JCTC Journal of Chemical Theory and Computation

What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study

Alessandra Villa* and Gerhard Stock

Institute for Physical and Theoretical Chemistry, J. W. Goethe University, Max-von-Laue-Strasse 7, 60438 Frankfurt am Main, Germany

Received May 5, 2006

Abstract: Classical molecular dynamics simulations of a 14-mer UUCG RNA hairpin are performed to study its conformational dynamics and corresponding NMR relaxation parameters. The direct calculation of the relaxation rates from the trajectory yields good agreement with experiment, indicating the validity of the theoretical model. Various ways to provide a link between theory and experiment are considered, including the "model-free approach" of Lipari and Szabo and Gaussian axial fluctuation model of Brüschweilwer. It is studied if the underlying assumptions of these approaches are satisfied in the case of a flexible RNA hairpin. Being consistent with the analysis of the NMR experiments, Lipari−Szabo fits of the first 100 or 1000 ps of the internal correlation functions lead to a nice agreement between calculated and experimental order parameters and internal correlation times. Finally, the relation between NMR order parameters and the underlying internal motion of the RNA hairpin is discussed in detail. A principal component analysis reveals that the principal motions of the molecule account only partially for the measured NMR order parameters, because the latter are insensitive to internal dynamics occurring on a nanosecond time scale due to molecular tumbling.

1. Introduction

Conformational dynamics may play a key role in the function of biomolecules such as proteins, DNA, and RNA. The flexible parts of a protein or ribonucleic acid, such as loop regions, are often involved in mediating specific interactions, for example, between protein and RNA during a binding process.[1−3] In the case of RNA, the flexibility of the loop may directly affect both the specificity and the affinity of the binding.[4] To account for the function of RNA systems, a site-specific dynamic description is therefore an important complement to static structural information.[5,6]

Internal motions in biomolecules occur on a wide range of time scales, from femtoseconds to seconds. Molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) spin relaxation measurements are valuable tools to gain access to fast (i.e., subnanosecond) internal motions. MD simulations[7] provide directly information at the atomistic level on inter- and intramolecular motions using an empirical force field as molecular model. NMR relaxation measurements[8] yield the dipolar correlation function, from which dynamical quantities such as generalized order parameters S^2 and effective correlation times τ_e can be extracted.[9−12] In particular, the combination of NMR and MD investigations has been shown to provide a comprehensive description of fast conformational dynamics of proteins.[13−20] On one hand, we may use experimental NMR results as benchmark data to study the accuracy of the MD description; on the other hand, we may employ the MD trajectory to provide a microscopic interpretation of the NMR experiments. While the description of structure and dynamics of proteins is well established, RNA systems have been comparatively little studied using NMR relaxation[21−25] or MD simulation. (For general reviews on RNA simulations see refs 26−33.)

* Corresponding author phone: +49-69-79829711; fax: +49-69-79829709; e-mail: villa@theochem.uni-frankfurt.de.

10.1021/ct600160z CCC: $33.50
Published on Web 08/22/2006

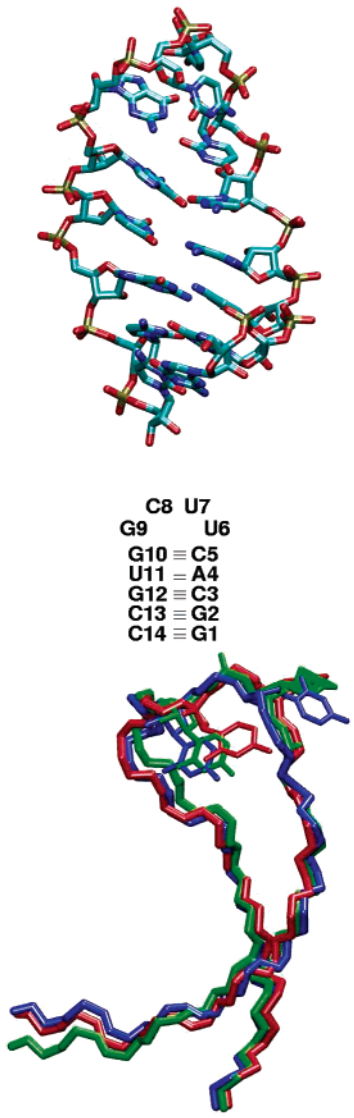

Figure 1. The 14-mer RNA loop. Top: representative MD snapshot at 50 ns. Center: secondary structure including base-pair hydrogen bonds and residue numbering. Bottom: backbone atoms and residues U7 and C8, describing the main hairpin motion along the first three principal components of the trajectory. Shown are snapshots at 10 ns (blue), 30 ns (green), and 50 ns (red). The pictures were performed using the graphical package VMD.[59]

In this work we present an MD study of the fast dynamics of the RNA hairpin ggcacUUCGgugcc (Figure 1). This hairpin belongs to one of the most stable tetraloop families[34] and is therefore a popular model system for theoretical[35−38] and experimental[39−41] investigations. The present study has also been motivated by recent work on UUCG loops by Duchardt and Schwalbe,[24] who performed detailed NMR ^{13}C relaxation measurements for the carbon atoms $C_{1'}$ in the ribose moiety and for the carbon atoms C_6 and C_8 in the pyrimidine and purine residues. Previously, ^{15}N relaxation measurements have been used by Akke et al.[21] to characterize the base dynamics of the UUCG loop.

First, the MD results are compared to the structures of available NMR[39,41] and crystallographic[40] studies on similar UUCG loops. Second, the MD trajectory is used to directly calculate the NMR relaxation rates,[18] in order to avoid most of the assumptions usually employed in the experimental analysis. Then we consider various ways to provide a link between theory and experiment, including the so-called model-free approach developed by Lipari and Szabo[9] and the Gaussian axial fluctuation model.[42] In particular, we study if the underlying assumptions of these approaches (such as the separation of overall and internal motion and the use of monoexponential Lipari−Szabo fits of the internal correlation function) are satisfied in the case of an RNA hairpin. Finally, the relation between NMR order parameters and the underlying internal motion of the RNA hairpin is discussed in detail.

2. Theory and Methods

2.1. Molecular Dynamics Simulations. The MD simulations were performed using the GROMACS suite of programs (version 3.2).[43,44] The AMBER force field (parm98)[45,46] was employed to describe the 14-mer UUCG RNA loop. The loop was placed in a rhombic dodecahedron box (edge length approximately 5 nm), which was subsequently filled with 2713 TIP3P water molecules.[47] To neutralize the system, 13 sodium ions were placed randomly in the simulation box.

A twin range cutoff was used for the Lennard-Jones interactions, that is, interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every 5 steps. The particle mesh Ewald method[48] was employed to treat Coulomb interactions, using a switching distance of 1.0 nm, a grid of 0.12 nm, and a beta value of 3.1 nm^{-1}. Constant pressure p and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and 298 K, using the Berendsen barostat and thermostat, respectively.[49] The RNA, the ions, and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps, and the isothermal compressibility was $4.5 \cdot 10^{-5}$ bar^{-1}. The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm.[50] All other bond distances were constrained using the LINCS algorithm.[51] A leapfrog integrator with an integration time step of 2 fs was used.

To obtain the starting structure of the UUCG hairpin, the loop was modeled based on the crystallographic structure of Ennifar et al.,[40] while the stem structure was built using tools of the AMBER6 software.[52] Following 20 ns of equilibration, the system was simulated for 50 ns. Analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of a hydrogen bond, an acceptor−donor distance smaller than 0.35 nm was requested.

2.2. NMR Relaxation Parameters. We have focused on the relaxation of the ^{13}C nuclear spin through dipolar interaction with the attached ^{1}H. According to the relaxation theory of Bloch, Wangness, and Redfield[8], the spin−lattice (R_1), the spin−spin (R_2) relaxation rates, and the nuclear Overhauser enhancement (NOE) are given by

$$R_1 = \frac{d}{4}[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_C + \omega_H)] + cJ(\omega_C) \quad (1)$$

$$R_2 = \frac{d}{8}[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H) + 6J(\omega_C + \omega_H)] + \frac{c}{6}[4J(0) + 3J(\omega_C)] \quad (2)$$

$$\mathrm{NOE} = 1 + \frac{d}{4R_1}\frac{\gamma_C}{\gamma_H}[6J(\omega_C + \omega_H) - J(\omega_H - \omega_C)] \quad (3)$$

where $d = (\mu_0^2/4\pi^2)(\hbar^2\gamma_C^2\gamma_H^2/r_{CH}^6)$, $c = \omega_C^2\Delta\sigma_C^2/3$, and $J(\omega)$ represents the spectra density defined in eq 5. Here, μ_0 is the vacuum permeability, $\hbar$ is Planck's constant, and γ_X is the gyromagnetic ratio of nucleus X. r_{CH} is the intermolecular distance between the two nuclei; for the bond $C_{1'}-H_{1'}$ a distance of 0.109 nm and for C_6-H_6 and C_8-H_8 a distance of 0.108 nm was used. $\Delta\sigma_C$ is the ^{13}C chemical shift anisotropy; for $C_{1'}$ a value of 45 ppm, for C_6 a value of -179 ppm, and for C_8 a value of -134 ppm were used. A value of 600.13 MHz was used for ω_H and 150.90 MHz for ω_C. All these values have been chosen in line with the NMR relaxation experiments of Duchardt and Schwalbe.[24]

2.3. Correlation Functions. The NMR relaxation due to the dipole−dipole interaction between two nuclei (i.e., carbon and hydrogen) can be described by the correlation function[8]

$$C(t) = \langle P_2(\vec{\mu}(0)\cdot\vec{\mu}(t))\rangle \quad (4)$$

where $\vec{\mu}$ is a unit vector pointing along the C−H bond, $P_2(x) = (1/2)(3x^2 - 1)$ is the second Legendre polynomial, and $\langle..\rangle$ denotes an equilibrium average. The spectral density

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)\,dt \quad (5)$$

which determines the relaxation parameters in eqs 1−3, is given by the Fourier transform of the correlation function.

Assuming that overall and internal motions of the molecule are independent, the total correlation function $C(t)$ can be factorized in the correlation functions for overall motion, $C_O(t)$, and for internal motion, $C_I(t)$, respectively:

$$C(t) = C_O(t)C_I(t) \quad (6)$$

The total correlation functions were calculated for the $C_{1'}-H_{1'}$, C_6-H_6, and C_8-H_8 dipoles of all residues according to eq 4. To obtain the internal correlation functions, each conformation was translated and rotated to give the best fit to a reference structure. Since no large conformational arrangement took place during the 50 ns simulation, the molecule-fixed frame is unambiguously defined by this approach. Subsequently, the correlation functions for overall motion were calculated using eq 6. Assuming that the overall motion of the molecule is isotropic, this correlation function is given by

$$C_O(t) = \frac{1}{5}e^{-t/\tau_c} \quad (7)$$

where the rotational correlation time τ_c is proportional to the inverse of the rotation diffusion constant.

In the model-free approach of Lipari−Szabo,[9] the internal and overall motions are assumed to be independent, and the internal correlation function is given by the following relation

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad (8)$$

where S^2 is the order parameter and τ_e is the effective (or internal) correlation time τ_e for the C−H dipole. Insertion of eq 8 in eqs 5−7 yields the spectral density

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1 + \tau_c^2\omega^2} + \frac{(1 - S^2)\tau}{1 + \tau^2\omega^2}\right) \quad (9)$$

with $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$.

2.4. Order Parameters. We have employed three different approaches to calculate the order parameters.

Lipari−Szabo Fit. Employing the Lipari−Szabo form of the internal correlation function (eq 8), S^2 was fitted using the first 100 ps or the first 1 ns of the MD internal correlation.

Equilibrium Average. Using the general property of correlation functions that $\lim_{t\to\infty}\langle A(0)B(t)\rangle = \langle A\rangle\langle B\rangle$, the order parameter can be determined by[8]

$$S_{eq}^2 = \lim_{t\to\infty} C_I(t) = \frac{4\pi}{5}\sum_{m=-2}^{m=2}|\langle Y_{2m}(\theta, \varphi)\rangle|^2 \quad (10)$$

where Y_{2m} is the spherical harmonic function of rank 2, $\theta(t)$ and $\varphi(t)$ are the polar angles defining the orientation of the dipole C−H at each snapshot of the trajectory, and $\langle\cdots\rangle$ denotes the average over all snapshots. This corresponds to a Lipari−Szabo fit using the full time range of the internal correlation function. It should be stressed that eq 10 avoids the cumbersome calculation of time-dependent correlation functions. In particular, this allows us to use highly efficient Monte Carlo schemes (e.g., like the popular replica exchange MD[53]) to calculate the equilibrium average in eq 10.

GAF Model. Assuming that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ, the order parameter can be related to motions around the $C_{1'}-N_1$ or $C_{1'}-N_9$ bonds. Assuming furthermore a Gaussian distribution for the dihedral angle χ, the Gaussian axial fluctuation (GAF) model[42] leads to the following expression for the order parameter:

$$S_{gaf}^2 = 1 - 3\sin^2\chi\left\{\cos^2\chi(1 - e^{-\sigma_\chi^2}) + \frac{1}{4}\sin^2\chi(1 - e^{-4\sigma_\chi^2})\right\} \quad (11)$$

Here the dihedral angle χ is defined by $O_{4'}-C_{1'}-N_1-C_2$ in the pyrimidine and by $O_{4'}-C_{1'}-N_9-C_4$ in the purine, and σ_χ is its standard deviation.

2.5. Principal Component Analysis. Principal component analysis is an efficient method to represent the motion of a $3N$-dimensional system in terms of a few "principal" components.[54−56] The basic idea is that the correlations of the motions are represented by the covariance matrix

$$\sigma_{ij} = \langle(q_i - \langle q_i\rangle)(q_j - \langle q_j\rangle)\rangle \quad (12)$$

where $q_1, ..., q_{3N}$ are the mass-weighted Cartesian coordinates of the solute molecule, and $\langle\cdots\rangle$ denotes the average over all sampled conformations. By diagonalizing σ, we obtain $3N$ eigenvectors υ_n and eigenvalues λ_n, which are rank-

ordered in descending order, i.e., λ_1 is the largest eigenvalue. We may expand the MD trajectory $q(t) = \{q_i(t)\}$ on the basis of the eigenvectors ν_n according to

$$x^{(n)}(t) = \sum_{i=1}^{n} [\nu_i \cdot q(t)] \nu_i \qquad (13)$$

While for $n = 3N$ the expansion becomes exact [$x^{(3N)}(t) = q(t)$], for small n (in practice, $n = 1-5$) $x^{(n)}(t)$ approximates the motion of the system in terms of a few principal components representing the "essential" dynamics of the system.[56] The projection of the MD trajectory on the first n eigenvectors is used in the calculation of the order parameters.

3. Results and Discussion

3.1. Structural Features of the UUCG Loop. The 14-mer tetraloop was simulated for 50 ns in explicit water at 298 K. The root-mean-square deviation for all atom coordinates, after fitting to a reference structure, had an average value of 0.19 (±0.03) nm. As an illustration, Figure 1 shows a representative snapshot of the hairpin at the end of the trajectory. The residues forming the stem are all involved in Watson−Crick base-pairs and stacking interactions. The UUCG loop is mainly stabilized by hydrogen bonds between the residues U6 and G9. The residue U7 is looped out, and the residue C8 is unpaired. All bases are in an anti conformation, except for G9 which shows a syn conformation and for U7 which is in an equilibrium between syn and anti conformations.

The structures sampled by the MD simulation are in overall agreement with previous investigations on similar UUCG loops, including MD studies[35−38] as well as NMR[39,41] and crystallographic[40] experiments. In particular, experiments and simulation largely agree on the interactions that stabilize the secondary structure of the hairpin. The loop residues U6 and G9 are involved in base−base and base−sugar hydrogen bonds, and there is a hydrogen bond between the C8 base and the U7 phosphate oxygen as observed in the experimental structures. Moreover, a weak interaction is observed between the 2′-OH group of U7 and the base G9. In the crystallographic structure, the U7 sugar oxygen is hydrogen bonded to the G9 base oxygen, while this is not the case in the NMR structures. The analysis of the calculated backbone dihedral angles of the loop residues U6−G9 indicates the presence of at least two conformational states (see section 3.4). Two clusters of structures are also obtained by the NMR refinement, while only one structure is reported in the X-ray study. One of the two MD conformers shows a better agreement with the crystal structure and with one of the NMR clusters as shown in Tables 4 and 5 in the Supporting Information. In the simulation, dihedral angle transitions are observed for the residues U6−U7, while the two groups of NMR structures differ in residues C8−G9.

3.2. NMR Relaxation Parameters. The internal and total correlation functions of all $C_{1'}-H_{1'}$ sugar bonds and C_6-H_6 and C_8-H_8 base bonds have been calculated according to eq 4, using the 50 ns trajectory with and without subtracting the overall motion, respectively (see Methods). The total correlation functions decay on average within a nanosecond (see Figure 2), while the internal correlation functions generally show a decay on a time scale of 10 ps (see Figure 3 for some representative examples). Monoexponential fits of the internal correlation function are seen to be appropriate in most cases, except for the loop residues U6−G9. In the latter cases, the internal correlation functions exhibit a multiexponential decay on pico- and nanosecond time scales.

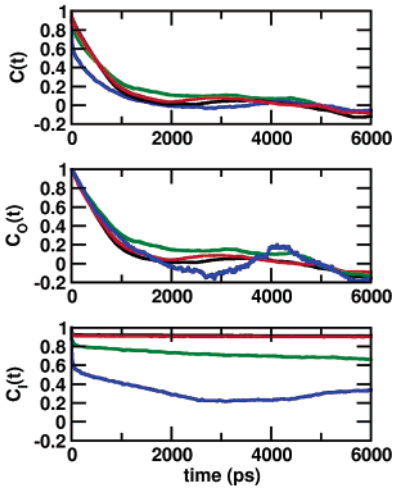

Figure 2. Total (top), overall-motion (middle), and internal (bottom) correlation functions of the dipoles $C_{1'}-H_{1'}$ (black and green line) and C_6-H_6 (red and blue line) for the residues C3 and U7, respectively.

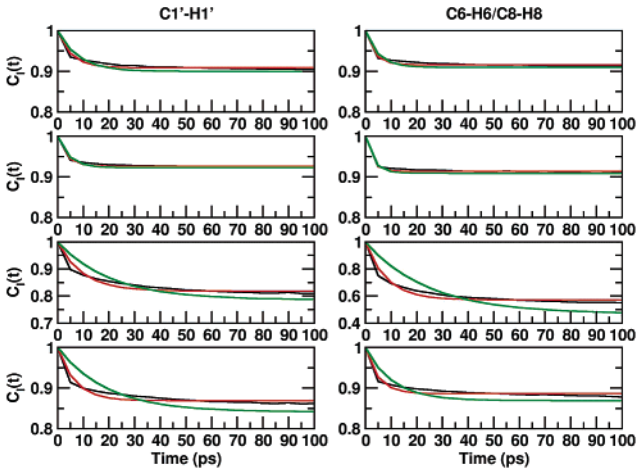

Figure 3. Internal correlation functions (black lines) of the dipoles $C_{1'}-H_{1'}$ (left side) and C_6-H_6 and C_8-H_8 (right side) for the residues G2, C3, U7, and G9 (from top to bottom). Exponential fits using the first 100 ps and the first 1 ns of the correlation function are shown in red and green, respectively.

Let us first study the validity of the assumption that overall and internal motions are separable, which leads to the factorization of the total correlation function $C(t)$ into components $C_O(t)$ and $C_I(t)$ describing overall and internal motion, respectively. As a representative example, Figure 2 shows the three correlation functions for the sugar and base dipole motions of the relatively rigid stem residue C3 and the most flexible loop residue U7. The overall and internal correlation functions of the stem residue exhibit an obvious separation of time scales (1 ns vs 10 ps) and are therefore clearly separable. In the case of the loop residue U7, on the

other hand, both correlation functions decay on a nanosecond time scale, and one may expect a coupling of overall and internal motions. However, for the relatively short times (0.1−1 ns) that are relevant in the analysis of the experimental NMR data, the internal correlation function $C_I(t)$ may be approximated by a 24 ps decay time (see Figure 3) and $C_O(t)$ decays just as the other overall-motion correlation functions. That is, for short times the factorization approximation is not expected to change the results of the calculation of NMR data.

The correlation times τ_c were obtained by fitting the overall-motion correlation function of each dipole to the monoexponential function e^{-t/τ_c}. The fitted τ_c has an average value of 0.7 ns, which is clearly shorter than values obtained by using a hydrodynamics model[24] for the whole hairpin (2.35 ns for $C_{1'}$ and 2.17 ns for C_6/C_8). The main reason for this deviation appears to be the different viscosity of the solvent in experiment and simulation. The diffusion constant for TIP3P water is around 5.56−5.70 10^{-5} cm^2 s^{-1},[57,58] that is, about two times larger than the corresponding experimental value. To quantitatively calculate NMR observables, we therefore cannot directly take the MD correlation functions. Instead, we assume that the correlation function can be factorized (eq 6) and use the correlation time τ_c obtained from the NMR analysis.[24]

Using the experimental correlation time τ_c and the calculated internal correlation functions, we have employed eqs 1−7 to compute the relaxation parameters R_1, R_2, and NOE for all investigated dipoles. As shown by black circles in Figure 4, the calculated values are in good agreement with the experimental results.[24] The relative error $\sum_i|x_i^{MD} - x_i^{exp}|/\sum_i x_i^{exp}$ is 0.04 for R_1, 0.14 for R_2, and 0.03 for NOE, respectively. Employing an axially symmetric diffusion model of overall rotation (with $\tau_c = 2.44$ and 2.18 ns and a diffusion anisotropy of 1.32 and 1.37 for $C_{1'}$ and C_6/C_8, respectively),[21,24] the relative errors are almost identical with respect to the isotropic diffusion model (0.04 for R_1, 0.13 for R_2, and 0.03 for NOE). The relatively large discrepancy obtained for the spin−spin relaxation rate R_2 may be related to the fact that in numerous cases the experimental R_2 values have been corrected for conformational exchange contributions during the experimental analysis.[24] To assess the validity of the Lipari−Szabo approach, the relaxation constants have also been obtained by fitting the first 100 and 1000 ps of C_I(t) to the Lipari−Szabo spectral density (red squares in Figure 4). Here, the relative errors for the 100 ps and 1 ns fit are 0.05 and 0.06 for R_1, 0.12 and 0.13 for R_2, and 0.02 and 0.03 for NOE, respectively. The relative errors of both fits are virtually identical and do not differ from the errors obtained in the direct MD evaluation of $C_I(t)$.

The qualitative agreement between the directly calculated MD data and the NMR relaxation parameters indicates that the force field and the simulation time scale used in this study are appropriate to describe the relaxation of the C−H bonds monitored in the NMR experiment. Thus, the MD simulation may be employed to reveal the dynamic information included in the experimental NMR data. Furthermore, we may use the MD data to compare and validate various methods to calculate the order parameters of the RNA hairpin.

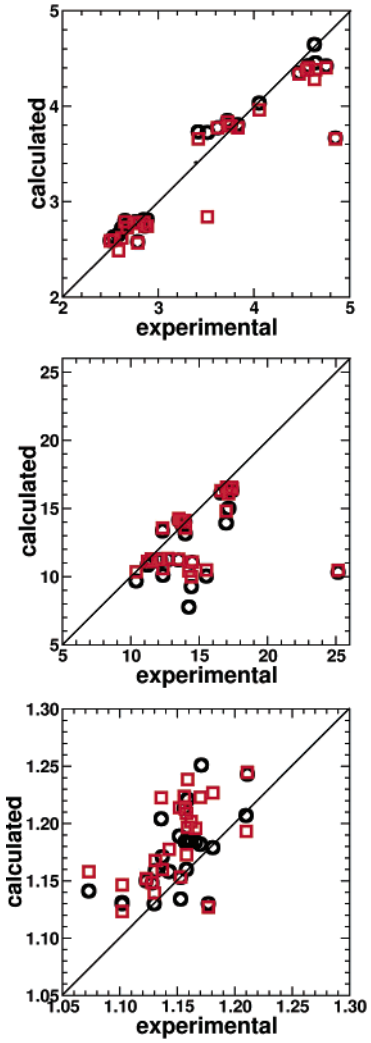

Figure 4. Experimental[24] vs calculated values of spin−lattice (top) and spin−spin (middle) relaxation rates (s^{-1}) and NOE (bottom) for all C−H dipoles. The black circles and red squares correspond to calculations using the spectral densities obtained directly from the correlation function and from a 100 ps Lipari−Szabo fit, respectively.

3.3. Order Parameters. Experimental and calculated order parameters for the $C_{1'}-H_{1'}$, C_6-H_6, and C_8-H_8 dipoles of all residues are compared in Figure 5 and in Tables 1 and 2. In the experimental study,[24] two models have been used to analyze the relaxation data, assuming either isotropic or axially symmetric diffusion. Both models are seen to give quite similar results for the order parameters and show that the loop residues, in particular U7, exhibit enhanced conformational fluctuations. The experimental effective correlation times τ_e listed in Tables 1 and 2 are mostly below 10 ps, with the exception of the residues C5, U7, and C14.

As detailed in the methods section, three different approaches have been used to calculate NMR order parameter from the MD trajectory. In the first approach, the order parameters S^2 and the internal correlation times τ_e were obtained by fitting the first 100 ps and the first 1 ns of the MD internal correlation to eq 8. Both Lipari−Szabo fits yield a good agreement with the experimental order parameters. In Figure 5 the results for the 100 ps fit are reported together with the experimental values. Tables 1 and 2 show that the

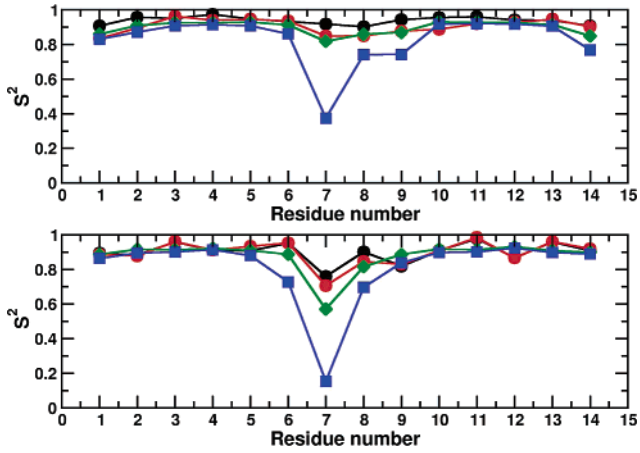

Figure 5. Order parameter S^2 for $C_{1'}$ (top) and C_6/C_8 (bottom) as a function of the residue number. The experimental values[24] are shown in black (using the isotropic diffusion model) and in red (using the axially symmetric diffusion model). Calculated values are shown in green (100 ps fit) and in blue (using eq 10).

Table 1: Internal Correlation Times τ_e and Order Parameters of $C_{1'}$ for the 14 Residues of the Hairpin[a]

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 10)	experiment	
	τ_e (ps)	S^2	τ_e (ps)	S^2	S_{eq}^2	τ_e (ps)	S^2
G1	6.3	0.859	9.1	0.847	0.830	<10	0.835
G2	5.2	0.909	8.1	0.900	0.871	<10	0.896
C3	3.6	0.926	4.4	0.922	0.908	<10	0.963
A4	2.7	0.925	3.2	0.923	0.914	<10	0.939
C5	3.2	0.928	4.4	0.924	0.908	<10	0.946
U6	5.4	0.913	14.6	0.897	0.861	<10	0.936
U7	9.7	0.818	20.6	0.786	0.374	15.74	0.848
C8	9.9	0.858	23.8	0.829	0.741	<10	0.850
G9	6.6	0.869	19.1	0.841	0.742	<10	0.877
G10	3.2	0.931	4.3	0.926	0.917	<10	0.888
U11	2.5	0.927	2.9	0.925	0.919	<10	0.920
G12	3.1	0.926	3.6	0.924	0.918	<10	0.928
C13	3.8	0.914	4.7	0.910	0.906	<10	0.946
C14	8.3	0.849	52.9	0.782	0.767	412.14	0.902

[a] Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using eq 10) and from the NMR experiment[24] (isotropic model).

100 ps fit is more appropriate to be compared to the experimental data, since it better reproduces the experimental effective correlation times.

In the second approach, we have used the equilibrium average in eq 10 to calculate the order parameters. This corresponds to a Lipari−Szabo fit using the full time range of the internal correlation function. Figure 5 demonstrates that the resulting order parameters only agree for the relatively rigid stem residues but not for the flexible loop residues. The reason for this discrepancy is that, by using the entire internal correlation function, the order parameters contain information also on internal motions occurring on a nanosecond time scale. For example, the loop conformational rearrangement and the anti−syn transitions of the U7 base represent internal motions on a time scale longer than 5 ns. The experimental analysis is limited by the decay of the overall-motion correlation function due to molecular tumbling (≈2 ns). As a consequence, possibly existing internal motions on a nanosecond time scale are not reflected in the experimental data. Although S_{eq}^2 correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to NMR order parameter.

Table 2: Internal Correlation Times τ_e and Order Parameters of C_6/C_8 for the 14 Residues of the Hairpin[a]

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 10)	experiment		GAF
	τ_e (ps)	S^2	τ_e (ps)	S^2	S_{eq}^2	τ_e (ps)	S^2	S_{gaf}^2
G1	4.4	0.886	6.4	0.877	0.864	<10	0.886	0.998
G2	3.5	0.916	4.9	0.910	0.894	<10	0.878	0.995
C3	2.5	0.914	3.2	0.909	0.900	<10	0.961	0.993
A4	2.3	0.921	2.8	0.918	0.913	<10	0.912	0.987
C5	2.6	0.909	3.5	0.904	0.880	121.59	0.933	0.977
U6	4.1	0.887	19.1	0.858	0.727	<10	0.953	0.928
U7	8.3	0.571	24.3	0.469	0.153	13.38	0.706	
C8	7.3	0.816	18.7	0.780	0.696	<10	0.845	0.938
G9	4.9	0.886	10.5	0.869	0.838	<10	0.830	0.770
G10	2.4	0.917	3.3	0.911	0.897	<10	0.909	0.991
U11	2.3	0.914	2.7	0.911	0.901	<10	0.984	0.986
G12	1.9	0.931	2.2	0.929	0.923	<10	0.866	0.998
C13	2.9	0.909	3.7	0.904	0.898	<10	0.960	0.993
C14	3.6	0.899	4.6	0.894	0.889	38.08	0.919	0.959

[a] Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using eq 10), from the NMR experiment[24] (isotropic model), and from the GAF model.

Finally, we have applied the GAF model (eq 11), which assumes that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ. To this end, we have calculated the distribution function of χ, which exhibits a single peak for all residues except U7. In the latter case, the base adopts both anti and syn conformations during the simulation, and the GAF model is not applicable (see Methods). Table 5 in the Supporting Information lists the mean and the variance of χ for all residues as calculated from the MD trajectory. All results are found to be in good agreement with the experimental data.[39,24] The order parameters obtained from the GAF model are reported in Table 2. Except for the residue G9, the values for S_{gaf}^2 are significantly larger (from 0.928 to 0.998) than the ones obtained from experiment and the Lipari−Szabo fit. The failure of the GAF model to correctly reproduce the order parameters of the UUCG hairpin clearly demonstrates that the motion of the base C−H dipole is not only caused by fluctuations of the base but is also due to the flexibility of the sugar ring and the backbone. For the UUCG hairpin under consideration, the GAF model on average accounts for about 20% of the fluctuations contributing to the order parameter.

Experimentally, the order parameters depend on the reorientation of the dipole and on the reorientation of the chemical shift anisotropy tensor. The latter contribution may affect the data on the aromatic carbon more than the dipole relaxation and is not easy to discriminate in practice. To support our analysis above, we have also calculated the order parameters and the internal correlation times of the N−H dipoles, by fitting the first 100 ps of the MD internal correlation function to eq 8. In Table 3, the calculated values

Table 3: Internal Correlation Times τ_e and Order Parameters of N_1/N_3 for the Guanine and Uracil Bases of the Hairpin[a]

	MD (100 ps fit)		experiment[21]		experiment[24]	
	τ_e (ps)	S^2	τ_e (ps)	S^2	τ_e (ps)	S^2
G1	3.7	0.890		0.74		
G2	2.5	0.907		0.787		0.938
U6	3.5	0.851	9.0	0.773		
U7	7.2	0.544				
G9	3.2	0.885	4.0	0.807		0.955
G10	2.0	0.917		0.776		0.907
U11	1.7	0.902		0.760		0.905
G12	1.7	0.918		0.780		0.905

[a] Reported values are derived from the MD simulation (by 100 ps fitting) and from the NMR experiment[21,24] (axially symmetric diffusion model except for G1).

for N_1-H_1 and N_3-H_3 dipoles of guanine and uracil bases are reported together with the available experimental data.[21,24] The N−H values fully agree with the values calculated for C−H dipole of the same base. Hence it may be expected that they characterize the identical base dynamics. Moreover, the calculated S^2 for N−H dipoles show a similar trend as in the experimental data, but they quantitatively agree only with one of the two available experimental sets.

3.4. Internal Dynamics of the UUCG Loop. In practice, the above studied GAF model is not used to calculate the order parameter but to rationalize the internal motions described by the experimental order parameter. Having validated our theoretical model in section 3.2, we are now in a position to use the 50 ns all-atom trajectory for this purpose. In what follows, we first characterize the main motions of the RNA hairpin. Then we analyze to what extent these motions are reflected in the calculated order parameter.

Principal component analysis (PCA) represents a standard method to identify the "principal" motions of a molecular system.[54−56] The approach represents the motion in terms of an orthogonal basis, the principal components (PCs), which are ordered according to their content of root-mean-square fluctuations (see Methods). For the 50 ns simulation of the UUCG loop, the first three PCs already contain 60% of the overall fluctuations of the hairpin, and to cover 80, 90, and 95% of the fluctuations, only 10, 25, and 50 out of 1442 PCs are required, respectively. In this sense, the first few PCs represent the main motions of the system. As an illustration, Figure 1 shows the motion along the first three PCs, which mainly consists of a conformational rearrangement involving the loop region. The presence of (at least) two conformational states is also confirmed by the analysis of the backbone dihedral angles of the loop residues U6−G9, see Table 4 in the Supporting Information. To assess the influence of this conformational rearrangement on the NMR order parameters, we have recalculated S^2 for the first and second half of the trajectory. Both 100 and 1000 ps Lipari−Szabo fits yield virtually identical results for S^2. The highest discrepancy is shown by the U7 base in the 1 ns fit (0.43 vs 0.50). This finding indicates that a simulation time of 50 ns is enough to investigate the C−H relaxation of the RNA loop and that its backbone rearrangement at 30 ns does not affect the order parameter values.

Table 4: Results of $1 - S^2$ Obtained for the Complete Trajectory (all) and by Including the First Three (PCA3) and the First Ten (PCA10) Principal Components, Respectively[a]

	$1 - S^2$ (100 ps fit)			$1 - S_{eq}^2$ (eq 10)		
sugar	all	PCA3	PCA10	all	PCA3	PCA10
U6	0.087	0.006	0.024	0.139	0.022	0.065
U7	0.182	0.053	0.089	0.626	0.511	0.578
C8	0.142	0.043	0.073	0.259	0.144	0.200
G9	0.131	0.033	0.056	0.258	0.140	0.186
base	all	PCA3	PCA10	all	PCA3	PCA10
U6	0.113	0.018	0.047	0.273	0.063	0.191
U7	0.429	0.082	0.232	0.847	0.195	0.680
C8	0.184	0.035	0.070	0.304	0.121	0.192
G9	0.114	0.013	0.035	0.162	0.033	0.074

[a] The order parameter are calculated either from a 100 ps Lipari−Szabo fit or from an equilibrium average (eq 10).

Let us now study to what extent the principal motions of the system account for the order parameters. To this end, we have expanded the MD trajectory in its first *n* PCs (see Methods) and calculated S^2 from this approximated trajectory, using a 100 ps Lipari−Szabo fit. Figure 6 in the Supporting Information shows the resulting order parameters as a function of the number of included PCs. As may be expected, the first PCs in general make the largest contribution to the decay of S^2. Compared to the rapid convergence of the overall fluctuations, however, the order parameters converge relatively slowly to their value obtained for the complete trajectory. Focusing on the loop residues U6−G9, Table 4 compares the results of $1 - S^2$ obtained for the complete trajectory (all), for the first three PCs (PCA3), and for the first ten PCs (PCA10). On average, the first three and 10 PCs yield about 20 and 40% of the value of $1 - S^2$ for the complete trajectory, respectively. Recalling that the first three and ten PCs contain 60 and 80% of the overall fluctuations, respectively, the order parameter apparently accounts only partially for the principal motions of the system. This is because the motion along the first few PCs may (i) be only weakly correlated with the orientation of the C−H dipoles and (ii) contains slow motion which is not seen by the Lipari−Szabo fit.

The latter issue can be studied by recalculating $1 - S_{eq}^2$ via an equilibrium average that covers all time scales of the trajectory. As shown in Table 4, in this case the first three and ten PCs on average yield about 40 and 70% of the value of $1 - S_{eq}^2$ for the complete trajectory, respectively, that is, quite similar to the values obtained for the overall fluctuations. In particular, we find that the S_{eq}^2 for the sugars U7, C8, and G9 to a large extent (80 and 55%) are caused by the motion along the first three PCs shown in Figure 1. Hence, if all time scales of the trajectory are taken into account, S_{eq}^2 are well described by the principal motions of the system.

4. Conclusions

A 50 ns MD simulation has been used to study the fast dynamics of the 14-mer UUCG hairpin. The comparison of

the simulation with the NMR relaxation data allows us to achieve several interesting results. First, the force field and the 50 ns simulation time used in this study appear appropriate to describe the fast relaxation of the C−H bonds monitored in the NMR experiment. Although we find slow conformation rearrangements of the RNA loop which may not be sampled appropriately, these motions do not influence the calculated NMR parameters. Internal motions on a time scale longer than 1 ns have been shown not to affect the calculated NMR relaxation rates and, as a consequence, will not be reflected in the order parameters.

Second, the overall and internal motions of the hairpin are found to be virtually independent, and the factorization approximation of the correlation function holds. Even in the case of the flexible U7, the approximation is not expected to change the calculated NMR data, since only the dynamics at short times ($\leq$1 ns) affect the relaxation parameters. Moreover, the internal correlation function can be replaced by a monoexponential function, fitted on its first 100 or 1000 ps, with no change of the relaxation parameters. The latter two observations demonstrate the validity of the Lipari−Szabo approach for the RNA hairpin.

Different ways to calculate the NMR order parameter have been compared. The best approach was found to be Lipari−Szabo fits of the first 100 ps (or 1 ns) of the MD internal correlation functions. The resulting order parameters S^2 and internal correlation times τ_e nicely agree with the experimental values. The calculation of order parameters via an equilibrium average was shown to deteriorate for residues undergoing slow internal dynamics, since the latter cannot be seen in NMR relaxation experiment limited by molecular tumbling. Although the equilibrium-average calculation of order parameters correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to experimental results. Finally the GAF model only yielded about 20% of the correct value for $1 - S^2$. This indicates that the motion of the base C−H dipoles not only maybe caused by fluctuations of the base but also are due to the flexibility of the sugar ring and the backbone, in particular for the loop residue C8.

The principal component analysis of the 50 ns trajectory has confirmed that a conformational rearrangement involving the loop region represents the main motion of the system. This principal motion, however, accounts only partially for the measured NMR order parameters S^2, because the latter are not sensitive to internal dynamics on a nanosecond time scale. Calculating the order parameter via an equilibrium average that covers all time scales of the trajectory, we obtain a direct correspondence between calculated S_{eq}^2 and principal motions. In particular, we find that the order parameters S_{eq}^2 for the sugars of residues U7, C8, and G9 are to a large extent caused by the motion along the first three principal components shown in Figure 1.

Acknowledgment. The authors wish to express their thanks to Elke Duchardt and Harald Schwalbe for numerous inspiring and helpful discussions and for sharing their results prior to publication. They also thank Jessica Koplin and Yuguang Mu for their contributions in the early stages of this work. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, and the Deutsche Forschungsgemeinschaft (SFB 579 "RNA-ligand interactions").

Supporting Information Available: Tables presenting the calculated and experimental mean values of the main dihedral angles of the 14-mer RNA hairpin and a figure reporting the order parameters for $C_{1'}$ and C_6/C_8 as a function of the number of included principal components. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wand, A. J. *Nature Struct. Biol.* **2001**, *8*, 926−931.

(2) Leulliot, N.; Varani, G. *Biochemistry* **2001**, *40*, 7947−7956.

(3) Al-Hashimi, H. M. *ChemBioChem* **2005**, *6*, 1506−1519.

(4) Persson, T.; Hartmann, R. K.; Eckstein, F. *ChemBioChem* **2002**, *3*, 1066−1071.

(5) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. *Structure* **2005**, *13*, 1255−1267.

(6) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. *Science* **2006**, *311*, 653−656.

(7) Frenkel, D.; Smit, B. *Understanding molecular simulations: from algorithms to applications*; Academic Press: London, 1996.

(8) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press: New York, 2004.

(9) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

(10) Brüschweiler, R.; Case, D. A. *Prog. Nucl. Magn. Reson. Spectrosc* **1994**, *26*, 27−58.

(11) Fischer, M. W. F.; Majumdar, A.; Zuiderweg, E. R. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *33*, 207−272.

(12) Korzhnev, D. M.; Billeter, M.; Arseniev, A. S.; Orekhov, V. Y. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 197−266.

(13) Levy, R. M.; Karlpus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998−6011.

(14) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289−2302.

(15) Palmer, A. G., III; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059−9067.

(16) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.

(17) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 7305−7313.

(18) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.

(19) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325−331.

(20) Lange, O. F.; Grubmüller, H.; de Groot, B. L. *Angew. Chem.* **2005**, *44*, 3394−3399.

(21) Akke, M.; Fiala, R.; Jiang, F.; Patel, D.; Palmer, A. G., III. *RNA* **1997**, *3*, 702−709.

(22) Vallurupalli, P.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 6893−6901.

(23) Showalter, S. A.; Baker, N. A.; Tang, C.; Hall, K. B. *J. Biomol. NMR* **2005**, *32*, 179−193.

(24) Duchardt, E.; Schwalbe, H. *J. Biomol. NMR* **2005**, *32*, 295−308.

(25) Chiarparin, E.; Rüdisser, S.; Bodenhausen, G. *ChemPhysChem* **2001**, *2*, 41−45.

(26) Cheatham, T., III; Kollman, P. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.

(27) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311−317.

(28) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **2000**, *8*, 227−236.

(29) Auffinger, P.; Westhof, E. *Biopolymers* **2001**, *56*, 266−274.

(30) Orozco, M.; Perez, A.; Noy, A.; Luque, F. *Chem. Soc. Rev.* **2003**, *32*, 350−364.

(31) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.

(32) Norberg, J.; Nilsson, L. *Acc. Chem. Res* **2002**, *35*, 465−472.

(33) Giudice, E.; Lavery, R. *Acc. Chem. Res* **2002**, *35*, 350−357.

(34) Woese, C.; Winker, S.; Gutell, R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467−8471.

(35) Williams, J.; Hall, K. *Biophys. J.* **1999**, *76*, 3192−3205.

(36) Williams, J.; Hall, K. *J. Mol. Biol.* **2000**, *297*, 1045−1061.

(37) Miller, J.; Kollman, P. *J. Mol. Biol.* **1997**, *270*, 436−450.

(38) Nina, M.; Simonson, T. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.

(39) Allain, F.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.

(40) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.

(41) Duchardt, E.; Richter, C.; Ohlenschläger, O.; Görlach, M.; Wöhnert, J.; Schwalbe, H. *J. Am. Chem. Soc.* **2004**, *126*, 1962−1970.

(42) Brüschweiler, R.; Wright, P. *J. Am. Chem. Soc.* **1994**, *116*, 8426−8427.

(43) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Comm.* **1995**, *91*, 43−56.

(44) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.

(45) Cheatham, T.; Cieplak, P.; Kollman, P. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−861.

(46) 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.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(48) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(50) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.

(51) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(52) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 6*; University of California: San Francisco, CA, 1999.

(53) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(54) Ichiye, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.

(55) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.

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

(57) Mark, P.; Nilsson, L. *J. Comput. Chem.* **2002**, *23*, 1211−1219.

(58) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(59) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1990**, *14*, 33−38.

CT600160Z