

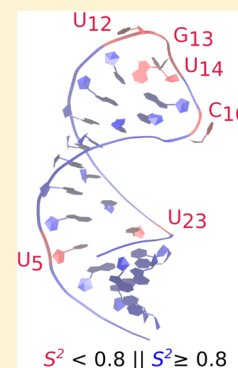
# Elucidating the Relation between Internal Motions and Dihedral Angles in an RNA Hairpin Using Molecular Dynamics

Alok Juneja, Alessandra Villa, and Lennart Nilsson\*

Department of Biosciences and Nutrition, Center of Biosciences, Karolinska Institutet, SE-141 83 Huddinge, Sweden

**S** Supporting Information

**ABSTRACT:** Molecular dynamics simulations were performed to characterize the internal motions of the ribonucleic acid apical stem loop of human hepatitis B virus. The NMR relaxation rates calculated directly from the trajectory are in good agreement with the experiment. Calculated order parameters follow the experimental pattern. Order parameters lower than 0.8 are observed for nucleotides that are weakly hydrogen bonded to their base pair partner, unpaired, or part of the loop. These residues show slow decay of the internal correlation functions of their base and sugar C–H vectors. Concerted motions around backbone dihedral angles influence the amplitude of motion of the sugar and base C–H vectors. The order parameters for base C–H vectors are also affected by the fluctuation of the glycosidic dihedral angle.



## INTRODUCTION

A large repertoire of cellular functions is associated with ribonucleic acid (RNA) systems. Despite the limited number of residue types, RNA molecules are capable of folding into a wide variety of secondary and tertiary structures. The variability in structure allows considerable diversity in biological function. Besides encoding and translational activity, RNA molecules are involved in enzymatic and regulatory functions. RNA functions cannot be explained solely based on characteristics of static structures, but dynamic properties are also of importance.<sup>1–3</sup> Dynamics accompanies and guides, for example, RNA cotranscriptional folding, ligand sensing and signaling, and site-specific catalysis in ribozymes. Moreover, RNA flexible regions, for example, loop regions, are usually the sites of interactions with small molecules, proteins, and other RNAs, playing a key role in cellular processes.<sup>4–7</sup>

RNA dynamics span a broad range of time scales from picoseconds, where vibrations and angular fluctuations occur, up to seconds, where catalytic function and global refolding take place. Experimentally, insight into the pico- to nanosecond motions can be obtained by NMR heteronuclear relaxation measurements.<sup>8</sup> In particular, <sup>13</sup>C relaxation experiments provide dynamics information for both nucleobase and ribose moieties in RNA systems. Usually dynamical quantities such as generalized order parameters ( $S^2$ ) and internal correlation times ( $\tau_c$ ) are extracted from the NMR relaxation rates following the model-free formalism.<sup>9</sup> Computationally, molecular dynamics (MD) simulation is an approach to gain access to RNA internal motions from the femto- to the nanosecond time scale.<sup>10–12</sup> Moreover, MD simulations directly provide information at the atomistic level on inter- and intramolecular motions using an atomistic empirical force field as the molecular model. The combination of the two techniques,

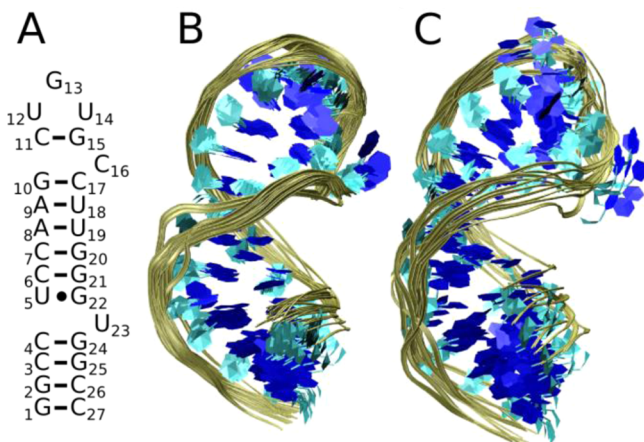
NMR spectroscopy and MD simulation, has been shown to provide a comprehensive description of the fast conformational motion in nucleic acid systems.<sup>13–17</sup>

Relaxation rates can be directly calculated from MD trajectories of a biomolecule, providing an unbiased comparison between simulation and experiments.<sup>17–20</sup> Various approaches have been successfully used to extract order parameters and correlation times from the MD trajectories. This allows biomolecular internal motions to be linked with the underlying order parameters. For proteins it has been shown, for example, that order parameters of the N–H spin pair are significantly affected by local fluctuations of the protein backbone  $\varphi$  dihedral angle.<sup>21</sup> In the case of an RNA hairpin, calculations of C–H order parameters from the fluctuations of the glycosidic dihedral (assuming a Gaussian distribution for the dihedral angle) lead to an overestimation (around 20%) of the values.<sup>17</sup> The authors interpreted the results as an indication that the base motion is not only affected by local fluctuations of the base but also by the flexibility of the sugar ring and backbone. Here, we want to go further and investigate in detail the relation between dihedral angles and C–H order parameters in an RNA hairpin. We used atomistic MD simulations of an RNA hairpin, for which <sup>13</sup>C relaxation measurements are experimentally available, to study in detail the internal motions and their relation with the calculated dynamical parameters (order parameters and correlation times). We aim to determine the influence of backbone and glycosidic dihedral angle fluctuations on the order parameters of C–H spins for sugar and/or base moieties and on their internal correlation times.

Received: March 10, 2014

Published: June 16, 2014

We focus on the highly conserved 27-mer apical loop of the encapsidation signal *epsilon* of human hepatitis B virus<sup>22</sup> (Figure 1), which is essential for viral replication *in vivo*.



**Figure 1.** Secondary and tertiary structures of the apical stem loop (27-nt) of human hepatitis B virus. (A) Secondary structure, (B) 23 overlaid structures from NMR (Protein Data Bank (PDB): 2IXY),<sup>23</sup> and (C) 10 overlaid structures from MD are shown. In the secondary structure diagram single lines (–) and filled circles (•) refer to Watson–Crick and wobble base pairs, respectively. Ten structures from MD are obtained by selecting a snapshot every 10 ns from the 100 ns production run. The tertiary structure figures were created with VMD.<sup>48</sup>

Hepatitis B virus is a deoxyribonucleic acid (DNA) virus, which replicates via reverse transcription of an RNA intermediate. The virus replication is triggered by the binding between the viral protein reverse transcriptase and *epsilon*. The secondary and tertiary structure of the apical loop have been determined by NMR experiments.<sup>23,24</sup> The hairpin contains two helical stems separated by a bulge residue (U23) and capped by a pseudotri-loop (C11–U12–G13–U14–G15–C16). Petzold and co-workers<sup>25</sup> studied in detail the internal dynamics of the hairpin by a combination of off-resonance ROESY and <sup>13</sup>C relaxation NMR experiments. The hairpin contains several mobile nucleotides, which are highly conserved among hepatitis B virus strains, making the hairpin a new example of regulatory RNAs where dynamical aspects are important for the biological function.<sup>4,5</sup> Interestingly, the residue motions in noncanonical structural elements are relatively slow (on the order of sub-nanoseconds), making the system a good test case to study the nature of the motions behind the observed NMR relaxations as expressed in the generalized order parameters and effective internal correlation times.

The apical loop is investigated at the atomistic level using MD simulations. <sup>13</sup>C relaxation rates for the sugar C1'–H1' and the pyrimidine C6–H6 and purine C8–H8 vectors have been calculated from the generated trajectory and compared with the available experimental values to validate the simulated dynamical properties of the system. Generalized order parameters and effective correlation times have been extracted from the simulation data in line with the Lipari and Szabo formalism<sup>9</sup> and discussed together with the values extracted from the experimental rates. We have looked at possible relations between the calculated dynamical parameters and the backbone and glycosidic dihedral angles. In particular, we have looked at the correlation between pairs of dihedral angles, and at dihedral angle fluctuations and their relation with sugar and/

or base generalized order parameters and effective correlation times.

## METHODS

**Simulation Protocol.** MD simulation was performed with the program CHARMM36<sup>26,27</sup> using the CHARMM36 nucleic acid force field<sup>28,29</sup> that includes an update of the 2'-hydroxyl torsion angle parameters.<sup>30</sup> The SHAKE<sup>31</sup> algorithm was applied to fix all covalent bonds containing a hydrogen atom allowing a 2 fs time step to be used in the integration of Newton's equations using the leapfrog algorithm. Fast table lookup routines for nonbonded interactions<sup>32</sup> were used when possible to increase the speed of the simulations.

The van der Waals interaction energies and forces were smoothly switched to zero from 8 to 12 Å, and the nonbonded list was constructed with a cutoff of 16 Å and was updated every time any atom moved by more than 2 Å since the last update. Long-range electrostatics was treated with the particle mesh Ewald method (PME),<sup>33,34</sup> using a grid of 0.8 Å, a  $\kappa$  value of 0.34, and a 12 Å cutoff for the real space part.

The MD simulations were performed under periodic boundary conditions at a constant pressure (1 atm). Constant pressure was maintained using the Langevin piston method<sup>35</sup> with piston mass 400 amu and collision frequency 20 ps<sup>−1</sup>. The average temperature of the system is checked every 2 ps and was found to remain within  $\pm 5$  K of 298 K after the heating MD run.

As a starting structure for simulation, we used the NMR structure model no. 1 of the apical loop (PDB: 2IXY).<sup>23</sup> The RNA system was solvated in a rhombic dodecahedron box with face-to-face distance of 62 Å, filled with pre-equilibrated TIP3P water molecules.<sup>36</sup> Water molecules with the oxygen atom less than 2.2 Å from any solute heavy atom were deleted. The negative charge of the nucleic acids was neutralized using Na<sup>+</sup> counterions, which were placed randomly.

The solvated system was energy minimized by 1000 steps of steepest descent with a harmonic restraint of 20 kcal/(mol/Å<sup>2</sup>) on RNA atoms and Na<sup>+</sup> ions allowing water molecules to adjust, followed by 3000 steps of the adopted-basis Newton–Raphson method without any harmonic restraint. The solvated system was heated to 298 K for 500 ps and then equilibrated for 1 ns, followed by a 100 ns production run. During the MD simulation, no harmonic restraints were imposed on the system and coordinates were saved every 1 ps.

**Relaxation Parameters.** NMR relaxation rates  $R_1$  (spin–lattice relaxation rate),  $R_2$  (spin–spin relaxation rate), and the nuclear Overhauser effect (NOE) of C–H spin vectors are computed from the simulated trajectory by calculating the spectral density function<sup>37,38</sup> from the correlation function<sup>18</sup> of the C–H bond using CHARMM. All of the constants in the expressions of the NMR relaxation rates were chosen in line with the experimental setup used by Petzold et al.<sup>25</sup>

Assuming that internal motions are much faster than the overall molecular tumbling of the molecule, the global correlation function can be separated into contributions from molecular tumbling (overall correlation function,  $C_{\text{overall}}(t)$ ) and internal motions (internal correlation function,  $C_{\text{int}}(t)$ ). The overall motion of the molecule is assumed to be isotropic with a rotational correlation time of 5.9 ns, the experimental value from Petzold et al.<sup>25</sup>

Overlaying the heavy atoms of each trajectory snapshot on a reference structure will eliminate the global motion (rotational

and translational) and allows calculation of the internal correlation function,  $C_{\text{int}}(t)$  for a C–H spin vector as

$$C_{\text{int}}(t) \equiv \langle P_2(\mu(t_0) \mu(t_0 + t)) \rangle \quad (1)$$

where  $\mu(t_0)$  is a time-dependent unit vector pointing along the C–H bond,  $P_2(x) \equiv (3x^2 - 1)/2$  is the second-order Legendre polynomial, and  $\langle \dots \rangle$  denotes a time average.

**NMR Dynamical Parameters.** The information on the internal motions contained in the NMR relaxation rate can be described by a generalized order parameter,  $S^2$ , which is the measure of the degree of spatial restriction of the motion and an effective correlation time,  $\tau_e$ .<sup>9</sup>

From the MD trajectory, the value of  $S^2$  is estimated by<sup>39</sup>

$$S^2 = \lim_{t \rightarrow \infty} C(t) = \frac{4\pi}{5} \sum_{m=-2}^{m=2} \langle Y_{2m}(\theta, \varphi) \rangle^2 \quad (2)$$

where  $Y_{2m}$  are the spherical harmonics of rank 2,  $\theta$  and  $\varphi$  are the polar angles defining the orientation of the dipole C–H vector at each snapshot of the trajectory,  $\mu$  are the Cartesian coordinates of the normalized C–H spin interaction vector, and  $\langle \dots \rangle$  denotes the average over all of the snapshots.

The effective correlation time for internal motions ( $\tau_e$ ), is given by

$$\tau_e = \frac{1}{1 - S^2} \int_0^{t_{\text{max}}} (C_{\text{int}}(t) - S^2) dt \quad (3)$$

where  $t_{\text{max}}$  is the time when  $C_{\text{int}}(t)$  reaches the plateau value ( $S^2$ ).

**Averaging and Error Analysis.** Only internal dynamics that is faster than the overall tumbling time,  $\tau_c$ , will directly affect the relaxation rates ( $R_1$ ,  $R_2$ , and NOE) and hence the dynamical parameters ( $S^2$  and  $\tau_e$ ). Slower processes that lead to a change of the environment, on the microsecond to millisecond time scale,<sup>40</sup> can also influence  $R_2$ , but our trajectory analysis does not include such effects on  $R_2$ . Experimentally slow exchange can be accounted for by an additional exchange term in the relation between  $R_2$  and the spectral density; in the apical loop NMR analysis exchange terms were employed for eight nucleotides, mainly for the C1' carbon.<sup>25</sup> Since we assume  $\tau_c = 5.9$  ns,<sup>25</sup> we break the production run of 100 ns into 20 windows of 5 ns and calculate the internal correlation functions for each window independently in order to mimic the experimental situation. The reported relaxation rates are averaged over all 20 windows.

To measure the variance in the relaxation rates of the nucleotide obtained from the MD simulation, we computed the relative standard deviation. For each nucleotide, the relative standard deviation of the relaxation rate  $x$  is defined as  $[(1/N) \sum_i^N (x_i - \bar{x})^2]^{1/2} / \bar{x}$ , where  $\bar{x} = (1/N) \sum_i^N x_i$  and  $N$  is the number of windows used to compute the average and standard deviation.

To quantify the agreement between relaxation rates from experiment and MD simulation, we computed the relative error. For the relaxation rate  $x$ , the relative (unsigned) error is computed over the total number of nucleotides  $M$  and is defined as  $\sum_i^M |x_i - x_i^{\text{exp}}| / \sum_i^M x_i^{\text{exp}}$ .

## RESULTS AND DISCUSSION

**Structural Features.** The apical stem loop (27-nt) of the encapsidation signal *epsilon* contains two helical stems separated by the bulged U23 nucleotide and pseudotriloop (C11-U12-G13-U14-G15-C16), which comprises of capping

loop region (U12-G13-U14) with C11-G15 closing base pair and C16 being bulged out (Figure 1). Most of the bases in the stems form Watson–Crick base pairs, with the hydrogen bonds present for at least 90% of the simulation. The nucleotide U5 in the upper stem is involved in hydrogen bonding with its wobble base pair partner G22 and with the unpaired bulged nucleotide U23 (14% and 45% of the time, respectively).

The root-mean-square deviation of the backbone heavy atoms after fitting to the reference structure (NMR model no. 1<sup>23</sup>) had an average value of  $2.9 \pm 0.4$  Å for the apical stem loop and  $2.9 \pm 0.3$  Å for the loop region (U12-G13-U14). By visual inspection of 10 overlaid structures from MD simulations (obtained by picking a structure every 10 ns) and 23 structures from NMR studies,<sup>23</sup> it was ascertained that the backbone of the simulated structures acquired similar conformations as the experimental structures; the conformation of the loop region, the unpaired U16, and the end residues (G1 and C27) show more variability in the simulated structures (Figure 1). In the experiment, the data do not lead to a single well-defined conformation for the triloop but to a pool of conformations, characterized by major–minor–major geometry. In the simulations, loop residues G13 and U14 point to the minor and major grooves, respectively, in agreement with NMR conformations, while U12 oscillates between the major and minor grooves. The U5 is involved in hydrogen bonding either with G22 or U23 and, indeed, an imino signal is observed for U5 and no signal for G22 and U23 in the NMR experiment.

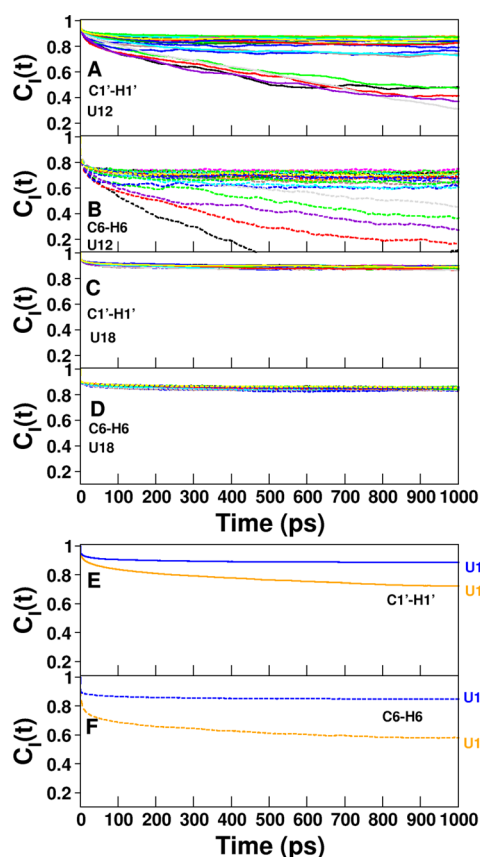
Both global and local structural features in RNA are sensitive to details of the force field, and there have been recent reparametrizations of the AMBER<sup>41–43</sup> and CHARMM<sup>30,44</sup> force fields that address specific aspects of their behavior. To judge the overall structural stability of our simulations, we monitored the root-mean-square deviation (RMSD) for the sugar–phosphate backbone of the whole stem loop as well as for just the loop residues 12–14. In both cases the RMSD is quite stable throughout the simulation (Supporting Information Figure S0), showing that overall the simulations are well-behaved with no indication of denaturation, in agreement also with other recent haripin simulations performed with the CHARMM36 force field.<sup>30</sup>

**Relaxation Rates.** From the MD simulation data, we computed <sup>13</sup>C longitudinal ( $R_1$ ) and transversal ( $R_2$ ) relaxation rates and NOEs for the sugar C1'–H1', the pyrimidine C6–H6, and the purine C8–H8 vectors of the apical stem loop (Table S1 in the Supporting Information). The relative standard deviation is around  $\sim 0.07$ ,  $\sim 0.03$ , and  $\sim 0.09$  for  $R_1$ ,  $R_2$ , and NOE, respectively, for the majority of the residues. Nucleotides with higher deviations are the wobble base paired nucleotide U5 (hydrogen bonded partially to its wobble base pair partner G22 and to the unpaired bulged nucleotide U23), the loop region nucleotides (U12-G13-U14), and the unpaired bulged nucleotides (C16 and U23). The calculated relaxation parameters from the simulations are in good agreement with experimental values (Figure S1 in the Supporting Information).<sup>25</sup> The relative error is 0.13 for  $R_1$ , 0.19 for  $R_2$ , and 0.09 for NOE. We observed some outliers. The worst estimation, we observed, is for  $R_2$  of C1'–H1' of residue C11. The difference (around  $20 \text{ s}^{-1}$ ) between calculated and experimental values is due to a slow exchange term that is not present in the simulation analysis (as discussed in Methods), but which is accounted for in the experiment; for C11 the exchange term contribution to  $R_2$  is  $15 \text{ s}^{-1}$ .<sup>25</sup>



The agreement between the directly calculated MD simulation data and the NMR relaxation data shows that the force field and simulation time scale used in MD simulation are appropriate to describe the relaxation of the C–H vectors monitored in NMR. Thus, the MD simulation can be employed, first to calculate the dynamical parameters of C–H vectors and then to reveal the motions underlying these dynamical parameters.

**Generalized Order Parameters and Effective Correlation Times.** We observed that  $C_{\text{int}}(t)$  of C–H vectors, after a rapid initial decay, generally reaches a plateau and does not decay to zero. The initial decay is fast compared with the overall tumbling rate. Most of the functions are characterized by a single-phase decay behavior on a time scale of 10–50 ps; some have a multiphase decay behavior on a pico- and nanosecond time scale (Figure 2). The plateau value, which is



**Figure 2.** Internal correlation functions of the sugar C1'–H1' and base C6–H6 vectors of residues U12 and U18. Internal correlation functions are obtained from  $20 \times 5$  ns (A–D) and 100 ns (E, F) MD simulations. Internal correlation functions are shown for the first 1 ns.

the generalized order parameter ( $S^2$ ), is due to the spatial restriction of C–H vector motions, caused by intramolecular interactions, and it can be calculated using eq 2 (Table S2 in Supporting Information).  $S^2$  can adopt values between 0 and 1, with 1 signifying total rigidity, while a value of 0 indicates a totally unrestricted motion of the C–H vector.

We have calculated the C–H internal correlation functions both from the full 100 ns trajectory and from 5 ns windows (for a total of 20 windows). In Figure 2 we show as an example the correlation functions of U12 and U18 C–H vectors. In the case of U18 all of the calculated functions have a similar shape, with

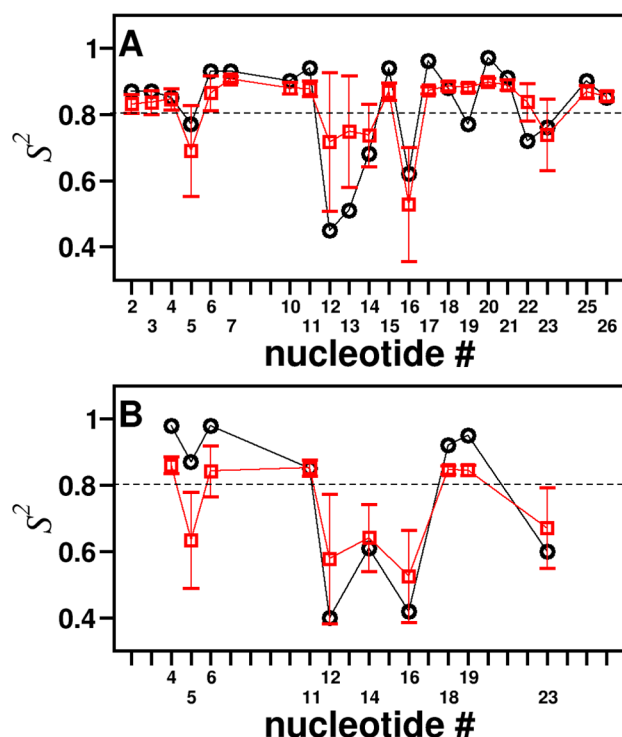
an initial fast (<20 ps) decay followed by a plateau at a value of >0.85. For U12, the picture is more complex; the correlation function calculated on the whole trajectory has a fast decay (<50 ps) followed by a slow decay (around 500 ps), reaching a plateau value between 0.8 and 0.6, while the correlation functions calculated from the 5 ns windows can be divided into three groups according to their decay behavior and plateau value. In most cases, which constitute the first group, there is just a fast decay on a time scale of <50 ps (a plateau value at >0.8), in the second group there is also a single, but slower decay on a time scale of 100–200 ps (a plateau value at around 0.8), and the last group has a multitime scale decay with a fast decay (around 100 ps) and a slow decay (>1 ns) (a plateau value at <0.6). The correlation functions in the third group are obtained in windows where several conformational transition events take place, for example, the windows between 0 and 15 ns for U12 (Figure 7). Note that the correlation functions are not changing systematically along the trajectory, and there are no indications of denaturation during the simulation, as also shown by the values of the order parameters as a function of time in Figure 4.

On one hand, the results support our approach to represent the effect of the overall molecular rotation by dividing the trajectory into several windows of time length comparable with the molecular tumbling, and calculating the C–H order parameters as an average over the values obtained in the individual windows. On the other hand, the results also show how complex the dynamical picture and the shape of the correlation functions could be for vectors that are involved in diverse motions with different time scales. While the C–H autocorrelations of U18 can be described using a mono-exponential function, this is not valid for the U12 autocorrelation functions.

In general, the calculated ( $S^2_{\text{MD}}$ ) and experimental<sup>25</sup> ( $S^2_{\text{exp}}$ ) order parameters for C1'–H1' and C6–H6 show a similar variation along the nucleotide sequence (Figure 3). Nucleotides involved in a base pair have  $S^2_{\text{MD}} > 0.8$ , while nucleotides with  $S^2_{\text{MD}} < 0.8$  are only partially hydrogen bonded to their base pair partner, as in the case of the wobble base pair nucleotide (U5), or they are unpaired (U12, G13, U14, C16, and U23). These residues are the ones which were found to have higher relative standard deviations in their relaxation rates (Supporting Information Table S1), as reported above. We found that the root-mean-square positional fluctuations of these nucleotides ( $\sim 1.7$  Å) is always higher than for the base paired nucleotides (<1 Å), which is consistent with these nucleotides being more flexible, with enhanced motions of their C–H vectors.

Now we look closely at how the  $S^2_{\text{MD}}$  of C–H vectors are changing during the MD simulation (Figure 4). In general, the C–H order parameters of bases and sugars have very similar values; that is, the motions of the sugar and base vectors have the same amplitude. Exceptions to this are residues U12, G13, and C16, where the motions of C1'–H1' vectors are more restricted than those of the base C–H vectors most of the time. Only in a few windows is the order parameter for the sugar lower than that of the base.

We noticed that  $0.72 < S^2_{\text{MD}} < 0.92$  for the C–H vectors of bases and sugars in the base paired residues in the stems. This indicates that the C–H vector motions in these residues are mostly restricted in the simulation. For residues U5, U12, G13, U14, C16, and U23 with the average value of  $S^2_{\text{MD}} < 0.8$ , the order parameters are spread over a broader range, from 0.1 to 0.8, indicative of large amplitude motions. For example in case



**Figure 3.** Generalized order parameter  $S^2$  for C1'–H1' (A) and C6–H6 (B) vectors as a function of the nucleotide number. The experimental values<sup>25</sup> are shown as black circles. Calculated values from MD are shown as red squares, with the bars showing their standard deviations. The numbers on the x-axis are the nucleotides for which experimental  $S^2$  values are available.<sup>25</sup>

of the bulged residue C16,  $S^2_{\text{MD}}$  is  $<0.4$  in seven windows (2, 14, 15, 17, 18, 19, and 20).

Calculating the effective internal correlation times using eq 3 on the same set of 5 ns windows as for the order parameters did not give as satisfactory an agreement with the experimentally reported correlation times with the calculated values being both smaller and larger than the experimental values (Table S3 in the Supporting Information). There are errors in the calculations as well as in the experiments, and we have identified some of these which are of a systematic character. The calculated  $\tau_e$  is sensitive to noise in  $C_{\text{int}}(t)$  in that the first zero crossing of  $C_{\text{int}}(t)$  is used as the upper limit for the integration in eq 3, and the integral also depends on the value of  $S^2$  (Figure S2 in the Supporting Information). In the experiment  $\tau_e$  is not well-defined from the observed  $R_1$ ,  $R_2$ , and NOE values in all ranges. As can be seen from the relations between ( $S^2$  and  $\tau_e$ ) and ( $R_1$ ,  $R_2$ , and NOE), one set of observed ( $R_1$ ,  $R_2$ , and NOE) may provide a robust estimate of  $S^2$ , while still leaving  $\tau_e$  largely undetermined; this is particularly true for the short correlation times and larger order parameters (Figure 5). As an example, for the C1'–H1' vector ( $R_1 = 1$ ,  $R_2 = 19$ , and NOE = 1.3) defines  $S^2 = 0.7$  relatively well, whereas  $\tau_e$  could be anywhere in the range of 1–50 ps, without requiring a corresponding change (error) in ( $R_1$ ,  $R_2$ , and NOE).

The uncertainty in the calculated effective correlation times is further indicated by their standard deviations, which are of the same order of magnitude as the average effective correlation time. Therefore, we decided to consider the internal correlation function of C–H vectors from the full 100 ns simulation to explore if the decay of the internal correlation function shows

any pattern similar to the internal correlation times reported in experiment.<sup>25</sup>

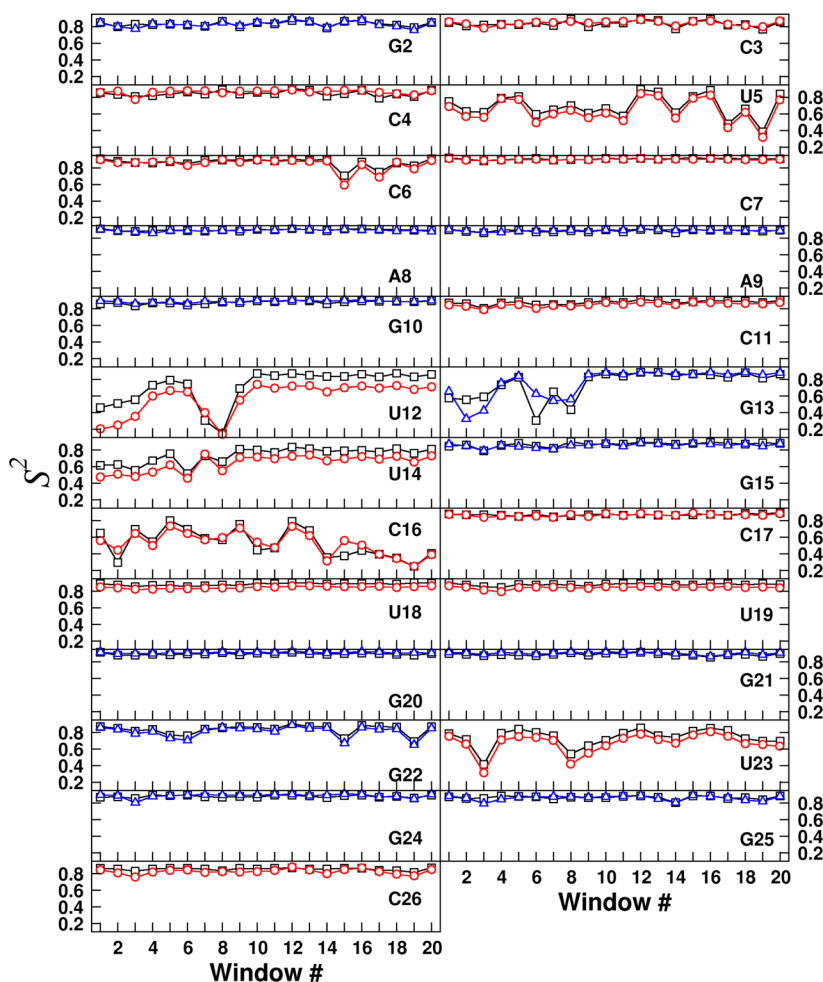
Considering the sugar C1'–H1' vector, we clearly notice two groups based on the decay of the internal correlation functions (Figure 6A). Residues U5, U12, G13, U14, C16, and U23 showed a fast decay followed by a slow decay (Figure 6A). In the experiment, these residues were reported to have slow sugar C1'–H1' vector motions with effective correlation times that range between 300 and 600 ps.<sup>25</sup> For pyrimidine base C6–H6 and purine base C8–H8 vectors, we noticed the existence of two groups based on the decay of the internal correlation functions (Figure 6B). In line with what we observed for the C1'–H1' vector of sugar, similar residues U5, U12, G13, U14, C16, and U23 were found to have a slow decay of the internal correlation functions of their base vectors (Figure 6B) in agreement with the available experiment values. Experiment reports slower motions of the C6–H6 vectors (between 300 and 600 ps) for the pyrimidine bases U5, U12, U14, and U23.<sup>25</sup> In addition to the residues discussed previously, residue G22 also shows a slow decay of the internal correlation function for its C8–H8 vector. Weak hydrogen bonding in the wobble base pair U5–G22 (14% of the time) might be identified as the source of the slow decay of the internal correlation function of the base C8–H8 vector of residue G22. Unluckily, no experimental data are available for the C8–H8 vector.

**Influence of Dihedral Angles on Generalized Order Parameters.** The overall structural stability of the system (Supporting Information Figure S0) is also reflected in the generalized order parameters, as well as in the dihedral angles, which show no long-term trends along the trajectory (Figure 4, Figure 7 and Supporting Information Figures S0 and S4–S29).

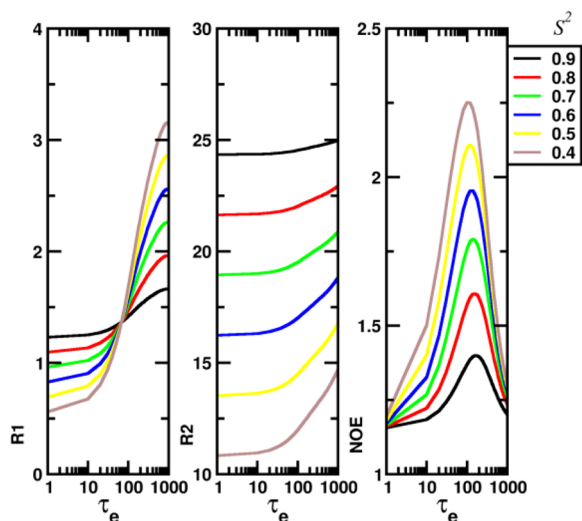
To characterize the motions underlying the order parameters of C–H vectors, we have looked for each nucleotide at the correlation of the motions around backbone and glycosidic dihedral angles along the trajectory and at their relation with the  $S^2_{\text{MD}}$  of C–H vectors of the corresponding nucleotide. To illustrate this relation, we pick residue U12, which has C–H vectors, characterized by  $S^2_{\text{MD}} < 0.8$  (Figure 3). We found two scenarios: one under the first 15 ns and at 30–40 ns and the other between 15 and 30 ns and after 40 ns.

During the first 15 ns several conformational exchange events occur involving the dihedral angles  $\alpha$  (P–O<sub>5'</sub>),  $\beta$  (O<sub>5'</sub>–C<sub>5'</sub>),  $\epsilon$  (C<sub>3'</sub>–O<sub>3'</sub>), and  $\zeta$  (O<sub>3'</sub>–P) (Figure 7). The C–H order parameters are  $<0.6$ , and the motions around these dihedral angles are correlated. In particular, a positive correlation coefficient is observed for dihedral pairs  $\alpha$ – $\beta$  ( $r \sim 0.45$ ),  $\beta$ – $\epsilon$  ( $r \sim 0.22$ ), and  $\epsilon$ – $\zeta$  ( $r \sim 0.21$ ). This might be identified as the source of less restricted motions of C–H vectors. Interestingly, a statistical analysis of the helical regions of yeast tRNA<sup>Phe</sup> have revealed a linear relation in these pairs of dihedral angles with a negative correlation coefficient.<sup>45</sup> We observed a similar situation between 30 and 40 ns, where several conformational rearrangements take place involving the backbone dihedrals  $\alpha$ ,  $\beta$ ,  $\epsilon$ , and  $\zeta$ . There, the order parameters drop to  $\sim 0.31$  (for sugar C1'–H1' vector) and to  $\sim 0.40$  (base C6–H6 vector) and a positive correlation between the motions of  $\alpha$  and  $\beta$  is observed (and no correlation for  $\beta$ – $\epsilon$  and  $\epsilon$ – $\zeta$  pairs).

Between 15 and 30 ns, the scenario is different. All of the dihedrals retain their achieved conformations, and  $S^2_{\text{MD}}$  increases to  $\sim 0.76$  for sugar C1'–H1' vector and  $\sim 0.64$  for base C6–H6 vector. Considering the motions around the dihedral angles, we found negative correlation for the dihedrals  $\alpha$ – $\gamma$  ( $r \sim -0.73$ ) and  $\beta$ – $\epsilon$  ( $r \sim -0.40$ ) and positive correlation



**Figure 4.** Generalized order parameters from MD simulation ( $S^2_{MD}$ ) for C–H vectors. The reported  $S^2$  values are calculated from each of the 5 ns window for C1'–H1' (black  $\square$ ), pyrimidine base C6–H6 (red  $\circ$ ), and purine base C8–H8 (blue  $\Delta$ ) vectors.



**Figure 5.** Relaxation parameters ( $R_1$  ( $s^{-1}$ ),  $R_2$  ( $s^{-1}$ ), and NOE) of the sugar C1'–H1' vector as a function of order parameter ( $S^2$ ) and internal correlation time ( $\tau_e$ ).

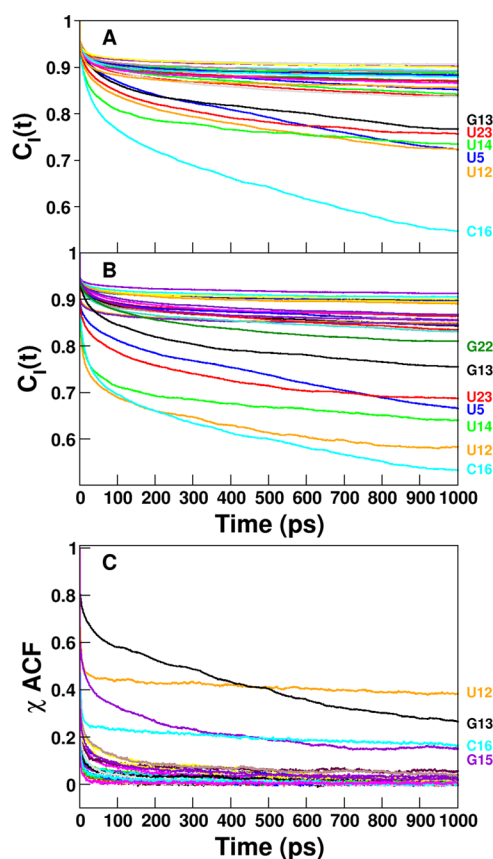
for dihedrals  $\chi$ – $\delta$  ( $r \sim 0.21$ ) and  $\epsilon$ – $\delta$  ( $r \sim 0.44$ ). A similar situation is observed during the last 60 ns of the simulation, where  $S^2_{MD}$  is  $\sim 0.85$  for the sugar C1'–H1' vector and  $\sim 0.71$  for the base C6–H6 vector. Here, a negative correlation

between the motions around dihedrals in the pairs  $\alpha$ – $\beta$  ( $r \sim -0.45$ ),  $\alpha$ – $\gamma$  ( $r \sim -0.78$ ) and  $\epsilon$ – $\zeta$  ( $r \sim -0.32$ ) was observed, while a positive correlation was seen in the case of  $\delta$ – $\chi$  ( $r \sim 0.25$ ) and  $\chi$ –P ( $r \sim 0.34$ ). The observed negatively correlated motion between the backbone dihedrals might result in motions of restricted amplitude for C–H vectors ( $S^2 > 0.6$ ) since the opposing motions of the two dihedrals will tend to cancel longer range structural effects. Interestingly correlations between dihedral angles observed in these two time windows are in qualitative agreement with statistical studies performed on nucleic acid structures.<sup>45–47</sup>

The performed analysis clearly indicates that concerted motions around the backbone dihedrals influence the amplitude of the motions of sugar and base C–H vectors, and thus the corresponding generalized order parameters. Then, we have moved on and look more in detail at the motions (amplitude and time scale) underlying the base order parameters. To achieve that, we have looked at the relation between glycosidic dihedral angle and the base generalized order parameters.

The glycosidic dihedral angle,  $\chi$  (purine, C<sub>1'</sub>–N<sub>1</sub>; or pyrimidine, C<sub>1'</sub>–N<sub>1</sub>), is the only dihedral that is connected to the base. To see if the motion around it influences the order parameters of the base C–H vectors, we performed linear regression fitting between the fluctuation around  $\chi$  and  $S^2_{MD}$  of base C–H vector. Fluctuations (standard deviation) in  $\chi$  were computed from the 100 ns production run. We observed that



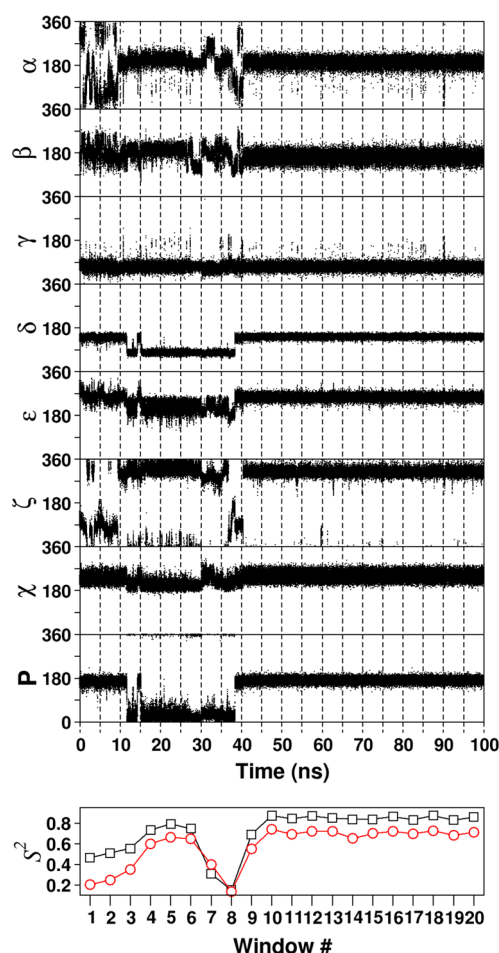


**Figure 6.** Internal correlation functions of the (A) sugar C1'–H1' and (B) base C6–H6/C8–H8 vectors of the residues obtained from the 100 ns MD simulation. (C) Autocorrelation functions (ACF) of the glycosidic dihedral angles,  $\chi$ , obtained from 100 ns MD simulation. Both the internal correlation function and  $\chi$  ACF are shown for the first 1 ns. Internal correlation functions and  $\chi$  ACFs with slower decay are labeled with the corresponding residue name.

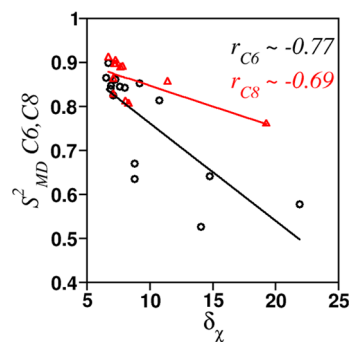
the order parameters of the pyrimidine (C6–H6) and purine (C8–H8) vectors form different groups, and therefore we performed linear regression fitting on their order parameters separately. Linear regression analysis dihedral fluctuations and the corresponding order parameters resulted in a negative correlation ( $r \sim -0.77$  for C6–H6 and  $r \sim -0.69$  for C8–H8) (Figure 8). The observed negative correlation suggests that the increase in the amplitude of motions of base C–H vectors ( $S^2$  value decreases) corresponds to an increase in the fluctuation around  $\chi$  and vice versa. To understand if there is a relation between the decay of the base C–H motion and the glycosidic dihedral angle, we have calculated the ACF of the motions around the glycosidic  $\chi$  (Figure 6C). Based on the decays of  $\chi$  ACFs, we can categorize the residues in two groups. Residues U12, G13, G15, and C16 belong to the group which has slow decay of  $\chi$  ACF, while the rest of the residues belong to another group where the  $\chi$  ACF has only one fast decay (<50 ps) and reaches a zero plateau value (Figure 6C). We found that the unpaired residues that were either part of the loop (U12 and G13) or bulged (C16) are the ones where the  $\chi$  ACF decay can be correlated to the slow decay of the internal correlation function of their base C–H vectors (Figure 6).

## CONCLUSION

We report the characterization of internal motions within the apical stem loop (27-nt) of *epsilon* of human hepatitis B-virus



**Figure 7.** Time series of the motions around dihedral angles and the pucker phase angle along with the order parameters obtained from each of the 5 ns windows of the MD simulation of residue U12. The reported  $S^2$  values are from sugar C1'–H1' (black  $\square$ ) and pyrimidine base C6–H6 (red  $\circ$ ) vectors. Backbone and glycosidic dihedral angles are defined in Figure S3 of the Supporting Information. Similar graphs for the rest of the residues are shown in Figures S4–S29 of the Supporting Information.



**Figure 8.** Scatter plot of the order parameters  $S^2_{MD}$  of C6–H6 (black  $\circ$ )/C8–H8 (red  $\Delta$ ) vectors and the fluctuations around the glycosidic dihedral angle,  $\chi$ . The black line is the linear regression line with correlation coefficient  $r \sim -0.77$  obtained from the order parameters of C6–H6, and the red line is the linear regression line with correlation coefficient  $r \sim -0.69$  obtained from the order parameters of C8–H8.

employing MD simulations. Although nuclear magnetic resonance can provide detailed information on amplitudes

and time scales of internal motions, it is not possible to describe the molecular nature of the underlying motions. In contrast, classical molecular dynamics simulations can generate models of internal motions that can be used to characterize the structural dynamics of biomolecules and their time scales. We used 100 ns atomistic simulation to calculate  $^{13}\text{C}$  relaxation rates that yield good agreement with experiment. This indicates that the force field and time scale used in our simulations are appropriate to describe the relaxation of the C–H vectors monitored in the NMR experiment.<sup>25</sup>

Order parameters calculated from the MD simulations have a pattern along the nucleotide sequence of the hairpin similar to the ones reported experimentally. For nucleotides with little flexibility ( $S^2 > 0.8$ ) the amplitude of motions of sugar and base C–H vectors in MD simulation are very similar, and nucleotides that are weakly hydrogen bonded to its base pair partner, being part of the loop or unpaired, are flexible and have lower order parameters of their base and sugar C–H vectors. These residues have been found to have higher root-mean-square positional fluctuations than the base paired nucleotides, which indicates that these nucleotides are more flexible and have enhanced motions of their C–H vectors.

The internal correlation functions of C–H vectors calculated from MD simulation are found to be the good indicators of the time scales of internal motions of C–H vectors reported in experiment. A slow decay of the internal correlation function of the sugar and base C–H vectors is observed for those residues which are weakly hydrogen bonded to their base pair partner; being part of the loop or unpaired is flexible. In experiment these residues have been reported to have slow internal motions of their sugar and base C–H vectors. For a few of these residues, the slow decay of the autocorrelation functions of glycosidic dihedral angles also captures the slow internal correlation times of the base C–H vectors.

Can the local motions within a nucleotide be related to the amplitude of motions of its C–H vectors? The conformational transitions lead to the lowering of the order parameter of C–H vectors, and in particular, the nonconcerted motions around dihedrals result in a lowering of the order parameters, while negatively correlated motions of dihedrals lead to a restricted amplitude of motions of C–H vectors. The analysis of the motion around the glycosidic dihedral angle,  $\chi$ , indicates that the fluctuation of  $\chi$  influences the amplitude of motion of the corresponding base C–H vectors, in agreement with a previous study on a smaller RNA hairpin.<sup>17</sup>

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

Tables listing relaxation parameters, order parameters, and average internal correlation times obtained from MD simulation and figures showing RMSD for the sugar–phosphate backbone heavy atoms of the whole stem loop and loop residues, internal correlation functions of the sugar C1'–H1' vectors of residues U12 and C16, torsion angles in a nucleoside, and individual time series of the motions around dihedral angles and the pucker phase angle along with the order parameters of all the nucleotides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +46-8-524 81099. E-mail: [Lennart.Nilsson@ki.se](mailto:Lennart.Nilsson@ki.se).

## Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) Hall, K. B. RNA in motion. *Curr. Opin. Chem. Biol.* **2008**, *12*, 612–618.
- (2) Al-Hashimi, H. M.; Walter, N. G. RNA dynamics: It is about time. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321–329.
- (3) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. Resolving the Motional Modes that Code for RNA Adaptation. *Science* **2006**, *311*, 653–656.
- (4) Al-Hashimi, H. M.; Gosser, Y.; Gorin, A.; Hu, W. D.; Majumdar, A.; Patel, D. J. Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.* **2002**, *315*, 95–102.
- (5) Hoogstraten, C. G.; Wank, J. R.; Pardi, A. Active site dynamics in the lead-dependent ribozyme. *Biochemistry* **2000**, *39*, 9951–9958.
- (6) Leulliot, N.; Varani, G. Current Topics in RNA–Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture†. *Biochemistry* **2001**, *40*, 7947–7956.
- (7) Rau, M.; Stump, W. T.; Hall, K. B. Intrinsic flexibility of snRNA hairpin loops facilitates protein binding. *RNA* **2012**, *18*, 1984–1995.
- (8) Shajani, Z.; Varani, G. NMR studies of dynamics in RNA and DNA by C-13 relaxation. *Biopolymers* **2007**, *86*, 348–359.
- (9) Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- (10) Hashem, Y.; Auffinger, P. A short guide for molecular dynamics simulations of RNA systems. *Methods* **2009**, *47*, 187–197.
- (11) Norberg, J.; Nilsson, L. Molecular Dynamics Applied to Nucleic Acids. *Acc. Chem. Res.* **2002**, *35*, 465–472.
- (12) Orozco, M.; Noy, A.; Perez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185–193.
- (13) Duchardt, E.; Nilsson, L.; Schleucher, J. Cytosine ribose flexibility in DNA: A combined NMR ( $^{13}\text{C}$ ) spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res.* **2008**, *36*, 4211–4219.
- (14) Ferner, J.; Villa, A.; Duchardt, E.; Widjakakusuma, E.; Wöhnert, J.; Stock, G.; Schwalbe, H. NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* **2008**, *36*, 1928–1940.
- (15) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. Structure and dynamics of an RNA tetraloop: A joint molecular dynamics and NMR study. *Structure* **2005**, *13*, 1255–1267.
- (16) Showalter, S. A.; Baker, N. A.; Tang, C. G.; Hall, K. Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J. Biomol. NMR* **2005**, *32*, 179–193.
- (17) Villa, A.; Stock, G. What NMR relaxation can tell us about the internal motion of an RNA hairpin: A molecular dynamics simulation study. *J. Chem. Theory Comput.* **2006**, *2*, 1228–1236.
- (18) Levy, R. M.; Karplus, M.; Wolynes, P. G. NMR relaxation parameters in molecules with internal motion: Exact Langevin trajectory results compared with simplified relaxation models. *J. Am. Chem. Soc.* **1981**, *103*, 5998–6011.
- (19) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J. Am. Chem. Soc.* **1992**, *114*, 2289–2302.
- (20) Case, D. A. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* **2002**, *35*, 325–331.
- (21) Brüschweiler, R.; Wright, P. E. NMR order parameter of biomolecules: A new analytical representation and application to the Gaussian axial fluctuation model. *J. Am. Chem. Soc.* **1994**, *116*, 8426–8427.
- (22) Ganem, D.; Varmus, H. E. The Molecular Biology of the Hepatitis B Viruses. *Annu. Rev. Biochem.* **1987**, *56*, 651–693.



- (23) Flodell, S.; Petersen, M.; Girard, F.; Zdunek, J.; Kidd-Ljunggren, K.; Schleucher, J.; Wijmenga, S. Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.* **2006**, *34*, 4449–4457.
- (24) Flodell, S.; Schleucher, J.; Cromsigt, J.; Ippel, H.; Kidd-Ljunggren, K.; Wijmenga, S. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* **2002**, *30*, 4803–4811.
- (25) Petzold, K.; Duchardt, E.; Flodell, S.; Larsson, G.; Kidd-Ljunggren, K.; Wijmenga, S.; Schleucher, J. Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res.* **2007**, *35*, 6854–6861.
- (26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (27) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (28) Foloppe, N.; MacKerell, J. A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (29) MacKerell, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (30) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929–1943.
- (31) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (32) Nilsson, L. Efficient table lookup without inverse square roots for calculation of pair wise atomic interactions in classical simulations. *J. Comput. Chem.* **2009**, *30*, 1490–1498.
- (33) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (34) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (35) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (37) *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J., Rance, M., Eds.; Elsevier Academic Press: Burlington, MA, USA, 2006.
- (38) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. *J. Am. Chem. Soc.* **1984**, *106*, 1923–1930.
- (39) Akke, M.; Brueschweiler, R.; Palmer, A. G. NMR order parameters and free energy: An analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k. *J. Am. Chem. Soc.* **1993**, *115*, 9832–9833.
- (40) Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **2004**, *104*, 3623–3640.
- (41) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. Benchmarking AMBER Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised  $\chi$  Torsions. *J. Phys. Chem. B* **2011**, *115*, 9261–9270.
- (42) Zgarbová, M.; Otyepka, M.; Šponer, J. i.; Mládek, A. t.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- (43) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820–16825.
- (44) Denning, E. J.; MacKerell, A. D. Intrinsic Contribution of the 2'-Hydroxyl to RNA Conformational Heterogeneity. *J. Am. Chem. Soc.* **2012**, *134*, 2800–2806.
- (45) Kitamura, K.; Mizuno, H.; Amisaki, T.; Tomita, K.-I.; Baba, Y. Locally oscillatory motion of RNA helix derived from linear relationships of backbone torsion angles. *Biopolymers* **1984**, *23*, 1169–1184.
- (46) Kitamura, K.; Wakahara, A.; Mizuno, H.; Baba, Y.; Tomita, K. Conformationally 'concerted' changes in nucleotide structures. A new description using circular correlation and regression analyses. *J. Am. Chem. Soc.* **1981**, *103*, 3899–3904.
- (47) Schneider, B.; Neidle, S.; Berman, H. M. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers* **1997**, *42*, 113–124.
- (48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.