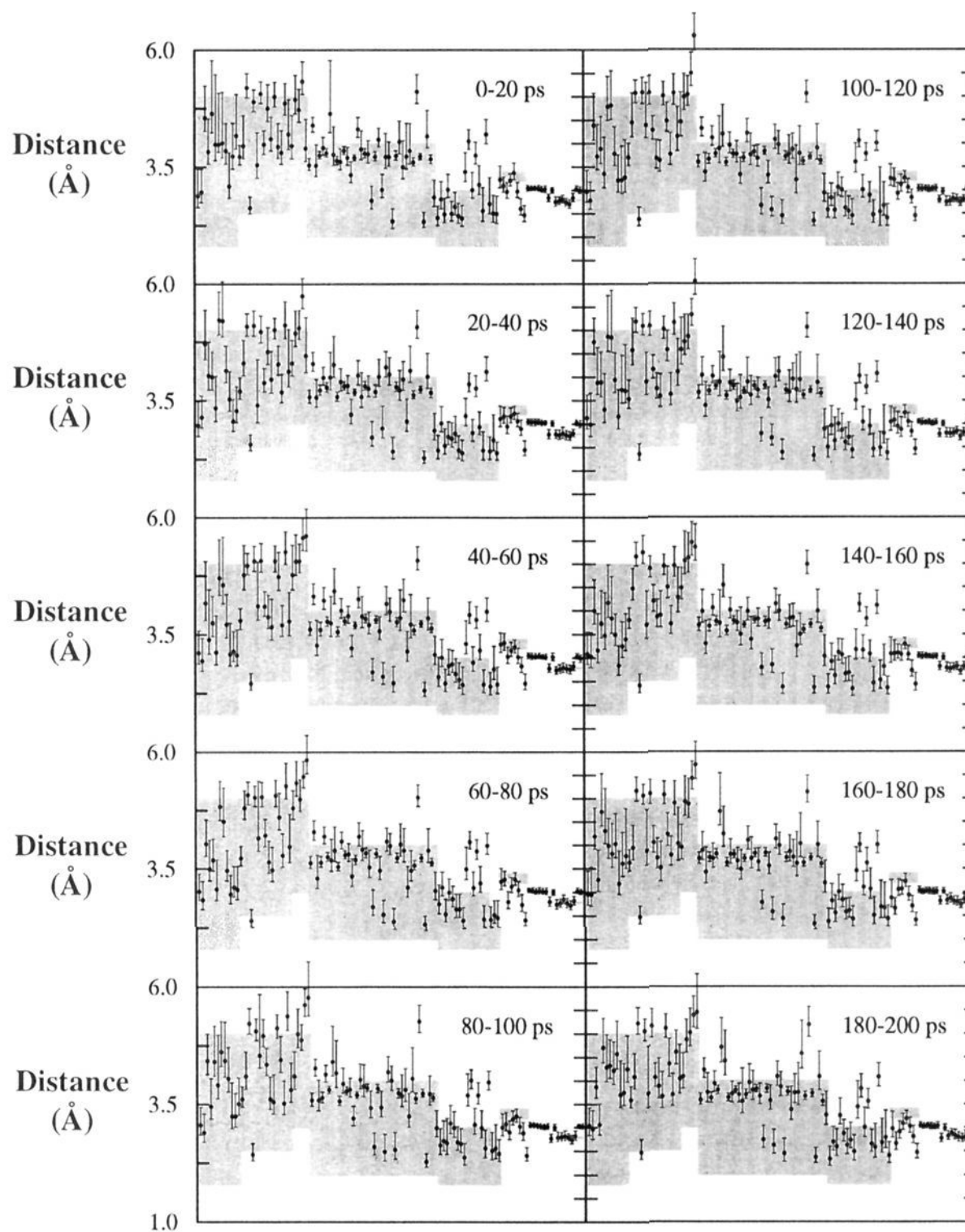
Figure 4. Calculated interproton distances (filled circles) vs NMR-derived ranges (gray boxes) for 20 ps time periods. A total of 111 individual distances are plotted along the *x*-axis with a vertical gray box indicating the NMR-derived distance range and a filled circle at the computed interproton distance, $\langle r^{-6} \rangle^{-1/6}$, for each NMR measurement. The vertical bars are a measure of the range of distances derived from the simulation for the proton pairs computed from $\langle r \rangle \pm \langle \delta r^2 \rangle^{1/2}$, $\delta r^2 = \langle r^2 \rangle - \langle r \rangle^2$.

original conformer by least-squares superposition of the ribose—phosphate backbone atoms.[36] A plot of the rms deviation as a function of time is shown in Figure 6. A very rapid equilibration to a set of states where the rms difference fluctuates around 0.94 Å for the backbone atoms and 1.0 Å for the base atoms is seen to occur between 0−50 ps. Following this, a distinct rearrangement in the molecule leads to a set of conformational states characterized by 1.3 and 1.6 Å rms differences of backbone and base atoms which persist for the remainder of the simulation, from 50−200 ps. As will be seen, the jump in rms deviation from 0.94 to 1.3 Å occurs after several backbone torsion angle transitions. The rms deviations seen here for the simulated structure relative to the NMR structure are comparable to those reported for protein simulations relative to experimental structures.[7,8] Such rms differences are also similar in magnitude to those computed for the structures obtained through NMR refinement.

It is of interest to compare the individual atomic deviations of the dynamical structure from the initial NMR structure in more detail. Figure 7 shows a plot of atomic rms deviations averaged over the entire 200 ps trajectory. Several trends are

apparent from the data. There is an overall base line of about 1.0 Å rms difference throughout the structure. Roughly 0.6 Å of the base line is due to instantaneous thermal fluctuations from the mean atomic positions, see below. The nucleotide base atoms for the terminal residues G_1 and U_{12} and the loop residues 5−8 show larger than average deviations from the NMR structure, with loop bases C_6 and A_8 exhibiting the largest deviations. The NMR data is not able to resolve the position of C_6, most likely because this nucleotide base is not well localized with respect to the rest of the molecule. The central stem bases, consisting of residues 2−4 and 9−11, show the smallest differences from the NMR conformation with rms differences <1.0 Å. The backbone atoms between A_8 and G_9 and to a lesser extent, those between G_2, G_3, and G_4, show evidence of some rearrangement.

In order to understand the trends seen in the rms difference of the simulated structure relative to the NMR structure, the instantaneous radius of gyration (R_G) for each principal inertial axis has been computed.[37] These axes are illustrated in Figure 8. The *c* axis lies along the center of the hairpin, between the

(36) Kabsch, W. *Acta Crystallogr.* **1976**, *A32*, 922−923.

(37) Flygare, W. H. *Molecular Structure and Dynamics*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1978; pp 43−44.
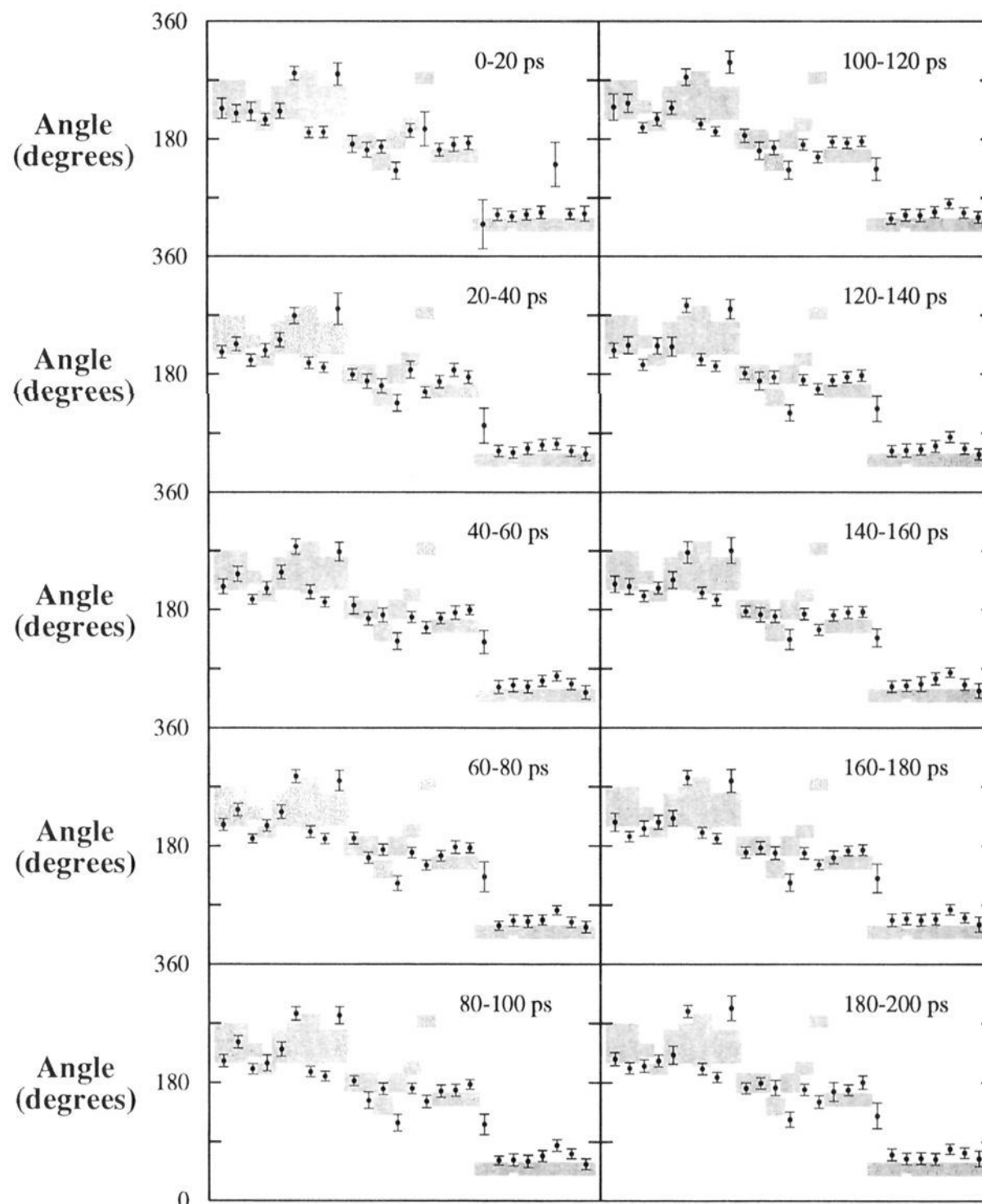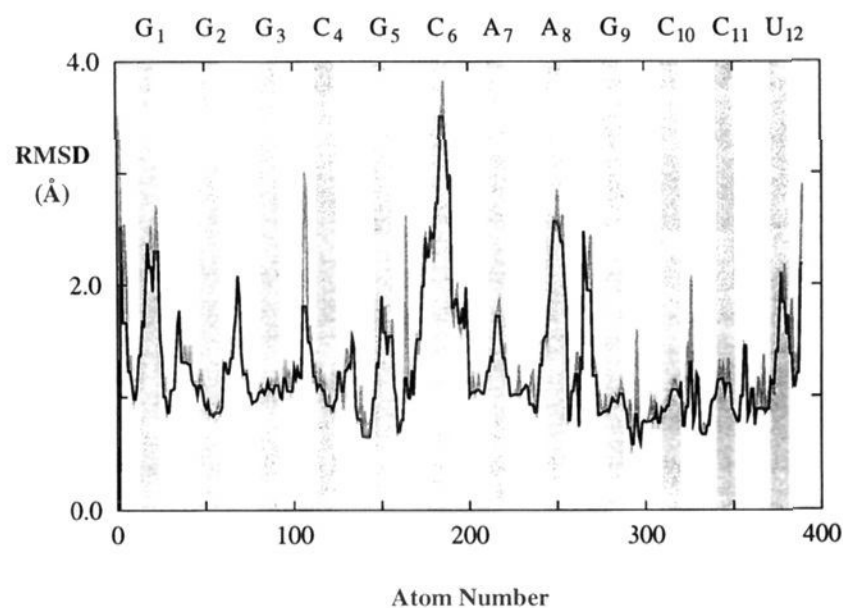
Figure 5. Calculated backbone torsion angles (filled circles) vs NMR-derived ranges (gray boxes) for 20 ps time periods. The vertical bars are computed from the simulation by $\langle\phi\rangle\pm\langle\delta\phi^2\rangle^{1/2}$, $\delta\phi^2 = \langle\phi^2\rangle - \langle\phi\rangle^2$. The first nine torsion angles correspond to ϵ 1-7, 11, 12; the next nine are β 2, 3, 6-12; the final eight are γ 1-3, 6, 7, 9, 11, 12.
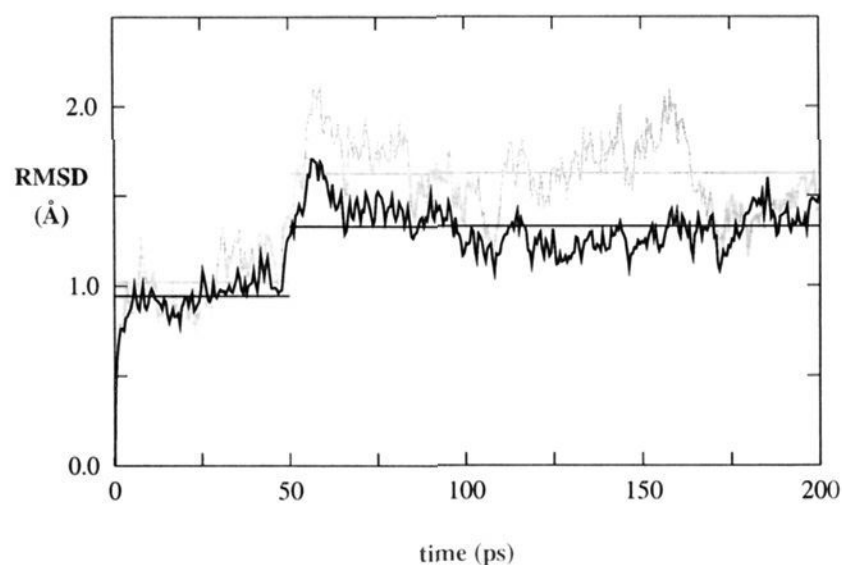

Figure 6. Calculated rms deviation (rmsd) of the simulated structure from the initial NMR structure for non-hydrogen ribose—phosphate backbone atoms (black line) and nucleotide base atoms (gray line) vs time. The straight lines indicate the overall average rms deviations from 0—50 ps and 50—200 ps for backbone (black) and base atoms (gray).


Figure 7. Individual atomic rms deviation (rmsd) of non-hydrogen atoms (black line) and hydrogen atoms (gray line) for the simulated structure compared to the initial NMR structure averaged over the 200 ps trajectory. The light gray shaded regions correspond to those atoms within the nucleotide bases labeled above the graph.

hydrogen-bonded base pairs. This roughly corresponds to the helical z axis of the short stem. The b axis lies approximately along the vector connecting hydrogen-bonded base pairs while the a axis is directed outward from this vector. The time course

of the R_G's, presented in Figure 9, closely reflects the features seen in the rms deviations. Initially, the R_G's for all three axes increase in the first few ps and plateau for ~25 ps. The R_G for
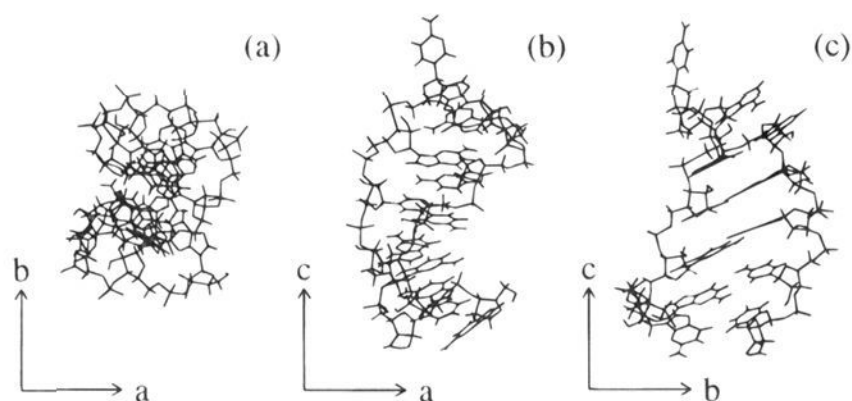
Figure 8. Three orthogonal views along the principal axis system of the RNA molecule relative to its center of mass. The molecule, corresponding to the NMR structure, is viewed in the (a) $a-b$ plane, (b) $a-c$ plane, and (c) $b-c$ planes.

time (ps)
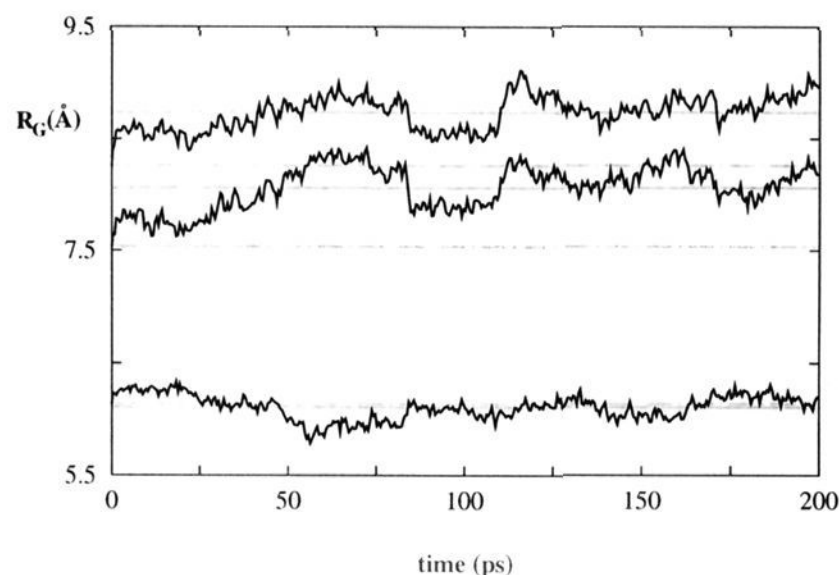
Figure 9. Radius of gyration, R_G, along the trajectory for the a axis (lower curve), b axis (center curve), and c axis (upper curve). The initial value of R_G for each axis is displayed (gray dot−dash line) as well as the average values computed from the trajectory (gray solid line).

the a axis fluctuates close to the initial value of 6.14 Å with a mean of 6.09 Å. However, R_G's for the b and c axes increase ~0.5 Å from their initial values of 7.53 and 8.26 Å during the next 50 ps and fluctuate about their new average values of 8.04 and 8.71 Å for the remainder of the simulation. This slight expansion of the hairpin along its two largest inertial axes accounts for the jump in rms deviation seen above. It also partly explains the larger atomic rms differences for the terminal and loop bases compared to the central stem, since these two ends of the molecule are most affected by the slight expansion in the b and c inertial axes.

It is interesting to note that the oscillations in R_G along the b and c axes appear to be quite well correlated, both expanding and contracting in tandem. A wagging motion of the solvent-exposed loop residue C_6 is consistent with this behavior. Motion which brings C_6 closer to A_7 can lead to a contraction of the molecule along the b and c axes, while motion of C_6 away from the molecule can have the opposite effect along these directions. Similarly, the fraying in the terminal G_1-U_{12} base pair can result in mass being shifted in concert along the b and c directions. The correlation in motion observed for the R_G's along the b and c axes, with a computed correlation coefficient of 0.84, is contrasted by a distinct anti-correlation with the breathing motion along a, which is seen to contract (expand) while b and c expand (contract). Viewing the molecule in the $a-c$ plane in Figure 8b, it is evident that a bending motion about the b axis, essentially bending about the major groove, would lead to such an anti-correlation in R_G for the a and c axes, evidenced by a correlation coefficient of −0.26. It is not clear what leads to such an effect for the a and b axes, with a large correlation

coefficient of −0.65, although the wagging of C_6 and fraying of the G_1-U_{12} base pair could contribute to the observed behavior.

Average Secondary and Three-Dimensional Structure. The overall secondary and three-dimensional structure of the molecule is well preserved throughout the entire simulation. This can most easily be seen by examining time-averaged structures over the course of the 200 ps trajectory. Figure 10 displays five structures viewed in the $b-c$ principal axes plane (Figure 8c), which were generated by averaging the atomic coordinates over consecutive 20 ps time blocks. Included in the figure is an overlay of the ten 20 ps average structures spanning the trajectory onto the original set of NMR coordinates. This sequence of structures clearly demonstrates the preservation of the original three-dimensional structure. The 11 stacked bases remain stacked, and the C_6 loop residue remains extended in solution. The base of residue C_6 is seen to wag, mostly along the b direction. The hydrogen-bonded base pairs are all well preserved; however, the terminal G_1-U_{12} pair does exhibit some fraying. The fraying of the terminal base pair and the loop residue wagging, which account for the slight increase in R_G for the b and c axes, are graphically illustrated in the overlay of Figure 10.

A more quantitative assessment of the secondary structure is found by analyzing the intramolecular hydrogen bonding of the oligonucleotide, see Figure 1b. A hydrogen bond is defined here strictly geometrically; a hydrogen bond exists if the Donor− Acceptor separation (R_{DA}) is less than 3.5 Å and the angle between the Donor−Hydrogen−Acceptor (θ_{DHA}) is greater than 120°. The distributions of R_{DA} and θ_{DHA} were computed for all intact hydrogen bonds throughout the 200 ps simulation and are presented in Figure 11. The observed frequency a particular bond is made, $\langle f \rangle$, the average values of $\langle R_{DA} \rangle$, $\langle \theta_{DHA} \rangle$, and their widths ($\langle \delta x^2 \rangle^{1/2}$, $\delta x = x - \langle x \rangle$, $x = R_{DA}$ and θ_{DHA}) are summarized in Table 1. On the basis of these hydrogen bond results, it is clear that the secondary structure of the hairpin stem is well maintained throughout the simulation. The three Watson−Crick G−C base pairs remain fully hydrogen bonded with sharply peaked distributions at 2.8 Å for R_{DA} and 170° for θ_{DHA}, consistent with hydrogen bond geometries observed experimentally.[38] The effects of an in-plane rocking motion of the G−C base pairs[39] is manifest in the hydrogen bond geometries by broader distributions which peak at shorter lengths and less linear angles for the outer, amino−carbonyl oxygen hydrogen bonds compared to the inner, imino−nitrogen hydrogen bond, see Table 1. The terminal G_1-U_{12} "wobble" base pair shows significant fraying of the $G_1:N_1$-H to $U_{12}:O_2$ hydrogen bond, which is intact only 37.5% of the time, whereas the $G_1:O_6$ to $U_{12}:N_3$-H hydrogen bond is made 91.5%, preserving the integrity of this base pair.

The unusual G_5-A_8 base pair is stabilized by three hydrogen bonds during the course of the simulation. Two of these hydrogen bonds, the $G_5:N_2$-H to $A_8:N_7$ and $G_5:N_1$-H to a phosphate oxygen of A_7pA_8, were identified in the NMR study. In the simulation, these bonds have less ideal geometries than the hydrogen bonds for the Watson−Crick pairs, with average bond lengths of 3.1 and 2.9 Å and average angles of 148.5° and 148.0°. These hydrogen bonds are intact in 85% and 78% of the configurations. A third hydrogen bond between $G_5:N_2$-H and the same phosphate oxygen of A_7pA_8 as above develops during the first 10 ps of the simulation and remains thereafter. This hydrogen bond is found in 97% of the trajectory configura-

(38) Saenger, W. *Pinciples of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.

(39) Harvey, S. C.; Prabhakaran, M.; Mao, B.; McCammon, J. A. *Science* **1984**, *223*, 1189−1191.

0-20 ps

40-60 ps

80-100 ps

120-140 ps

160-180 ps
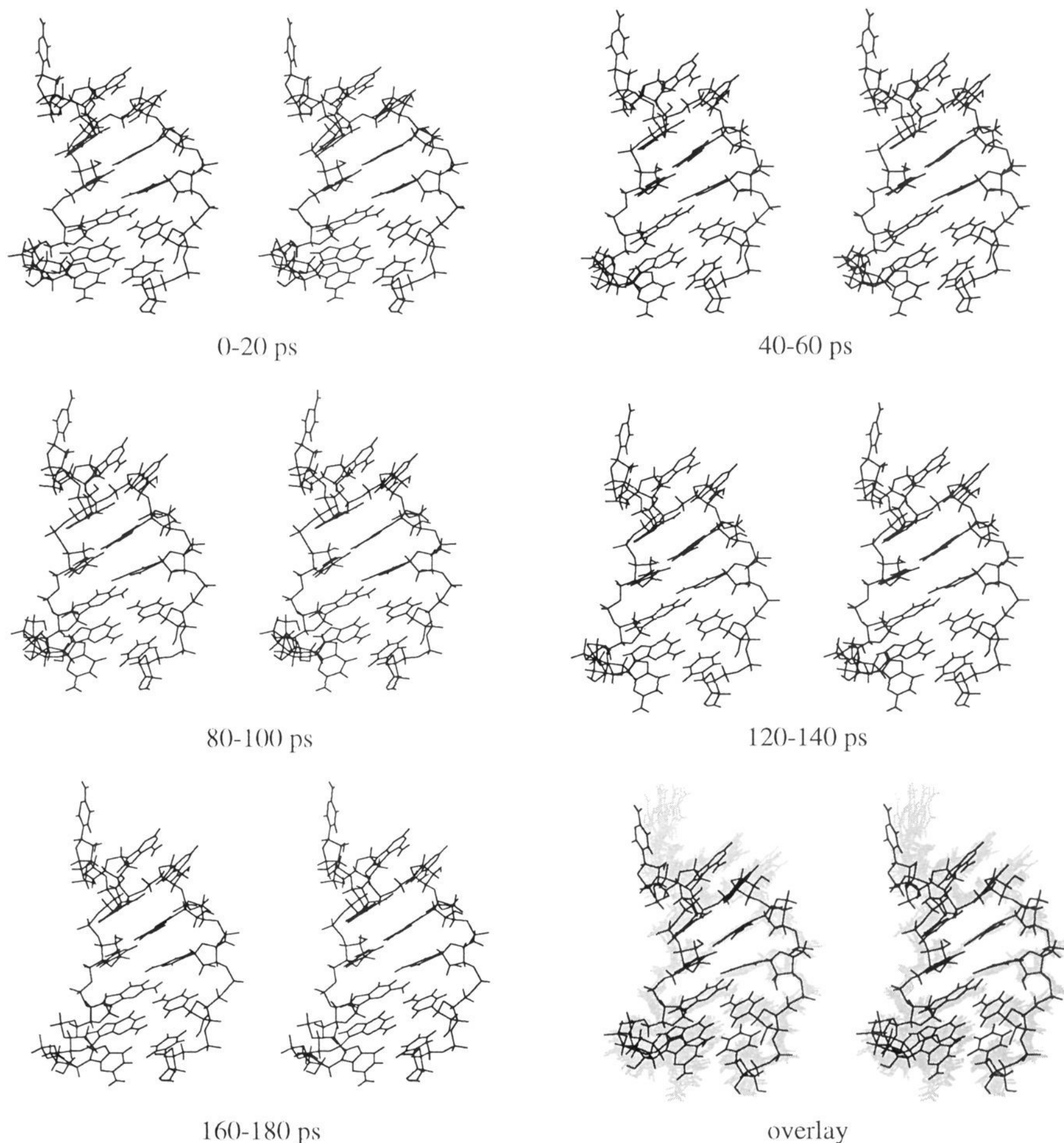
overlay

Figure 10. Stereoviews of five mean structures computed by averaging over 20 ps time periods centered on 10, 50, 90, 130, and 170 ps, along with an overlay of all ten 20 ps average structures (gray lines) spanning the entire simulation with the initial NMR conformation (black lines).

tions. The geometry of this interaction, not reported in the NMR study, is similar to the other phosphate−base hydrogen bond with an average length of 2.9 Å and angle of 150.5°. In addition to these three hydrogen bonds, a base−ribose hydrogen bond is seen to form intermittently between $A_7:N_6$−H and $G_5:O_{2'}$. This interaction is made 27% of time here and was postulated to exist in the NMR study, although the initial conformation did not contain this hydrogen bond by the present geometric definition.

A second hydrogen bond between loop residues G_5 and A_8 identified in the NMR study, $A_8:N_6$−H to $G_5:N_3$, was never directly observed during the simulation. Rather, a water-mediated hydrogen bond develops between these two groups to further stabilize the noncanonical G−A base pair. In order to accommodate this bridging water molecule, the base A_8

reorients by pivoting around the existing hydrogen bond and opening the N_6−N_3 separation from an initial distance of 4.2 Å to a range of 5.5−6.0 Å. A time history of the N_6−N_3 separation is presented in Figure 12 along with the distance history of the bridging water to atoms N_6 and N_3. This rearrangement, which develops over the first 60 ps, is reflected by the large rms difference of 2.5 Å for base A_8 in the simulated molecule relative to the initial NMR structure, see Figure 7. Subsequent to the base pair opening, a water molecule, which had been interacting with N_6 intermittently, rapidly diffuses into a bridging position at 75 ps. This single water molecule remains proximate to these two residues for the duration of the simulation. Once the bridge is established, the frequency of this single water molecule forming a hydrogen bond with each residue *simultaneously* is found to be 65%. In fact, this water
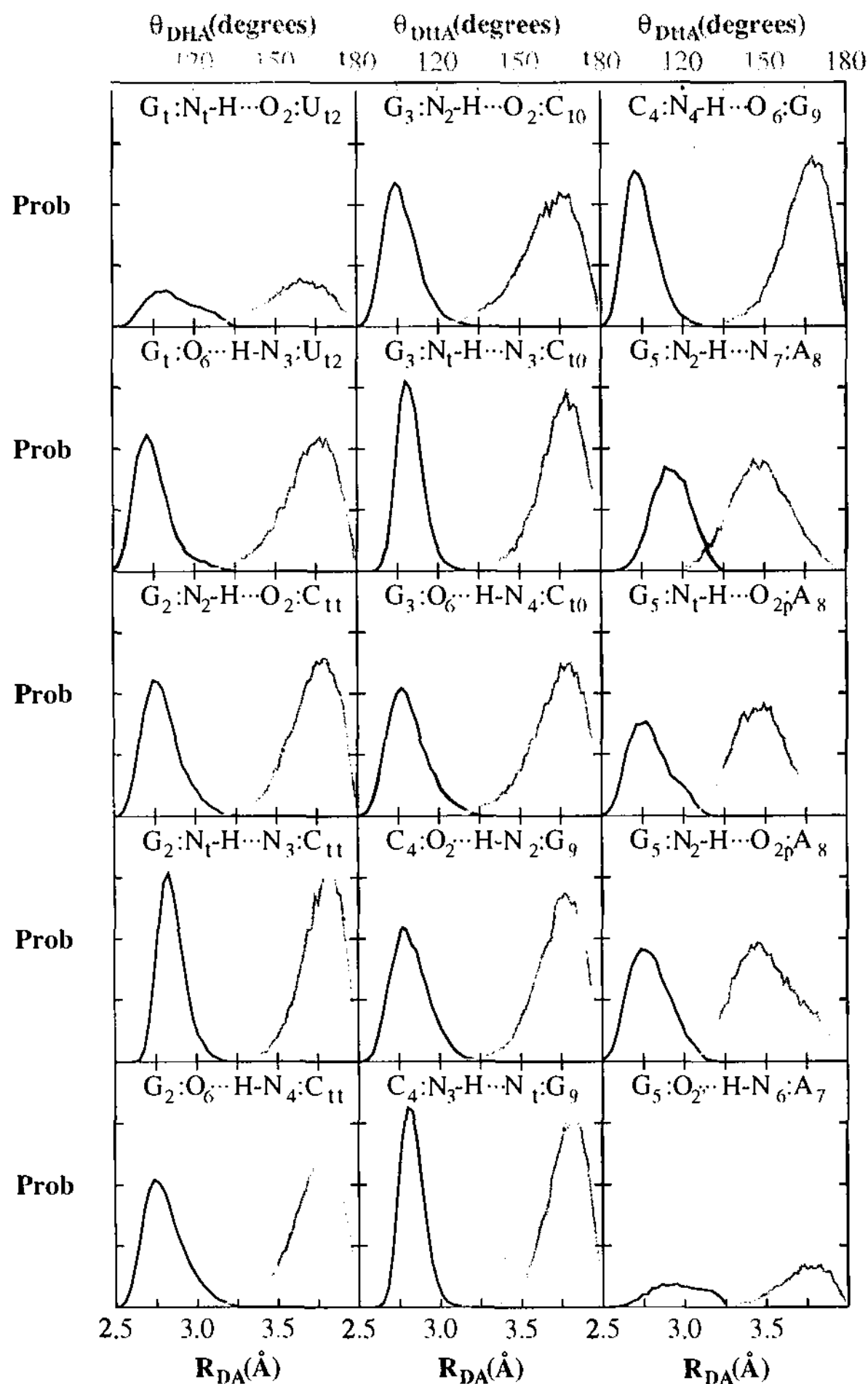
Figure 11. RNA intramolecular hydrogen-bond geometry distributions for bond lengths R_{DA} (black curve) and bond angles θ_{DHA} (gray curve). The probability distributions have been normalized to $\langle f \rangle$, the frequency with which a particular bond is found in the 200 ps trajectory.

molecule remains hydrogen bonded to one of the two residues for almost all subsequent times. Occasionally, two water molecules were observed to form the bridge between G_5 and A_8, leading at one point to an exchange in the identity of the bridging molecule. This exchange, occurring near 120 ps, was quickly reversed, however, with the original bridging water re-establishing the interbase interaction. It is interesting to note that a similar water-mediated base pair interaction was observed for a C—U mismatch in the X-ray crystal structure of a double-stranded RNA duplex.[40]

B. Dynamical Features. In the previous section it was shown that the RNA structure derived from the unconstrained

simulation agrees quite well with the secondary and three-dimensional structural features of the hairpin molecule that were derived from NMR studies. In this section some features of the simulation which are not amenable to direct observation via NMR spectroscopy are explored. First, the time evolution of the torsion angles of the molecule are presented followed by an analysis of mean atomic displacements and their cross-correlations.

Torsion Angle Evolution. The 200 ps time histories of the seven torsion angles (see Figure 1c) and the pseudorotation phase angle, P, are displayed in Figure 13 as polar diagrams with the angle at the initial time plotted in the center and the final value on the outer radius. Consistent with the NMR data, all nucleotides have anti-glycosidic bond conformations, i.e., χ

(40) Holbrook, S. R.; Cheong, C.; Tinoco, I., Jr.; Kim, S.-H. *Nature* **1991**, *353*, 579–581.

Table 1. Intramolecular Hydrogen Bond Geometry Summary

| hydrogen bond	$\langle f \rangle^a$	$\langle R_{DA} \rangle$ (Å)	$\langle \delta R_{DA}^2 \rangle^{1/2}$ (Å)	$\langle \theta_{DHA} \rangle$ (deg)	$\langle \delta \theta_{DHA}^2 \rangle^{1/2}$ (deg)
$G_1:N_1-H\cdots O_2U_{12}$	0.375	3.00	0.20	157.4	11.3
$G_1:O_6\cdots H-N_3:U_{12}$	0.915	2.83	0.15	161.6	10.2
$G_2:N_2-H\cdots O_2:C_{11}$	0.995	2.90	0.15	163.3	9.1
$G_2:N_1-H\cdots N_3:C_{11}$	0.999	2.96	0.11	165.9	7.6
$G_2:O_6\cdots H-N_4:C_{11}$	0.988	2.90	0.16	163.2	9.5
$G_3:N_2-H\cdots O_2:C_{10}$	0.984	2.87	0.14	160.2	11.0
$G_3:N_1-H\cdots N_3:C_{10}$	0.999	2.94	0.11	164.5	8.5
$G_3:O_6\cdots H-N_4:C_{10}$	0.981	2.92	0.16	162.6	10.0
$C_4:O_2\cdots H-N_2:G_9$	0.994	2.93	0.15	163.7	8.7
$C_4:N_3\cdots H-N_1:G_9$	1.000	2.94	0.10	166.0	7.5
$C_4:N_4-H\cdots O_6:G_9$	1.000	2.84	0.13	164.2	8.8
$G_5:N_2-H\cdots N_7:A_8$	0.851	3.08	0.15	148.5	10.8
$G_5:N_1-H\cdots O_{2p}:A_8$	0.780	2.90	0.16	148.0	9.5
$G_5:N_2-H\cdots O_{2p}:A_8$	0.967	2.89	0.16	150.5	11.5
$G_5:O_2\cdots H-N_6:A_7$	0.268	3.11	0.20	162.7	9.4

^{a} Frequency for an intact hydrogen bond averaged over the 200 ps trajectory.
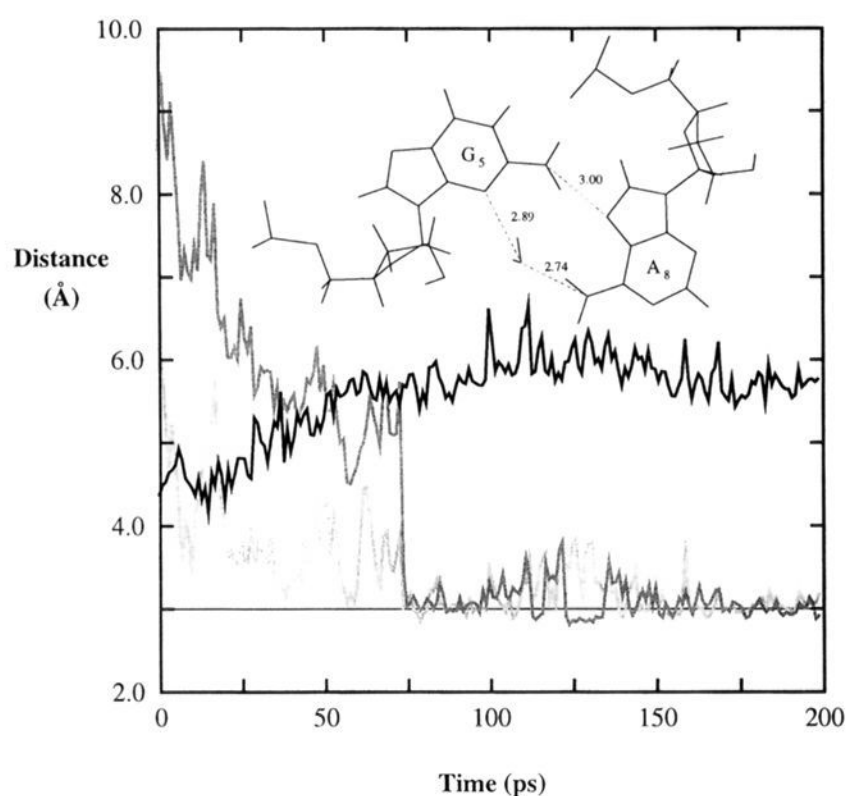
Figure 12. Interatomic separation of $A_8:N_6$ and $G_5:N_3$ (black curve), water oxygen and $A_8:N_6$ (dark gray curve), and water oxygen and $G_5:N_3$ (light gray curve), as a function of time. The solid black line denotes a separation of 3.0 Å, roughly the distance for a water–RNA hydrogen bond. A typical configuration of the water-mediated interaction is displayed above the curves.

$\approx 180°$. The ribose of residue C_6 is in a $C_{2'}$ endo conformation, ($\delta = 120°$ and $P = 120°$), while the A_7 ribose is $C_{3'}$ endo ($\delta = 90°$ and $P = 20°$). These were the initial conformations, and both are consistent with the NMR data; however, A_7 is predominantly in $C_{2'}$ endo according to NMR data. Although no transitions between these pucker conformations were seen, residues G_1, C_6, and U_{12} are seen to make significant excursions in P, covering over 120° during the course of the simulation. In general, the torsion angles for the stem residues oscillate near those values for an ideal A-form helix.[38] Two exceptions are notable. Firstly, the pucker on residue C_{11} is $C_{2'}$ endo, rather than $C_{3'}$ endo. Again, this pucker is in the initial conformation from NMR refinement and does not undergo a transition during the course of the simulation. Secondly, residue C_4 undergoes a rotational transition which moves α and γ from $-sc$ and $+sc$, characteristic of A-helices, to both $+ap$. This rotational transition, a so-called "crankshaft motion", results in a three-bond flip along the phosphodiester backbone, $P-O_{5'}-C_{5'}-C_{4'}$, with virtually no net displacement of the P and $C_{4'}$ atoms.

Crankshaft transitions occur four separate times during the
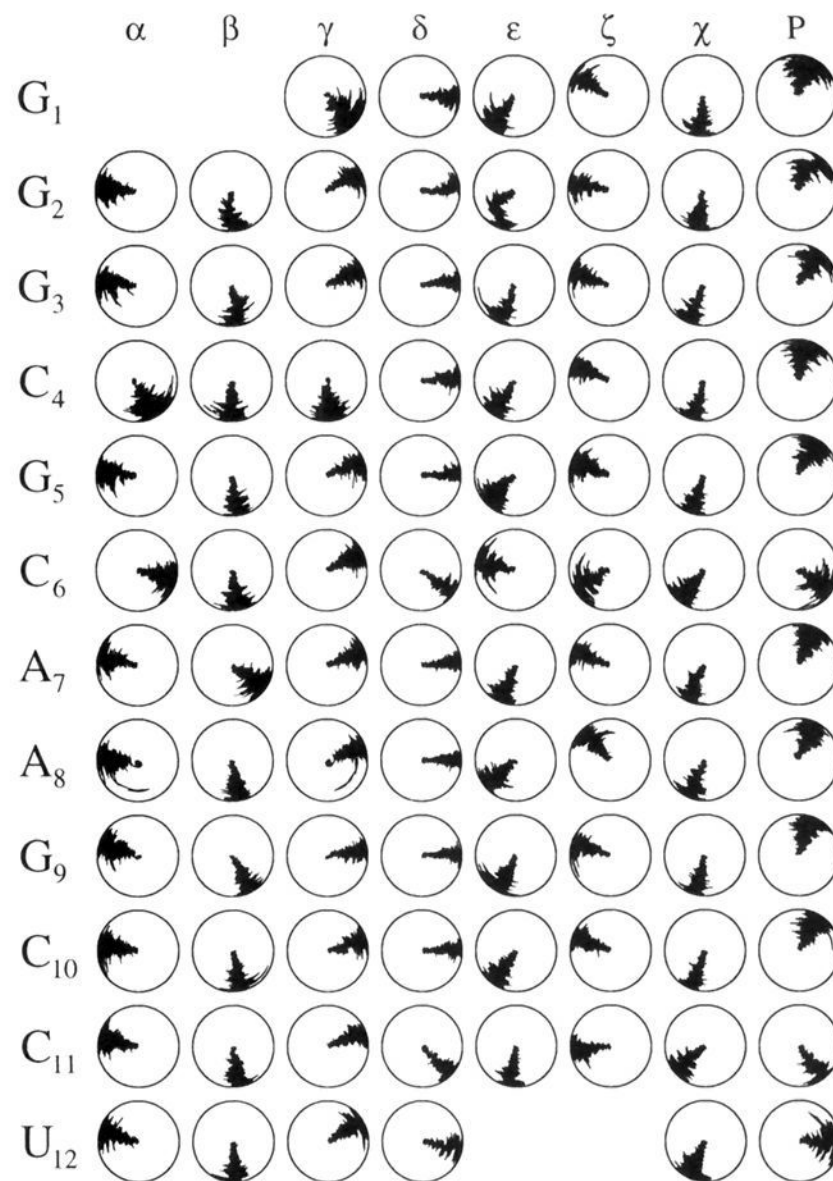
Figure 13. Polar plot of the torsion angle trajectories over the course of the trajectory. Angles are measured in a clockwise direction from the vertical, which denotes an angle of 0°. The radial component corresponds to time, with the initial angle, $t = 0$ ps, displayed at the center and the final angle, $t = 200$ ps, displayed on the bounding circle.

simulation and involve three residues, namely, A_8, C_4, and G_9. These transitions occur within the first 40 ps and also include a slight rotation of β from $+ap$ to $-ap$. The concerted motion of these three angles (α, β, γ) during the bond flips is well illustrated by the time histories of these torsion angles, see Figure 14. Within the first picosecond, loop residue A_8 undergoes a crankshaft rearrangement, rotating from a $(-sc, +ap, +sc)$ conformation characteristic of A-type helices to $(ap, -ap, ap)$. At roughly 15 ps, stem residue G_9 undergoes the reverse rotation from an $(ap, -ap, ap)$ conformation to $(-sc, +ap, +sc)$. Although this rotation primarily involves the α_9 and β_9 torsion angles, ϵ_8 and ζ_8 are seen to rotate ~90° before returning to their original positions. This transition is followed at ~20 ps by another crankshaft rotation for stem residue C_4, which is base paired to G_9 and originally in an A-type helix conformation. The final such transition occurs at 35 ps, returning residue A_8 to its initial orientation $(-sc, -ap, +sc)$. At 150 ps residue A_8 attempts another flip, but it is very short-lived, returning quickly to its initial conformation, see Figure 13. It should be noted that crankshaft transitions have also been observed in double-stranded B-DNA simulations.[22]

Atomic Fluctuations. The individual atomic motions within the RNA molecule are examined in detail here. The fluctuations of an atom relative to its average position will reveal the extent of motion experienced by this atom. The rms deviation of atomic displacements from their average positions, $\langle \delta r_i^2 \rangle^{1/2}$, $\delta r_i = r_i - \langle r_i \rangle$, have been computed for the last 100 ps of the trajectory. The results are displayed in Figure 15. The calculated atomic fluctuations for this RNA molecule are similar in magnitude to those found in typical protein simulations.[7,8]
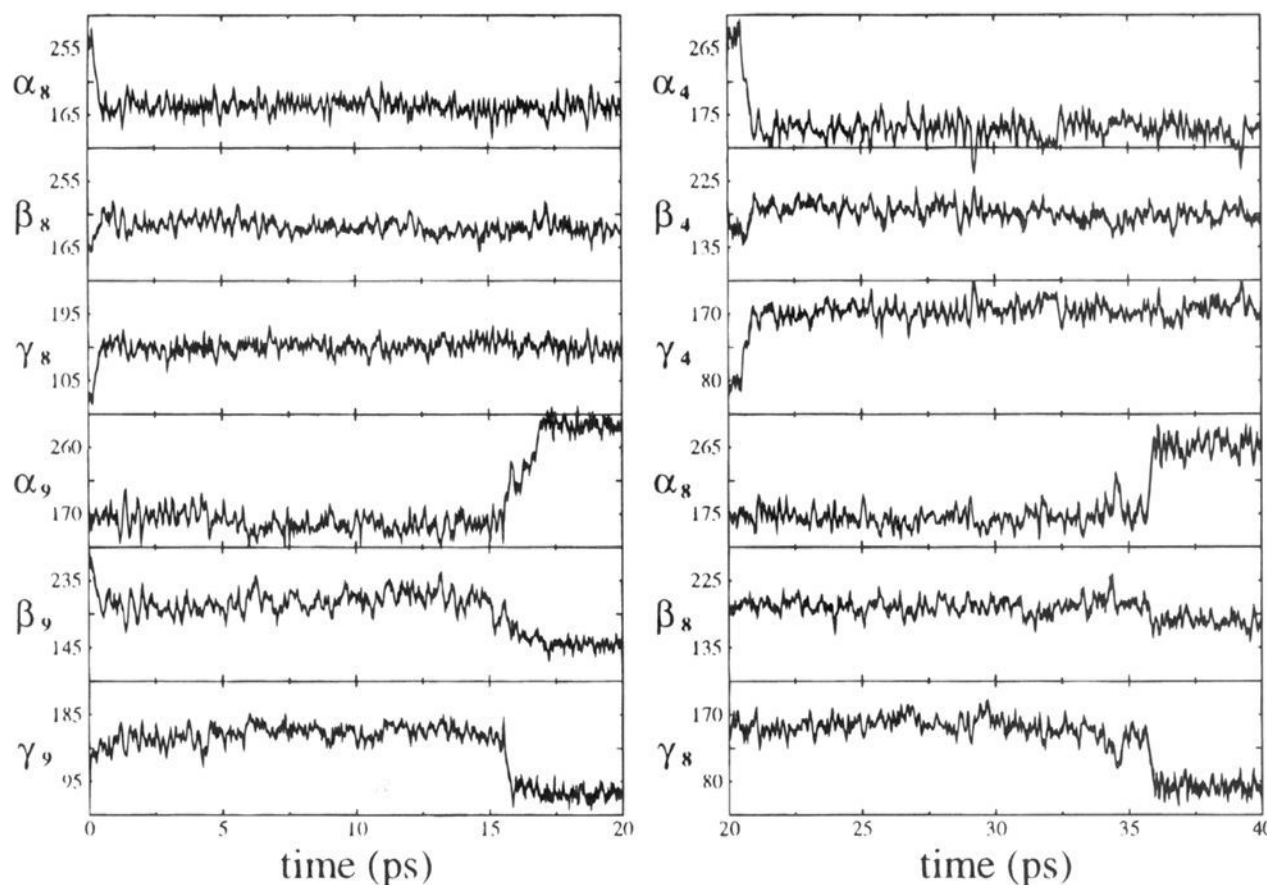
Figure 14. Torsion angle trajectories for the four "crankshaft" transitions during the first 40 ps of the RNA hairpin molecule. The angles are measured in degrees.
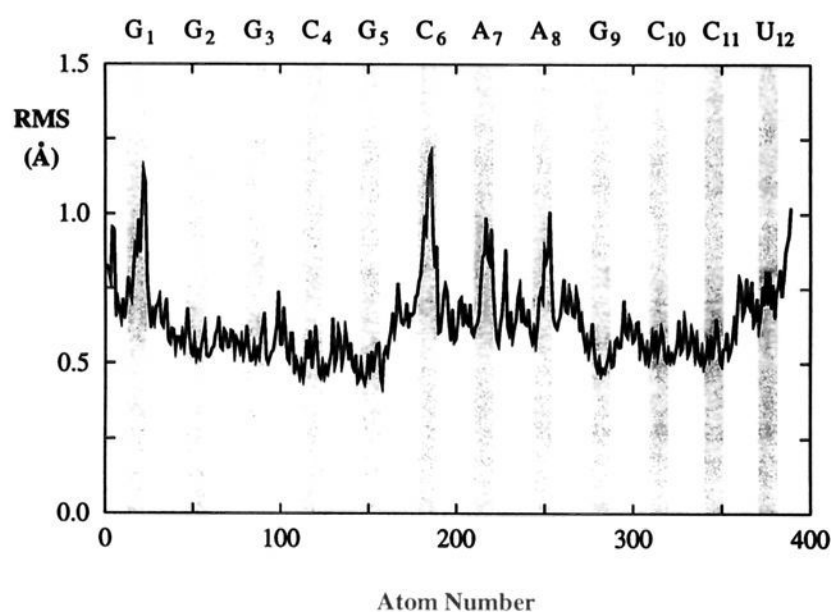
Figure 15. Atomic rms fluctuations of RNA atoms relative to their average position for the last 100 ps of the trajectory. The light gray shaded regions correspond to atoms within the nucleotide bases labeled above the graph.

Overall, there is a base line of ~0.6 Å for amplitudes of atomic motions, with some atoms exhibiting significantly increased displacements from this base line. The atoms in the central stem region comprised by residues 2−5 and 9−11 exhibit the smallest amplitude motions of ~0.6 Å rms deviation, while the terminal base pair and loop residues exhibit the largest amplitude motions, the largest being ~1.5 Å rms fluctuations for the G_1 and C_6 bases. The fraying observed in the hydrogen bonding for the terminal base pair is manifested by increased motion of the nucleotide bases G_1 and U_{12} compared to the central stem, with G_1 showing amplitudes of motion ~50% larger than U_{12}. The nucleotide bases C_6, A_7, and A_8 are seen to fluctuate further from their average positions than the ribose−phosphate backbone atoms, in contrast to the central stem residues 2−4 and 9−11. It is interesting to note that loop residue G_5 exhibits quite small amplitude motion, similar in magnitude to the central stem bases, while the adjacent loop residue C_6 experiences quite large amplitude motion. The unusual hydrogen-bonding pattern and backbone conformation of G_5 may contribute to the low

amplitude of motion, while clearly, the lack of stacking and intermolecular hydrogen bonding of C_6 give rise to its large motion.

In order to assess cooperative motion within the molecule, the cross-correlation of atomic displacements from their average positions, C_{ij}, was computed, where

$$C_{ij} = \langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j \rangle / \langle \delta\mathbf{r}_i^2 \rangle^{1/2} \langle \delta\mathbf{r}_j^2 \rangle^{1/2}$$

Since the atoms are numbered sequentially along the oligo-nucleotide, the matrix of C_{ij} will clearly exhibit correlated and anti-correlated motion within the molecule. Similar analyses have been performed on cytochrome c[41] and HIV-1 protease[42] and were shown to be useful in identifying domain motions within these proteins. An average C_{ij} matrix for the RNA molecule was obtained by averaging the five C_{ij} matrices computed from sequential 20 ps blocks spanning the last half of the trajectory. The results are presented in Figure 16 with the positive correlations plotted in the upper left triangle and the negative correlations plotted in the lower right triangle. The average C_{ij} matrix computed here does not differ significantly from those computed with 10 and 40 ps block averages.

Distinct groups of atoms are seen to have well-correlated motions. The atoms composing the stem residues 1−4 (atoms 1−131) display well correlated motions for both nucleotide base atoms and ribose−phosphate backbone atoms. The backbone atoms for residues 5, 6, and 7 (atoms 132−229) are correlated, as are the atoms from residue 6 backbone through residue 10 (atoms 166−328). The correlated motion within the secondary structure is easily identified by the strong peaks running perpendicularly to the diagonal. These correlations are centered on the nucleotide bases and roughly span three adjacent residues, the central correlation being the strongest. These central peaks correspond to the concerted motions of the three Watson−Crick base pairs composing the stem as well as the terminal G_1−U_{12} wobble pair and the unusual G_5−A_8 loop interaction. The atoms composing the nucleotide base G_5 (atoms 144−158) show strong

(41) McCammon, J. A. *Rep. Prog. Phys.* **1984**, *47*, 1−46.
(42) Swaminathan, S.; Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 2717−2721.
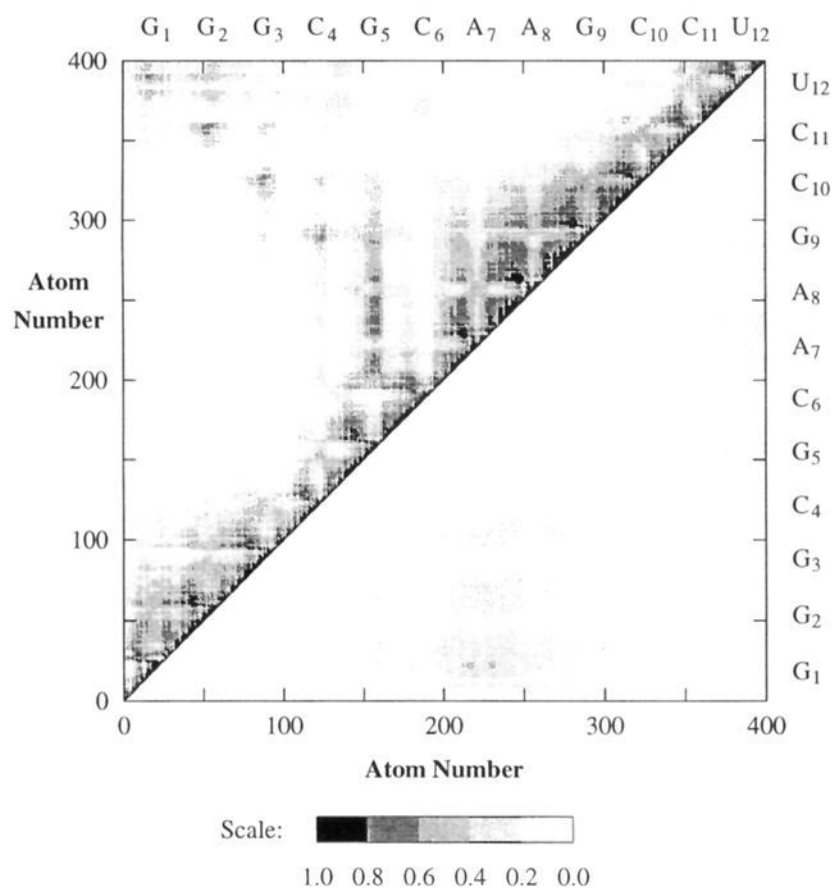
Figure 16. Atomic cross-correlations for the RNA hairpin molecule obtained from the mean of five 20 ps averages over the 100–200 ps range of the simulation. Positive correlations appear above the diagonal, and negative correlations appear below. The labels for the nucleotide bases are centered on those atom positions.

correlations with atoms from residues 7 through 10, while the C_6 base atoms (atoms 178–189) do not correlate well with any other group of atoms. In general, the correlated motions seen here indicate that the atoms composing the two distinct secondary structural units of the hairpin molecule, the stem and loop, behave dynamically as two distinct units.

There are two regions of modest anti-correlations observed, those of residues 6–8 with 1–4 and 6–8 with 11 and 12. Such an anti-correlation may be due to a wagging motion of the hairpin molecule, with the loop atoms moving as a group separate from the stem. Such a loop–stem wagging motion, with residues 5 and 9 acting as the hinge, is consistent with the oscillatory motion observed in the radius of gyration about the *b* and *c* principal axes. To examine these types of correlated motions more carefully, an analysis of the delocalized collective motions for the RNA molecule is currently being performed from the 200 ps trajectory.[43] This analysis will reveal the underlying cooperative motions of the molecule along the essential, many-atom directions and should offer additional insight into the general trends observed in the cross-correlation matrix.[44]

IV. Conclusions

The structural and dynamical features of an RNA hairpin molecule containing a tetranucleotide loop have been computed from an unconstrained MD simulation and compared with NMR solution studies for the same molecule. The agreement between the unconstrained simulation and the NMR experiments is quite good for a number of measures of structure, including inter-proton distances and average torsion angle values. The three-dimensional structure, determined from NMR data, was preserved throughout the 200 ps trajectory, with the simulated structure exhibiting rms deviations from the initial NMR-derived structure comparable to other simulations of macromolecules in solution. The secondary structure was well maintained throughout the simulation, with all hydrogen bonds of the stem, with the exception of the terminal G_1–U_{12} base pair, remaining intact throughout the simulation.

The main differences in the computed structure relative to the NMR-refined structure may be primarily due to solvent effects since the NMR refinement was carried out *in vacuo*. The main differences include a slight lengthening along the two longest molecular axes and a slight rearrangement of the nucleotides G_5 and A_8 in the loop. It is well-known that vacuum phase calculations can result in more compact structures compared to aqueous environments.[45] This may account for the slight expansion observed along the molecular axes. The rearrangement of residues G_5 and A_8 was made to accommodate a water molecule which bridged the two bases. It is obviously not possible to obtain such a result without explicit account of solvent molecules in the simulation.

Overall, the current results are quite encouraging for the use of MD simulations for studying structural features of single-stranded RNA molecules in solution. Even though a number of drastic simplifications and approximations in the representation of the system have been made, the computed results are seen to be quite reasonable. The use of Ewald sums in order to account for electrostatic screening effects may have contributed greatly to the stability of the oligonucleotide. Indeed, such stabilization due to long-range electrostatics has been observed in aqueous peptide simulations.[15] The interaction potentials embodied in the OPLS–Amber force field, although lacking explicit atomic polarizability, appear to be adequately realistic for the present purposes of equilibrium MD simulations. It is of great interest to explore quantitatively the energetic aspects of the force field by evaluating sequence-dependent thermodynamic differences of RNA duplexes for which a wealth of experimental data exists.[46] Free energy simulations may aid in the interpretation of such experimental data and will be the topic of future work.

(43) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699. Garcia, A. E.; Soumpasis, D. M.; Jovin, T. M. *Biophys. J.* **1994**, *66*, 1742–1755.

(44) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412–425.

(45) Reference 8, pp 137–139.

(46) See, for example: Turner, D. H.; Sugimoto, N.; Jaeger, J. A.; Longfellow, C. E.; Freier, S. M.; Kierzek, R. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 123–133.