THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Elucidating Molecular Motion through Structural and Dynamic Filters of Energy-Minimized Conformer Ensembles

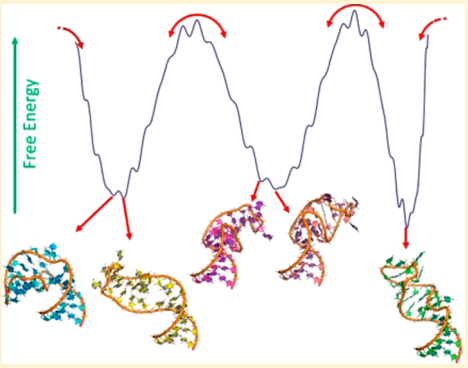
Prashant S. Emani,[†] Michael F. Bardaro, Jr.,[†] Wei Huang,[†,∥] Sergio Aragon,[§] Gabriele Varani,[†,‡] and Gary P. Drobny*[,†]

[†]Department of Chemistry, University of Washington, Box 351700, Seattle, Washington 98195-1700, United States

[‡]Department of Biochemistry, University of Washington, Box 357350, Seattle, Washington 98195-7350, United States

[§]Department of Chemistry and Biochemistry, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, United States

ABSTRACT: Complex RNA structures are constructed from helical segments connected by flexible loops that move spontaneously and in response to binding of small molecule ligands and proteins. Understanding the conformational variability of RNA requires the characterization of the coupled time evolution of interconnected flexible domains. To elucidate the collective molecular motions and explore the conformational landscape of the HIV-1 TAR RNA, we describe a new methodology that utilizes energy-minimized structures generated by the program "Fragment Assembly of RNA with Full-Atom Refinement (FARFAR)". We apply structural filters in the form of experimental residual dipolar couplings (RDCs) to select a subset of discrete energy-minimized conformers and carry out principal component analyses (PCA) to corroborate the choice of the filtered subset. We use this subset of structures to calculate solution T_1 and $T_{1\rho}$ relaxation times for ^{13}C spins in multiple residues in different domains of the molecule using two simulation protocols that we previously published. We match the experimental T_1 times to within 2% and the $T_{1\rho}$ times to within less than 10% for helical residues. These results introduce a protocol to construct viable dynamic trajectories for RNA molecules that accord well with experimental NMR data and support the notion that the motions of the helical portions of this small RNA can be described by a relatively small number of discrete conformations exchanging over time scales longer than 1 μs.



1. INTRODUCTION

It is now widely recognized that many RNA molecules are predisposed to forming complexes with proteins by fluctuating spontaneously through an ensemble of structural states. This dynamic mode of RNA−protein recognition is referred to as "conformational capture". A description of the physical principles involved in forming RNA−protein complexes via conformational capture requires complete description of the dynamics of these structurally labile RNA molecules, including a characterization of long- and short-lived conformational states sampled by the RNA. However, experimental characterization of transitory states is complicated by the fact that the rate of transition may be too fast to allow for a comprehensive catalogue of all states.[1] Thus, although progress has been made recently in experimentally isolating the partially folded states of proteins,[2,3] the complete elucidation of protein or nucleic acid dynamics requires analytical and computational modeling to complement experimental observations. A common approach toward this goal is to fit a limited set of model parameters to experimental data that are sensitive to dynamics, such as NMR relaxation rates, line shapes, or residual dipolar couplings.[4−7] This procedure involves guessing-and-checking, using physical constraints on possible motions of the labeled residue(s) to guide the model-building process. However, this semianalytic approach becomes complicated when further degrees of freedom and new free parameters are required by the model to fit the data adequately.

A purely computational approach to the description of molecular states requires an accurate potential energy function (PEF), followed by either molecular dynamics (MD) or energy-minimization calculations. Molecular dynamics simulations are able to generate dynamic trajectories of the molecule and, in principle, explore molecular parameter space if sufficient numbers of trajectories are available; examples are found using the AMBER[8−12] and CHARMM packages[13−15] and others, such as Lindorff-Larsen et al.[16] However, the extrapolation of dynamics in nucleic acids to time scales of the order of microseconds or longer, where many conformational changes are expected to take place, has only recently begun to be explored.[17,18]

Energy-minimization techniques also rely on a well-validated energy function but involve the subsequent alteration of the relative conformations of parts of the molecule in an iterative manner to find the global energy minima.[19−23] It is then

Received: September 19, 2013
Revised: January 24, 2014
Published: January 30, 2014

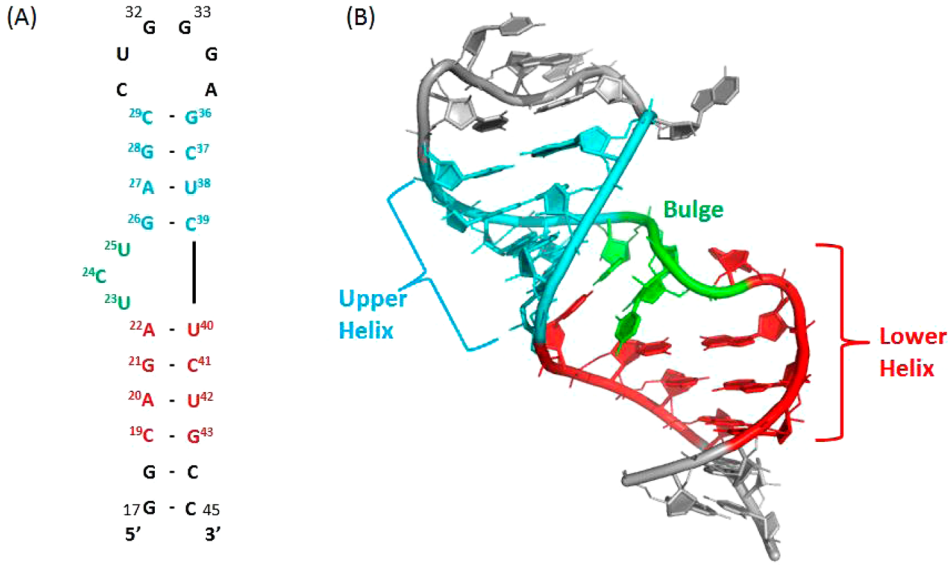


Figure 1. 29-nucleotide apical stem-loop of HIV-1 TAR RNA: (A) secondary structure; (B) sample tertiary structure. The different submotifs and the component residues used in the analyses are color-coded: upper helix (turquoise), lower helix (red), and the single-stranded bulge (green).

possible to generate multiple sample structures more easily than for a complete MD calculation. In the current manuscript, we utilize structures generated by energy-minimization techniques as a complement to MD-based analyses of dynamics. Use of energy-minimized structures is facilitated by the availability of structures on the Rosetta server.[24] Other servers such as the MC-SYM/MC-FOLD pipeline[25] are also available that allow the user to obtain all-atom RNA models.

We use the HIV-1 TAR (trans-activation response) RNA molecule as a model for a dynamics simulation based on a set of 500 low energy models generated using the program FARFAR[22] for the 29-nucleotide apical section of the RNA[26] (Figure 1). The TAR RNA binds the viral regulatory protein Tat,[27] a critical transcription elongation factor essential for viral replication. The Tat binding site surrounds the single-stranded, trinucleotide (UCU) bulge[27,28] and is contained within the 29-nucleotide construct. The bulge region interlinking the two helical stems is the primary binding site for the Tat protein[28,29] and exhibits considerable flexibility[4,30−32] allowing for the two helical regions to adopt a wide range of relative orientations.[4,33] Protein binding is believed to occur by "conformational capture", where the free TAR RNA exchanges between multiple conformers, one or more of which are amenable to Tat binding.[34] TAR RNA also provides a common RNA structural motif, where two helices are connected by a single-stranded bulge on one end and a backbone "hinge" at the other. Other RNA's exhibit similar structures, such as K-turns,[35] or the HIV-1 RRE (rev-response element).[36,37] It is therefore worthwhile to characterize the dynamics of such fundamental motifs, especially to characterize "large-scale" motions (as opposed to more localized motions) of one helical domain relative to another.

Earlier work in our group used solid-state NMR to identify intermediate rate motions for the residue U38 in the upper helix.[4] We more recently characterized the motions of U40 and U42 in the lower helix (Wei Huang, unpublished data). These data indicate that the two helices move relative to each other at slow rates ($\sim 10^5$ to 10^6 s^{-1}) relative to the rotational diffusion rate of the entire molecule ($\sim 10^8$ s^{-1}). These results motivated us to look at the distributions of the orientations of the upper helix relative to the lower helix among the set of lowest energy structures, as characterized by the Euler angles of an upper helix-attached reference frame relative to a different frame attached to the lower helix. Reduction to a set of three angles characterizes many features of the dynamical trajectory of this particular structural motif, while simplifying a very large-dimensional problem to a set of three essential coordinates. Local base librations, i.e., rotations of the bases around a base normal (representing vibrations of the base around the equilibrium base-paired orientation) or rotations of the base around the glycosidic bond (for the single stranded bases), are included in the simulations as well and provide atomistic detail regarding the motions of individual residues.

Here we extend our prior studies of nonrigid rotation and helix reorientation in TAR RNA to a full description of the concerted dynamics of the upper and lower helices and the bulged loop. Our approach describes a dynamic trajectory based on an ensemble of energy minimized structures with the Rosetta program FARFAR. Because FARFAR does not provide a Boltzmann-weighted distribution of states (and thus does not provide the entropy and free energy) our calculations rely on fitting phenomenological parameters to the data. To reduce the complexity of the problem, a predominant set of conformers is selected that describes its dynamic conformational ensemble. We used experimental residual dipolar couplings (RDCs) obtained from partial alignment of the TAR RNA by one alignment medium to achieve this selection, and other groups have recently used similar filtering techniques.[18,38,39] A summary of some of these techniques has been published recently.[40] In addition to filtering out a set of long-lived conformers along the stochastic trajectory, we acquire information regarding their relative populations. We complement the RDC filter with principal component analyses (PCAs) based on a judiciously chosen set of backbone torsion angles, as done to establish conformational clustering and free energy landscapes[41] of RNAs[12] and proteins,[39,42] and a PCA can serve as a useful aid in setting up dynamics calculations.

Finally, we use the outcome of that analysis to calculate solution spin−lattice relaxation times T_1 and the rotating-frame spin−lattice relaxation time $T_{1\rho}$ for ^{13}C spins in nucleotides

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

located in the upper and lower helices and in the bulge. The protocol would allow extensions to other observables as well, but the current work focuses on these to provide a clear proof of method. We direct the interested reader to more extensive work on other NMR relaxation parameters and their simulations.[30,43] We use two methods to perform these calculations. The first approach (slow exchange or "SE" model) assumes an effectively infinite time scale of exchange (i.e., the conformational exchange is much slower than any other relevant motional time scale). The second approach (general rate or "GR" model) allows for an arbitrary time scale of exchange (including one that overlaps with the rotational diffusion time scale). By combining complementary experimental and analytical techniques into a single framework, we have been able to construct a viable dynamic trajectory for the TAR RNA. We provide a flowchart of the methodology in Figure 2.

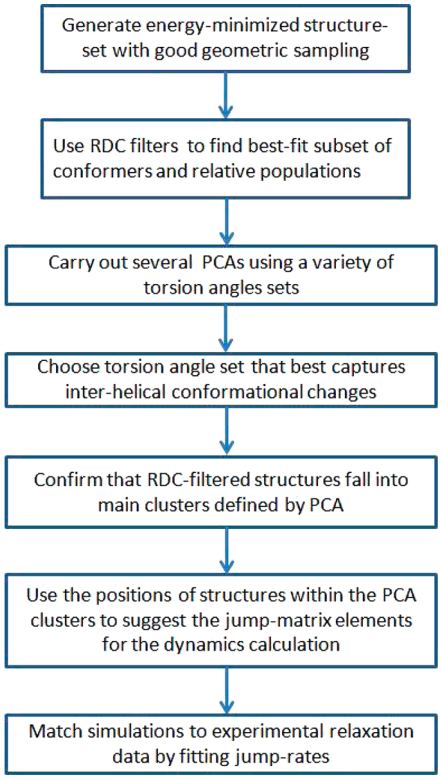


Figure 2. Protocol used in this manuscript for the simulation of domain-dynamics in a molecule.

2. THEORETICAL AND COMPUTATIONAL METHODS

A. Solid-State NMR Models for the Lower Helix of TAR. We present here the solid-state NMR (ssNMR)-derived models that served as motivations for the solution relaxation simulations. ssNMR studies of the dynamics of the uridine bases U38, U23, and U25 in TAR were carried out using samples selectively deuterated at the 5- and 6-carbon base sites.[4,44] U38 was chosen to represent the dynamics of the upper helical stem, whereas U23 and U25 are of interest on account of their positions in the single-stranded bulge. More recently, 5,6-^{2}H labels were introduced at the U40 and U42 positions in TAR, in correspondence with the lower helix (U42) and an unstable base pair that closes the bulge region (U40) (Wei Huang, unpublished data). The data included line shapes as well as T_{1Z} and T_{1Q} relaxation times collected on samples hydrated in all cases to 16 water molecules per nucleotide to reproduce conditions where motions were shown to be solution-like.[4,44] Each of the upper helical and single-stranded sites investigated had characteristically distinct spectral features. Within the lower helix, U40 and U42 showed results similar to each other, but distinct from the upper helix or single-stranded sites. Motional models generated to fit the data included a slower base motion consisting of jump between two equally populated sites, superposed on a faster motion occurring around the normal to the base-plane for the helical residues and around the glycosidic bond for the bulge residues. The analyses of the spectra recorded for the upper and lower helical sites were done independently resulting in two different sets of parameters.

The U38 base was modeled as undergoing a two-site ±4° jump process around the base-normal at a rate of 2×10^{8} s^{-1} in addition to a two-equivalent-site conformational exchange process, where the upper helix underwent a 9° bend and a 15° twist at a rate of ~10^{6} s^{-1} relative to a crystal-fixed frame.[4] A similar model could fit the data for the lower helix bases U40 and U42 as well, resulting in a 0−18° bending motion and a 18−25° twisting motion of the lower helix, but at slower rates on the order of 10^{5} s^{-1}. By analyzing relaxation data, we observed small amplitude (±6° to ±9°) local motions of the base at a rate of ~10^{8} s^{-1} for both U40 and U42. The U23 data were fit by significantly different models, involving a local two-site jump about the glycosidic bond of ±11° at a rate of ~10^{10} s^{-1}, and a 24° hop of the base at a rate of ~10^{8} s^{-1}, whereas U25 was modeled as experiencing a 30° jump at 6×10^{7} s^{-1} in addition to a much slower twisting of the base (6×10^{5} s^{-1}) with a large amplitude of ±40°.

We utilized the same two-site jump models for the local base motions of both the helical and bulge residues, under the rationale that solid-state sample conditions should be able to replicate solution conditions at least in the small-amplitude local motions at the hydration conditions of our studies. We also used the time scale of the local base jumps as a starting point for simulations of the solution conditions. Finally, we used the observation that conformational exchange motions in solution conditions occur on a time scale similar to that in solid-state samples. This is reflected in our use of the "slow exchange" formalism for part of the relaxation simulations, which assumes an exchange process much slower than all other motional rates. We did, however, also consider the more general case of an arbitrary rate of conformational exchange.

B. Solution-State Simulations. Structures were generated using the program FARFAR.[22] The torsion angles that were altered in the generation of the structures used here were those of the residues in the bulge and the bulge-adjacent base pairs of the two helical domains. The 500 lowest energy structures represented a distribution in energy of about 13 Rosetta units, where 1 Rosetta unit is approximately 1 $k_{B}T$ (Rhiju Das, private communication).

To reduce the dimensionality of the problem, we assumed that the simple motif of a helix−bulge−helix can be represented by a reduced set of relative helical orientation parameters. A set of three Euler angles transforming between the two helices is taken to be sufficient to discriminate between helical conformations (Figure 3) if the helices behave as rigid units.[33,45]

We have used experimental residual dipolar couplings (RDCs) for several bonds along the molecule. Because RDCs

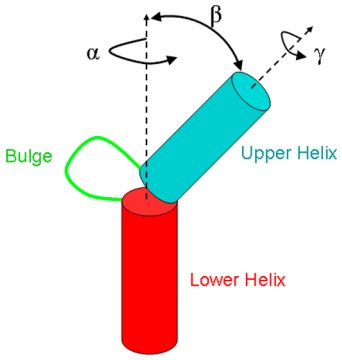


Figure 3. Relation between the Euler angles determined from the atomic coordinates and the associated domain motions.

are long-time weighted averages over all conformational states of the molecule, they potentially provide means of extracting both the best-fit conformers and their relative populations. The structures that were selected by this RDC-filtering process, along with the best-fit populations, were subsequently used to simulate the solution relaxation times.

We incorporated the local base motions obtained by fitting the ssNMR data into the RDC-selected structures and calculated the solution relaxation times using previously published methods.[6,7] When a small amount of variation for the local amplitudes and rates of motion, and for the rates of conformational exchange between the selected structures, was allowed, the solution relaxation times for the 17 residues in the molecule could be calculated (7 in the lower helix, 3 in the bulge, and 7 in the upper helix) with this method.

i. Helical Parametrization of Structures. The relative orientations of the upper and lower helices were quantified by considering the orientation of the normal vector of the U38 base relative to a frame defined by a *z*-axis aligned with the lower helical axis. The method of evaluating this relative orientation and the subsequent binning of structures within this scheme is as follows:

(a) Define the upper and lower helical axes for all structures using the program 3DNA.[46] The upper helical axis is taken to be the average local helical axis of the A27-U38::G28-C37 and G28-C37::C29-G36 dinucleotide steps, where the base pairs flanking the bulge were excluded due to possible distortions from ideal A-form helical structure. The lower helical axis is calculated similarly as the average over the C19-G43::A20-U42 and A20-U42::G21-C41 dinucleotide steps.

(b) Define the lower helix coordinate frame (LHF) by choosing the lower helical axis (calculated above) as the *z*-axis and the perpendicular from the *z*-axis to the G43 C8 atom as the *y*-axis. This choice of the *y*-axis was made as the 500 structures did not differ in the orientations of the first few base pairs (including C19-G43), so the *y*-axis would be the same across all structures. However, the specific choice of the G43 C8 atom for this purpose was arbitrary.

(c) Calculate the α angles for each structure, defined as the angle between the projection of the upper helical axis onto the *xy*-plane of the LHF and the *x*-axis of the LHF.

(d) Calculate the β angles, defined as the angle between the upper helical axis and lower helical axis.

(e) The γ angle is then defined by the orientation of the normal of U38 base (the vector perpendicular to the C4−C2 and C6−C2 bonds) about the upper helical axis. Extracting this information from the structures requires first removing the α and β dependence by rotating the original U38 base-normal vector $\vec{v}_{\mathrm{U38norm}}$ about the fixed LHF axes as follows: $\vec{v}'_{\mathrm{U38norm}} = \vec{R}_{y_{\mathrm{LHF}}}(-\beta)\vec{R}_{z_{\mathrm{LHF}}}(-\alpha)\vec{v}_{\mathrm{U38norm}}$. The resultant vectors are distributed around the LHF *z*-axis as a function of their γ angles. The Euler angles described above are related to the domain motions as shown in Figure 1.

(f) Bin the 500 structures as a function of the Euler angle set $\{\alpha, \beta, \gamma\}$. The bins were chosen in 10° increments for α ($0° \le \alpha \le 360°$) and β ($0° \le \beta \le 180°$) angles, resulting in 36 and 18 bins, respectively. Instead of binning the γ angle in terms of degree increments, we fixed the number of bins to 5 because of an observed correlation between the α and γ angles, which results in a shift in the values of γ for every α bin, as well as the restriction of the γ values to only a portion of the full phase space. Thus, trying to bin all possible γ angles would have unnecessarily increased the computational time.

ii. RDC Constraints. To select a subset of structures for multisite jump simulations from a starting set of 500 energy-minimized structures generated by FARFAR, we use experimental RDCs[47] in conjunction with the predictive program PALES.[48] We only consider the data from a PEG/Hexanol mixture[49] because this uncharged medium aligns the charged RNA only through steric hindrance of the overall rotation. This situation does not require a detailed characterization of the RNA charge density, as would be required for the simulation of RDCs measured in charged alignment media. Extension to charged media data may be considered in the future.

The purely steric version of PALES was used to calculate the eigenvalues and eigenvectors of the Saupe alignment tensor of the molecule by sampling multiple allowed orientations in the presence of a flat (infinitely large) obstruction. The alignment tensor eigenvalues are then used in conjunction with directional information for a particular bond relative to the alignment tensor principal axis frame (PAS). For the purposes of simulating the experimental RDCs, the bond orientations are considered to be static within each conformer. The only motions to be considered are then the exchanges between distinct conformers. To include dynamic sampling of multiple conformations, we use the weighted average of the RDCs of each conformer $\mathrm{RDC}_{\mathrm{IS}}^{\mathrm{Total}} = \sum_{i=1}^{N_{\mathrm{conformers}}} p_i \mathrm{RDC}_{\mathrm{IS}}^{i}$ for the RDC of a bond between spins I and S, with the population of each conformer i being represented by p_i.

The residues considered for this study were C19, A20, G21, A22, U40, C41, U42, and G43 from the lower helical stem, U23, C24, and U25 from the bulge, and G26, A27, G28, C29, G36, C37, U38, and C39 from the upper helical stem. The bond types included the C6−H6 bond (pyrimidines), the C8−H8 bond (purines), and the C5−C6 bond (pyrimidines) from the bases, the C1′−N1 (pyrimidines) and the C1′−N9 (purines) glycosidic bonds, the C1′−H1′ and C4′−H4′ bonds from the furanose rings, and the C5′−H5′ and C5′−H5″ bonds from the backbone. RDCs for some bond types were not available for some of the residues. Furthermore, the RDCs for the bulge residues were considered separately, and not in the same simulations as the helical residues. This is because the bulge is likely to be significantly more mobile than the helices and will sample a significantly larger number of

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

configurations, requiring a different set of simulation conditions.

Effective bond lengths of the species in question are required in the final PALES calculation. The values we chose for the bond lengths are the aliphatic C−H bond length, $r_{CH,Aliph}$ = 1.1 Å,[50,51] the aromatic C−H bond length, $r_{CH,Arom}$ = 1.09 Å (average of Ying et al.[51] and Allen et al.[52]), the aromatic C−C bond length, $r_{CC,Arom}$ = 1.4 Å,[53] the glycosidic C−N bond length for the cytosines, $r_{CN,GlycCyt}$ = 1.47 Å,[53] and the glycosidic C−N bond lengths for the remaining base types, $r_{CN,GlycRest}$ = 1.48 Å.[53]

To gain a geometric perspective on the best-fit structures from the RDC comparison, we binned the simulated couplings for each structure according to the $\{\alpha, \beta, \gamma\}$ angles determined for the 500 structures. The RDCs within each bin were then uniformly averaged to produce a single bin-RDC for each bond type and residue, rather than retaining the values for individual structures. This was done because, although the structures within each bin have similar helical conformations, they may differ in the orientations of bonds of certain residues relative to the large-scale conformations. By averaging over these variations in the bonds, we effectively included in the simulations the small amplitude thermal fluctuations of the atomic bond orientations.

To select the best-fit set of structures, we started with an arbitrary initial set of N structures and allowed the choice of the $(N + 1)$th angular bin to float while attempting to optimize the total χ^2 as well as the Pearson's correlation coefficient between the simulated and experimental RDCs. In addition, we varied the relative weights of the $(N + 1)$ bins. Once the best-fit parameters were obtained for this first iteration, the $(N + 1)$th bin so obtained replaced the lowest probability structure from the remaining N, and the search was repeated for a second iteration and beyond. To make the final choice of N, we started with $N = 2$ and after minimizing the χ^2 for a choice of two structures, we added a third and repeated the process. Proceeding thus, we found that $N = 5$ structures (i.e., 5 bins which also happened to have only one structure in each of them) were sufficient to produce the best-fit to the RDC data, with no improvement after the fifth iteration. In addition, as a separate, independent check of the results of the previous procedure and to allow for the variation of the number of bins more easily, we generated a Markov-chain Monte Carlo (MCMC) simulation to search through the bins and populations for a best-fit to the RDCs, with the potential for varying the number of bins. The number of parameters for an N-conformer search was $2N - 1$ (N bin choices + $N - 1$ probabilities to be floated). The Markov-chain Monte Carlo method did not yield better results than the iterative technique.

iii. PCA of the Torsional Degrees of Freedom. To determine the number of exchanging conformers required to describe the dynamics in TAR RNA, and to corroborate the results of our RDC filter, we performed a principal component analysis (PCA), following procedures applied to molecular ensembles of proteins,[39] and RNAs.[12,41,42] The covariance matrix $\sigma_{ij} = (q_i - q_i)(q_j - q_j)$ was calculated as described by Mu et al.,[42] i.e., by proposing the following variable set $\{q_{2j}\}$:

$$\left.\begin{aligned} q_{2j-1} &= \cos(\varphi_j) \\ q_{2j} &= \sin(\varphi_j) \end{aligned}\right\} \quad j = 1, \ldots, N_{torsion} \tag{1}$$

where φ_j is the jth torsion angle of interest, and $N_{torsion}$ is the number of torsion angles used in the analysis. The use of the cosine and sine functions removes complications associated with the periodicity of the torsion angles by helping to uniquely identify particular values of the angles.

The covariance matrix of these variables is calculated with averages over the full ensemble of structures and is subsequently diagonalized. The eigenvalues (and associated eigenvectors) are arranged in descending order, with the highest values representing modes with the largest contributions to the structural scatter. In our results, we have found that 2 or 3 modes contain a majority (∼70%) of the total variance in the data. In addition, histograms of the projections of the ensembles for each eigenvalue are examined for Gaussianity. As discussed in the above references,[12,41] a mode whose histogram consists of a single Gaussian-like peak only represents continuous fluctuations about a central structure, whereas multimodal distributions describe discrete conformations separated by free-energy barriers. Thus, from the perspective of assessing the conformational transitions of the molecule, only modes with multimodal distributions are considered relevant. This adds another layer of information regarding the molecular dynamics. Our results for PCAs with different choices of torsion angle sets (as described in the following) indicate that the first 2 or 3 modes were non-unimodal in distribution and therefore of primary significance in describing conformational exchanges.

We performed several PCAs, each with a different choice of coordinates. The first set included all torsional "suites"[54] along the bulge and hinge region (PCA method 1). The sugar-to-sugar suite as defined in the above reference consists of the set of 6 backbone torsion angles $\{\varepsilon, \zeta, \alpha, \beta, \gamma, \delta\}$ as well as the glycosidic torsion angle χ. In our study, we focused attention only on the 6 backbone torsion angles. The residues included were A22, U23, C24, U25, G26, C39, and U40. These were meant to encompass the conformationally relevant part of the molecule under the assumption of relatively rigid helices. The number of torsion angles considered was therefore 42 (7 bases ×6 torsion angles) resulting in an 84 dimensional PCA (because the cosine and sine functions of the angles are treated as separate variables). However, the PCA of this set resulted in several modes which contributed substantially, with no clear clusters in any of the largest modes.

The inclusion of the single-stranded bulge region may have caused the lack of structure in the PCA projections calculated in this first method, because single-stranded regions are significantly more flexible and may add a level of disorder to the torsional distribution. To isolate the interhelical motions, we evaluated the PCA of only those torsional "suites" that extend from U38 to C39 and from C39 to U40 (PCA method 2). These suites include only the hinge region of the molecule, and resulted in a 24-dimensional PCA.

To check the robustness of the clustering results obtained from method 2, we then carried out a series of PCAs by successively including an additional sugar-to-sugar backbone suite: U38 through C41 (PCA method 3), U38 through U42 (PCA method 4), C37 through U42 (PCA method 5), and C37 through U40 (PCA method 6). The results of these analyses and the relation to the results from the RDC fit will be discussed in the Results below.

iv. Relaxation Time Simulations. The evaluation of the solution relaxation times is based on two previously explored techniques.[6,7,55] Both methods involve the calculation of the

two-time correlation function for the orientation of an atomic bond located within a nonrigid Brownian rotator relative to a fixed laboratory frame. The evaluation proceeds by introducing an intervening reference frame associated with the principal axis system (PAS) for the rotational diffusion tensor, which is time dependent as a result of exchange between different structural conformers. The correlation function then becomes dependent on Wigner rotation matrices whose arguments are the Euler angles that orient the rotational diffusion tensor of the molecule relative to the laboratory frame. In turn, the Fokker–Planck equation in three-dimensions allows the evaluation of the transition probabilities from one set of Euler angle orientations to another. The form of the Fokker–Planck equation, which accounts for coupling between rotational diffusion and conformational changes, is

$$\frac{\partial}{\partial t}P(\vec{\Omega},\beta,t|\vec{\Omega}_0,\alpha) = -\sum_{i,j=1}^{3}\hat{L}_i D_{ij}^{\beta}\hat{L}_j P(\vec{\Omega},\beta,t|\vec{\Omega}_0,\alpha) + \sum_{\gamma=1}^{N_{\text{conformers}}} R_{\beta\gamma}P(\vec{\Omega},\gamma,t|\vec{\Omega}_0,\alpha) \quad (2)$$

where $P(\vec{\Omega},\beta,t|\vec{\Omega}_0,\alpha)$ is the probability that the molecule will transition from a diffusion tensor-to-laboratory frame orientation of $\vec{\Omega}_0$ at time 0 and in a conformational state α to an orientation of $\vec{\Omega}$ and a conformational state β at time t, given diffusion tensor elements of D_{ij}^{β} in the conformational state β and an exchange rate of $R_{\beta\gamma}$ between conformational states β and γ. This equation can also formally cover the case of continuous transitions to new conformational states by allowing the number of conformers to be infinite. However, in all cases considered in the original references, and here as well, only discrete jumps will be considered.

If the exchange rate is much slower than the rate of diffusion and all other time scales of exchange processes, yet faster than the relaxation rates themselves, the rotational diffusion problem is effectively decoupled from conformational exchange. Under these conditions, the SE model applies, and it is possible to calculate the relaxation rates for the molecule as the weighted-average over the relaxation rates for the individual conformational states of the molecule:[6]

$$\left\langle \frac{1}{T_{1,2}} \right\rangle = \sum_{i=1}^{N_{\text{conformers}}} w_i \left(\frac{1}{T_{1,2}^i} \right)$$

$$\langle \text{NOE} \rangle = 1 + \langle T_1 \rangle \left\{ \sum_{i=1}^{N_{\text{conformers}}} w_i \frac{(\text{NOE}^i - 1)}{T_1^i} \right\} \quad (3)$$

Expressions for the NOEs can also be derived by appropriate weighting of the NOEs of individual conformational states. The procedure to calculate the relaxation times for each conformer are provided in the original reference.

Solution of eq 2 for the case of a general exchange rate (i.e., the GR model) between conformers involves considerably more analytical and computational processing.[7,55] We have solved this problem for the case in which the eigenvectors of the diffusion tensors of the exchanging conformers are coincident at the moment of exchange (see Ryabov et al. for general case[56]) and correlation functions have been published.[55] This general rate analysis also indicates that the slow exchange regime occurs at time scales longer than about 1 μs.

For the carbons considered in this analysis, the rotating frame z-relaxation time $T_{1\rho}$ was measured instead of T_2.[4,30] Under the application of a weak spin-lock field and the assumption of Lorentzian spectral densities, $T_{1\rho}$ contains the same spectral density information as T_2.[57] Therefore, for the purposes of this paper we operate under this assumption and simulate the T_2 relaxation times.

The structure set and the corresponding populations preselected by the RDC filter are the basis for simulations of the relaxation times using eq 2. Eigenvalues and eigenvectors of the rotational diffusion tensor are first calculated using the public-domain program HYDRONMR.[58] Then orientations of the atomic bonds of the residues of interest are calculated with respect to this axis system, i.e., the principal axis system of the rotational diffusion tensor. The orientational parameters, together with the diffusion tensor eigenvalues are input into the two algorithms we have derived for simulations of the relaxation times of nonrigidly rotating macromolecules: (a) the "slow-exchange" formalism,[6] describing the case where the conformational jumps occur at a rate much slower than the rate of overall rotational diffusion of the molecule, and (b) the general rate formalism,[7,55] where arbitrary rates of exchange are allowed. The "slow-exchange" formalism, though merely a limiting case of the general rate theory, has the advantage of being significantly faster and easier to implement and so is considered here.

The residues considered were A20, G21, A22, U40, C41, U42, and G43 from the lower helical stem, U23, C24, and U25 from the bulge, and G26, A27, G28, C29, G36, U38, and C39 from the upper helical stem. In the current work, we have only simulated the motions of the bases of these residues: the C6–H6 bonds for the pyrimidines and the C8–H8 bonds for the purines.

The parameters used in the simulation include the atomic element radius (AER),[58,59] the radius of the beads used in the HYDRONMR calculation of the diffusion tensor, and the bond lengths. The AER was chosen to be 2.3 Å and the bond lengths for the carbon–hydrogen bonds for the aromatic bases were chosen to be 1.1 Å, both choices having been justified in previously published work.[6] The viscosity was chosen to be 1.096 cP[60] to correspond to the conditions of the solution experiments (99.9% D_2O at 25 °C).

We have also incorporated the two-site base motions inspired by simulations of the solid-state NMR (ssNMR) data: the so-called "base libration" occurs around a vector normal to the plane of the base in the case of helical residues, whereas the two-site motion is modeled to be around the glycosidic bond for the bulge residues. We floated the values of the rates and amplitudes of these two-site jumps relative to the ssNMR models, which were found to be on a time scale much shorter than that of the conformational exchange. For the slow exchange simulations, these internal, local motional rates and amplitudes were the only free parameters, whereas in the general rate simulations we floated the conformational exchange rates between the states. The fitting procedures were carried out in a combination of grid-searches and MCMC techniques.

3. RESULTS

A. RDC Filter of the Structural Ensembles. Five structures provided the best χ^2 values to fit the RDC data, as obtained by iteratively searching through the bins and updating the choices of bins and relative populations. They are shown in

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726–1742

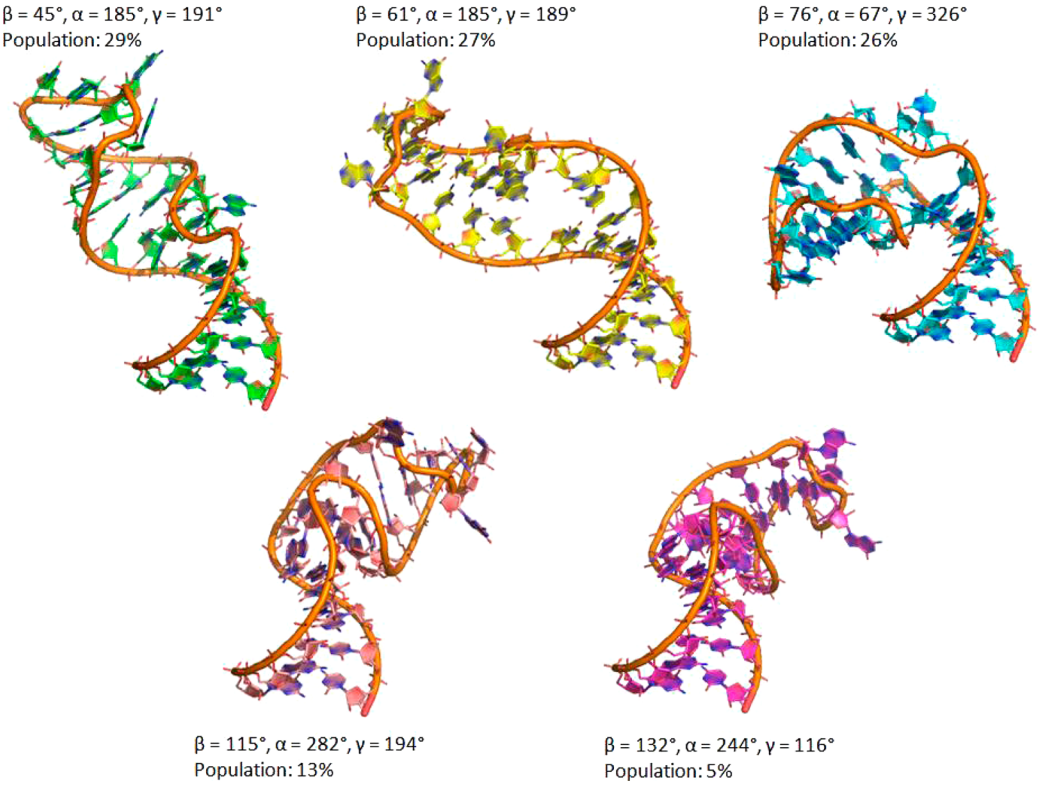


Figure 4. Five structures obtained by the RDC-filtering procedure to represent the ensemble of Tar conformation that describes the experimental data, together with their interhelical Euler angles and population percentages.

Figure 4, and key characteristics are summarized in Table 1. For simplicity, the structures will be referred to by their respective bend angles. Thus, the highest population structure will be called the "45° structure", the second highest populate structure as the "61° structure", and so on.

Table 1. Twist of Upper Helix Relative to Lower Helix (α), Interhelical Bend Angle (β), Twist of Upper Helix around Its Own Helical Axis (γ) (Figure 3)

structure	α (deg)	β (deg)	γ (deg)	% relative population
1	185	45	191	29
2	185	61	189	27
3	282	115	194	26
4	67	76	326	13
5	244	132	116	5

The χ^2 for the best-fit set of structures was 11 460 for a set of 48 RDCs, with 9 degrees of freedom (5 bin choices and 4 probabilities), for a reduced χ_r^2 = 302 (=11460/(48 − 9 − 1)), whereas the Pearson's correlation coefficient was 0.72. It is to be noted that the χ^2 represent unweighted values; i.e., the discrepancies between the experimental and simulation RDC values are not inversely weighted with the error bars of the RDC values (which have not been calculated). This assumption is equivalent to assuming that the error on all measurements is 1 Hz, which is likely to be a considerable underestimation of the true error bars. The comparison between experiment and simulation is shown in Figure 5. In the figure, the RDCs from different bond types are clustered together for each residue. We further attempted an MCMC fit procedure with the possibility of 6, 7, or 8 structures but were unable to improve upon the fit. It is possible that this fit may be improved by the continuation of the MCMC procedure to a greater numbers of iterations, but the PCA analysis reported in the following section provides further corroboration that the model found by RDC fitting represents the conformational landscape of the molecule well.

Plotting the calculated RDCs against the experimental values (Figure 6), we found that the trend, on average, is toward an underestimation of the RDCs by the simulations. The dashed blue line in Figure 6 is the best-fit line to the data and has a slope of 0.4 and a y-intercept of −3.9. An underestimation of the RDCs may arise from using a smaller degree of alignment in the simulations than in the actual experimental situation. One possible source of this discrepancy may be the current assumption of a simple steric model for the alignment of the molecule by PEG/hexanol. Recent work[61] has proposed that there are subtleties in the alignment process, including the possible contributions of complex alignment medium topology and electrostatic alignment, that are not incorporated in simulations using only the basic steric version of the PALES algorithm.

This result is a cautionary statement in the application of simple steric models in the simulation of potentially complex alignment media. We attempted to fit RDC data collected in glucopone/hexanol mixtures and in Pf1 filamentous bacteriophage media and found that the models selected were different (two or three of the structures chosen were the same compared to the PEG/hexanol model). One obvious reason for this was the availability of RDC data for different bonds in the different media. However, there is potentially a fundamental difference in the alignment properties of the media as well. For example, the Pf1 Phage medium is negatively charged[62] and, thus, as may be the case with PEG/hexanol as well, the alignment has an electrostatic component. This must be taken into account more carefully in future analyses.

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

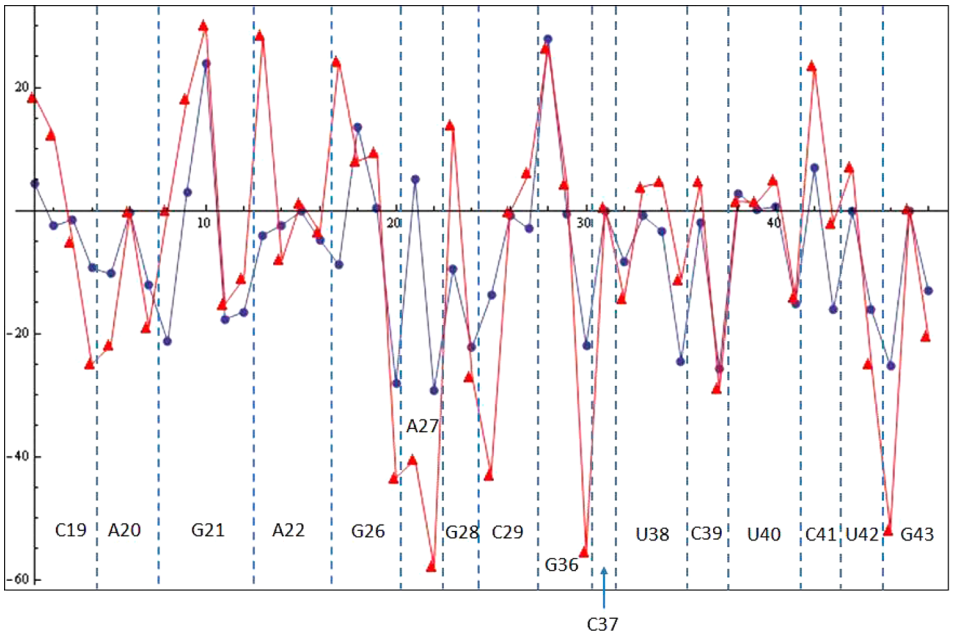


Figure 5. Comparison of the experimental RDCs (red triangles) with the RDCs generated by the best-fit simulation parameters (blue circles) for the helical residues in HIV-1 TAR RNA. The residues associated with the RDC values are labeled as well. The values shown include those for backbone (C5′−C5′, C5′−H5″), furanose (C1′−H1′, C4′−H4′), glycosidic (C1′−N1 for pyrimidines, C1′−N9 for purines), and base (C5−C6 and C6−H6 for pyrimidines, C8−H8 for purines) bonds.

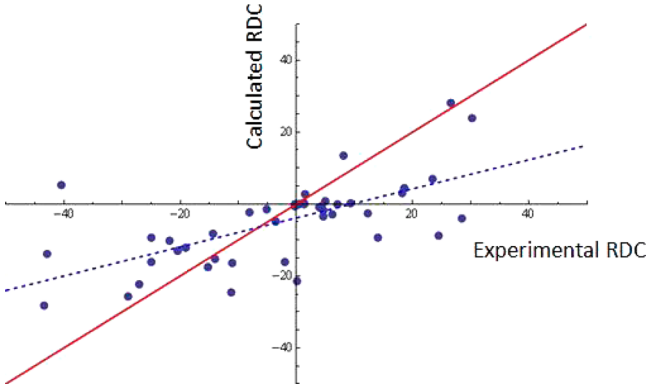


Figure 6. Plot of calculated RDC values vs experimental RDC values for the best-fit set of structures and populations. The dashed blue line is a linear fit to the data, and the solid red line is the ideal case where the calculated and experimental values match perfectly.

To further discern the causes of the discrepancy in the fit, we looked at all helical RDC values that contributed to a deviation of magnitude 10 or greater and found a set of 18 RDCs: 8 correspond to C1′−H1′ furanose ring bonds, 2 to C4′−H4′ furanose ring bonds, and the remaining 6 to C6H6/C8H8 base bonds, occurring in 12 helical residues in both the upper and lower stems as well as among all four nucleotide types. Removal of these RDCs gave $\chi^2 = 802$ and $\chi_r^2 = 40$. The large deviations observed for furanose ring bonds indicate the existence of additional motions localized to the furanose rings, as has been reported for DNA,[5] that have not been accounted for using the current set of 500 structures. These motions involve an exchange between the C2′-endo and C3′-endo conformations, and on the time scale of the RDCs these motions may be averaged out to produce an intermediate conformation. It is likely that the furanose ring samples these conformations even for the residues stacked in a helical configuration. A similar argument holds for those base bonds that show a large discrepancy in RDC values: there may be additional vibrations in the base orientations that are not adequately sampled by the 500 structures.

We also compared the five structures selected above to the set of 20 lowest energy TAR RNA structures recently generated on the basis of NOE data[47] that have not been constrained by RDC data. Upon calculating helical axes and orientations in the same manner as for the 500 FARFAR structures in this manuscript, we find that 11 out of 20 of the structures, including 3 out of the 5 structures that best fit the NOE data, occur within the same 10° bend angle bin as the highest population structure from the RDC fit described above. Thus, we believe that our approach identifies a predominant conformation set.

To test the robustness of the search algorithm, we performed two simulations of fitting a reduced data set upon the random removal of (a) 10 RDCs and (b) 15 RDCs (different RDCs were deleted in each of these two cases). Removing the first set of 10 resulted in the selection of the same 5 conformers as from the full set, along with a sixth new structure with a population of 4%. The populations of the 5 full-set best-fit structures were slightly different (maximum change of 12%). Removal of 15 RDCs reproduced 4 out of the 5 full-set best-fit structures, and two new conformers with populations of 7% and 5%. The maximum change in population among the 4 best-fit structures was 7% in this case. These results indicate that the choice of structures is robust to a reduction in the size of the experimental data set.

Jumps between the five conformers shown in Figure 4 require a combination of bending and twisting about either the lower or the upper helices. Among the entire set of energy-minimized structures, there was an observed correlation between the α and γ angles, as has been reported previously.[45,63,64] These correlations may be reflected even in the jumps among this set of five structures, representing a free

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

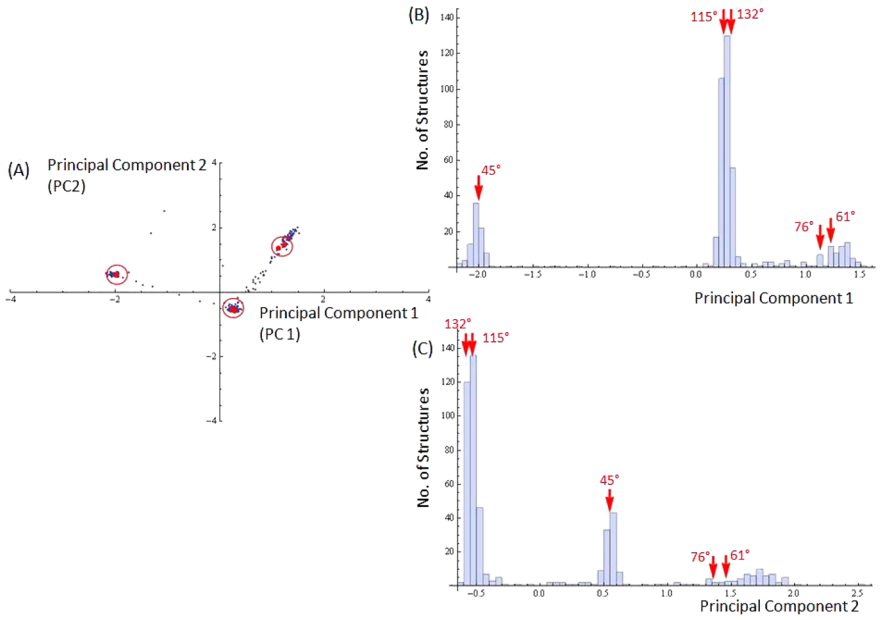


Figure 7. Principal component analysis of the torsion angle "suites" along the hinge (residues U38 through U40). (A) Principal components 1 and 2 for the 500 structures, represented by blue dots, with the positions of the five best-fit structures marked explicitly by red dots, and further encircled for clarity. (B) Histogram of the first principal component (corresponding to the largest eigenvalue of the covariance matrix), with the positions of the five best-fit structures marked explicitly by red arrows. (C) Histogram of the second principal component (corresponding to the next-to-largest eigenvalue of the covariance matrix), with the positions of the five best-fit structures marked explicitly by red arrows.

energy landscape where the exchanges between minima involve coupled shifts in Euler angle values.

B. PCA Clusters. We carried out a series of PCAs to (a) identify the choice of torsion angles that best captures interhelical motions, (b) corroborate the RDC-filtered set by overlaying the five chosen structures on the clusters obtained from the PCA of choice, and (c) identify jump matrix elements for the exchanges between the five RDC-filtered structures in the dynamics calculation.

Given the lack of clear clustering results from PCA method 1, which incorporates all backbone torsion angles in the bulge and hinge, we examined whether torsion angles on one side of the helical joint would suffice to describe the interhelical reorientation (PCA method 2, incorporating the backbone torsion angle suites between U38 and U40). When this was done, only the first two principal components (PCs) contributed significantly, accounting for 75% of the fluctuations in the molecule (Figure 7). Furthermore, these two PCs were the only ones with a multimodal probability distribution across the 500 structures. This non-single-Gaussian distribution signals the presence of conformational clusters in energy minima separated by significant free energy barriers.[12,41] The map of these two PCs is shown in Figure 7A and shows the presence of three to four major conformational clusters. The large red dots superimposed on the 2D plot correspond to the five structures selected from the RDC filtering procedure. Parts B and C of Figure 7 show the population distribution histograms for PC 1 and PC 2, respectively. Again, the positions of the five structures from the RDC filter are indicated by red arrows.

Three points are noteworthy. First, the five structures chosen as the best-fit set match up well with the main conformational clusters obtained from this PCA, suggesting that our RDC-filtered set has captured the relevant information about the major conformational clusters of the molecule. Second, structures with similar interhelical bend angles have similar values of each of the principal components. The principal component values of the two structures with bend angles of 115° and 132° occur in close proximity to each other. The same is true of the pair of structures with bend angles of 61° and 76° (it is true, however, that the 45° structure does not differ significantly from the 61° structure; the PCA suggests that there is a free energy barrier that separates even these two neighboring structures). This is important because other PCA methods (described below) separate structures with similar interhelical orientations, possibly due to the presence of additional degrees of freedom that do not contribute significantly to interhelical reorientation. Finally, the sums of the probabilities of the best-fit structures within each cluster are similar to each other: the 45° bend structure has a population of 29%, the 61° and 76° structures have a joint population of 40%, and the 115° and 132° structures have a joint population of 31%. Because the histogram heights do not correlate well with the nearly uniform probability distribution, we fit the jump rates numerically, as described below.

Subsequent attempts at testing the robustness of our cluster analysis yielded further interesting results. PCA method 3 (U38 through C41) showed a marked difference in clustering of the structures. Three PCs contributed significantly, yielding about 70% of the total fluctuations, with all three now being multimodal. However, the five structures do not all seem to fall within major clusters. Moreover, the 61° and 76° structures no longer fall within the same cluster, nor do the structures with bend angles of 115° and 132°. This occurs because of the intervention of the torsion angles associated with U40. When the helical parameters of the 500 structures were searched, about 37% of the structures U40 did not form a canonical Watson−Crick pairing with A22, indicating considerable conformational variability of this residue, at least within the physical picture generated by the energy-minimization

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

ensemble. To test this hypothesis, we extended the PCA up to U42 (PCA method 4) on the lower stem and up to C37 on the upper stem (PCA method 5). Both these methods gave very similar clustering to method 3, with three significant PCs. We interpret this observation as confirmation that, after the inclusion of U40, the remaining residues behave fairly rigidly and do not change the results of the PCA. As a final confirmation of this conclusion, we carried out a PCA including only torsion angles from C37 to U40 (PCA method 6). The clustering results for this PCA proved to be the same as for PCA method 2.

Given the change in clustering associated with the inclusion of the U40 degrees of freedom, we utilized the results of PCA method 2 to set up the rate matrix for the relaxation time simulations. In general, for N conformers, the number of combinations of pairwise rate constants that need to be fit is ${}^{N}C_2$. Given the clustering suggested by the PCA, we reduce the fit problem from 10 (=${}^{5}C_2$ for the five RDC filter structures) to five parameters in the following manner. For pairs of structures that occur within the same cluster in the two PC distributions, we allow for only one distinct exchange rate between both members of the pair in that cluster and any structure in another cluster. In addition, there is also one intracluster exchange rate for each cluster. Thus, the rates used in the fitting process were for the following exchanges:

$$(1)\{45°\} \leftrightarrow \{61°, 76°\},\ (2)\{45°\} \leftrightarrow \{115°, 132°\},$$
$$(3)\{61°\} \leftrightarrow \{76°\},\ (4)\{61°, 76°\} \leftrightarrow \{115°, 132°\},$$
$$\text{and } (5)\{115°\} \leftrightarrow \{132°\}$$

These exchange processes are shown graphically in Figure 8. The assumption in this parameter reduction is that all the structures within a cluster are separated from other clusters by similar free energy barriers.

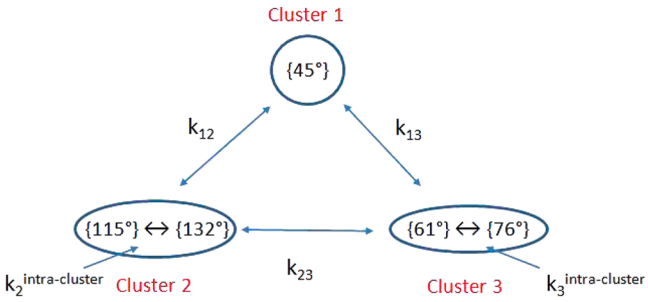


Figure 8. Exchange between clusters inferred from the PCA, together with the five rates used in the jump matrix for the relaxation time simulations.

C. Calculation of Solution Relaxation Times. The RDC-filtered conformer set of five structures, together with the relative probabilities, were used to calculate the T_1, T_2 and NOE values. We used two different approaches to calculate the C6–H6 (pyrimidine) and C8–H8 (purine) relaxation times: (a) the slow exchange method, where the assumption is that the exchanges occur at an infinitely slow rate (compared to the rotational correlation time), and (b) the general rate method, where we fit the relaxation times by allowing the exchange rates to vary arbitrarily.

i. Slow Exchange Method. The general rate analysis has shown[7] that the slow exchange regime in TAR RNA effectively occurs for time scales longer than about 1 μs. Thus, the results of this subsection assume conformational exchanges occur on a scale longer than 1 μs. As a starting point for the fitting process, we use rates and jump amplitudes for the base librations close to those obtained to fit the solid-state NMR data of the uridines,[4,6,7] and changed both the rates and amplitudes in small increments to improve on the χ^2. Using the fit to the T_1 values as a benchmark, we found that it suffices to fit only two base-libration rates, one for the upper helix residues and one for the lower helix residues. This simulation model was inspired by the results of the two solid-state NMR analyses. The exceptions, however, are the parameters for U40 and U42 in the lower helix. We obtain a significantly better fit by using the rates and jump amplitudes for the upper helix for these two residues, indicating that these residues are more similar in local base motion to the upper helix. We found that there were several nearly degenerate minima or best-fit "windows" in the χ^2 plot for the local motion parameters, with the upper and lower helical parameters behaving independently. The rates in these "windows" vary between $\sim 10^7$ s^{-1} and $\sim 10^9$ s^{-1}, whereas the amplitudes are less than 20°.

We did attempt simulations of additional models of base-libration such as a treatment of the rates of purines and pyrimidines independently, and the assumption of a constant rate across the entire molecule. The model described above was the best among the three considered. There is always the possibility that more complex models of base libration rates may fit the data better; for example, we may treat the libration rate of each individual residue as an independent parameter, or choose the purines and pyrimidines in the lower helix as independent from the purines and pyrimidines, respectively. However, this would increase the number of free parameters in the problem, and we chose the above model as a balance between an arbitrary increase in free parameters and an attempt at a physically realistic representation.

The following representative values of the local motion parameters simulate the relaxation times well:

(a) upper helix, U40 and U42 base-libration rate = 4.6×10^7 s^{-1}
(b) upper helix, U40 and U42 base-libration jump amplitude = 13.7°
(c) lower helix (without U40 and U42) base-libration rate = 6.6×10^8 s^{-1}
(d) lower helix (without U40 and U42) base-libration jump amplitude = 9.8°

The match between the experimental and simulation relaxation T_1 and T_2 values is shown in Figure 9. For quantitative comparison, we calculated the root-mean-square deviation (RMSD) across the 14 helical T_1 values to obtain an RMSD of 5.3 ms. The corresponding RMSD for T_2 comparisons was 1.5 ms.

The error bars shown in Figure 9 describe the statistical error in the measurements. However, the potential systematic error, not quantified by the authors,[30] is larger. Yet, even at this level of uncertainty there is a consistent difference between the best-fit local jump amplitudes for the upper and lower helices, with the lower helix amplitudes being larger on average (this is not reflected in the parameter set shown above but is true for the best-fit windows in general).

Thus, the solution relaxation times can be fit to an RMSD close to the statistical error in the experiments by assuming a slow exchange rate (i.e., slower than 10^6 s^{-1}) between the five structural configurations of TAR RNA shown in Figure 4 with

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726–1742

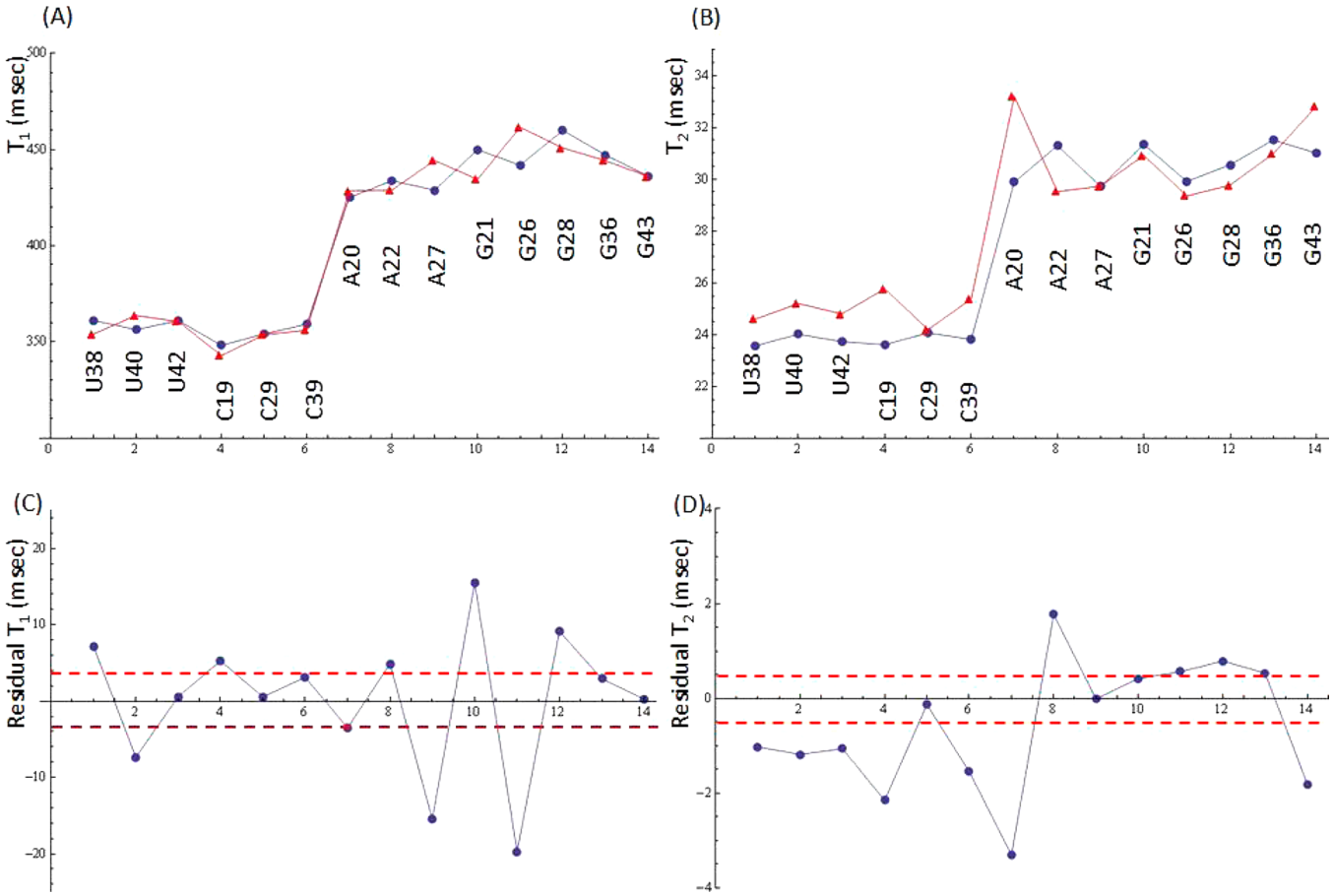


Figure 9. Relaxation time simulations for the C6−H6 (pyrimidine) and C8−H8 (purine) bonds using the slow exchange method, and comparisons of residuals (discrepancies relative to experimental values) to statistical error bars: (A) T_1 simulations (blue circles) compared to the experimental T_1 values (red triangles); (B) T_2 simulations (blue circles) compared to the experimental T_2 values (red triangles); (C) Difference between simulation T_1 and experimental T_1 values, together with the statistical error bars on experimental data (red dashed lines) at ±3.2 ms; and (C) Difference between simulation T_2 and experimental T_2 values, together with the statistical error bars on experimental data (red dashed lines) at ±0.5 ms.

Table 2. Intercluster (k_{12}, k_{13}, k_{23}) and Intracluster (k_2, k_3) Exchange Rates As Defined in Figure 8 Used To Produce Relaxation Time Simulations in Figure 10[a]

k_{23}	k_{12}	k_{23}	k_3	k_2
{45°} ↔ {61°, 76°}	{45°} ↔ {115°, 132°}	{61°, 76°} ↔ {115°, 132°}	{61°} ↔ {76°}	{115°} ↔ {132°}
1.2×10^4 s^{-1}	4.0×10^4 s^{-1}	6.1×10^4 s^{-1}	3.4×10^4 s^{-1}	4.6×10^5 s^{-1}

[a]Local base motion parameters are the same as for Figure 9.

populations determined by RDC filtering and further corroborated by principal component analysis. The model assumed base-libration rates that range between 10^7 and 10^9 s^{-1}, and libration-rate-dependent jump amplitudes less than about 20°. The solid-state NMR parameters for the local base-libration parameters fall within the windows described above, consistent with the notion that solid-state experiments are able to capture the solution-state local motions accurately.

ii. General Rate Method. Although the slow exchange method provides an approximate scale for exchange processes, the general rate method could further resolve the values of the exchange rates between conformers. For ease of comparison, we used the local base motion rates and jump amplitudes from the slow-exchange fit, whose results were shown in Figure 9. Although several combinations of 5 PCA-inspired rates were found that could be fit to the 14 helical T_1 and 14 helical T_2 values with comparable RMSD's, in all cases the inter- and intracluster exchange rates were on the order of 10^4−10^5 s^{-1}, which confirms the validity of the slow exchange approximation. For example, the set of inter- and intracluster exchange rates in Table 2 yield the base relaxation times shown in Figure 10 with an RMSD for T_1 values of 5.2 ms and an RMSD for T_2 values of 1.4 ms. Interestingly, there is no clear distinction between the intercluster and intracluster rates, as may be expected from a significant difference in the free energy barriers.

To summarize, if we fix the local motion parameters to the best-fit set obtained using the slow-exchange method but float the conformational exchange rate parameters, we obtain a slight improvement in the quality of the fit. Best fits to experimental relaxations using general rate theory are achieved with inter- and intracluster exchange rates on the order of 10^4−10^5 s^{-1}, thus justifying the slow exchange approximation of the prior section. We did not float both local motion parameters and conformational exchange parameters simultaneously, but it is reasonable to assume that the results will be similar, especially

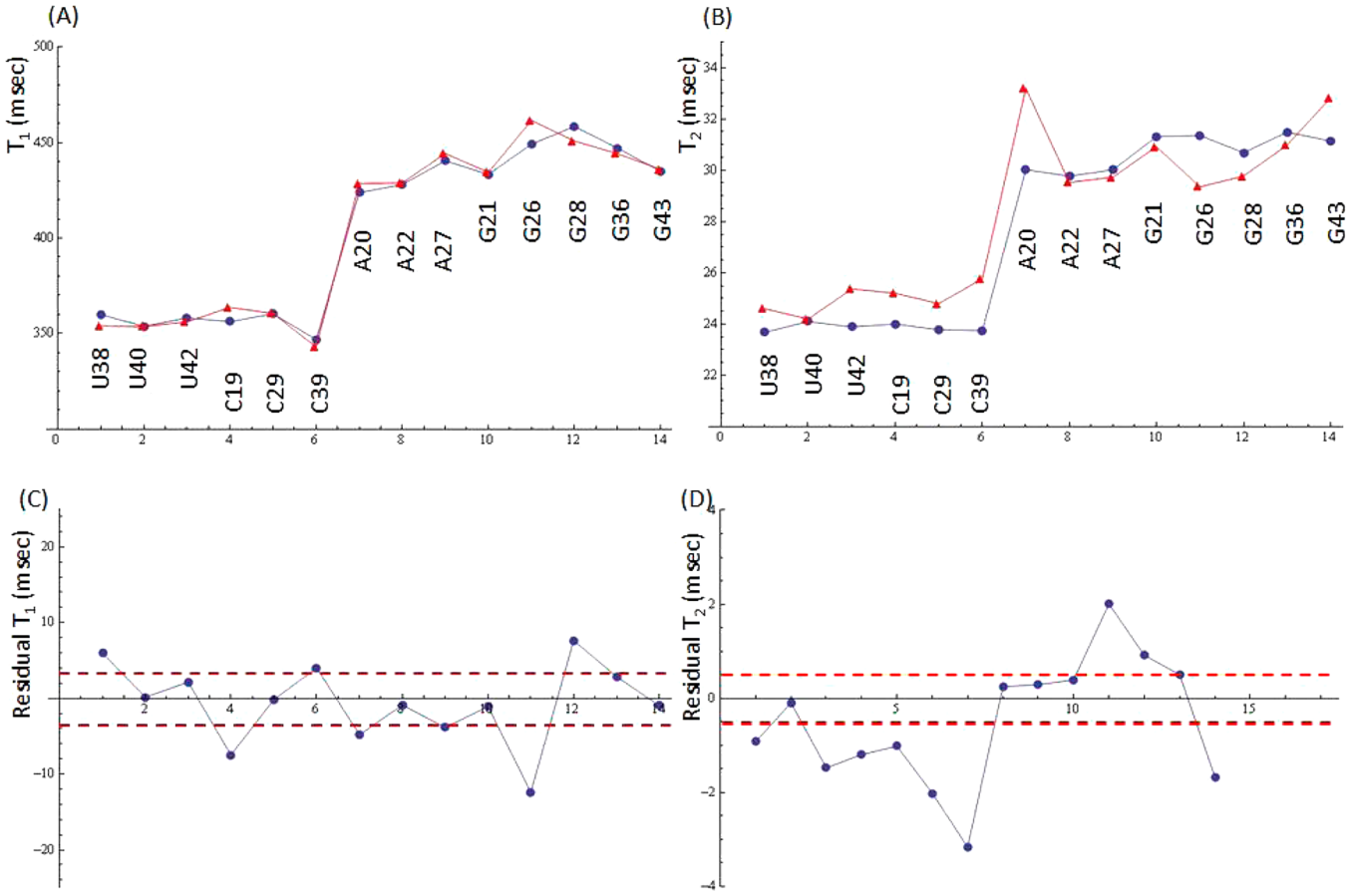


Figure 10. Relaxation time simulations for the C6−H6 (pyrimidine) and C8−H8 (purine) bonds using the general rate method with inter- and intracluster exchange rates from Table 2, and using the same local base motion parameters as for Figure 9: (A) T_1 simulations (blue circles) compared to the experimental T_1 values (red triangles); (B) T_2 simulations (blue circles) compared to the experimental T_2 values (red triangles); (C) difference between simulation T_1 and experimental T_1 values, together with the statistical error bars on experimental data (red dashed lines) at ±3.2 ms; (C) difference between simulated and experimental T_2 values, together with the statistical error bars on experimental data (red dashed lines) at ±0.5 ms.

Table 3. Relaxation Times for the Bulge C6−H6 (Pyrimidine) and C8−H8 (Purine) Bonds Simulated Using Both the Slow Exchange and General Rate Methods

residue	experimental T_1 (in ms)	experimental T_2 (in ms)	slow exchange T_1 (in ms)	slow exchange T_2 (in ms)	general rate T_1 (in ms)	general rate T_2 (in ms)
U23	328.3	29.1	322.5	24.3	321.9	24.3
C24	320.8	35.2	321.5	24.4	320.8	24.4
U25	no experimental values: assumed similar to C24		312.0	25.3	311.5	25.4

when we constrain the scale of the local base motion parameters to those observed under solid-state conditions.

iii. Bulge Residues. In addition to helical residues, we also simulated the relaxation times for the bulge residues using the slow exchange and general rate algorithms. In Table 3, we present relaxation time simulations assuming exchange between the conformations shown in Figure 4 using the slow exchange and general rate algorithms. The base rotation rate is selected as 5×10^8 s^{-1} and the amplitude as 15° (to yield good matches to the T_1 values). For the general rate simulations, inter- and intracluster exchange rates shown in Table 2 were assumed.

We obtain simulated T_1 values that are within 2% of experimental data, but simulated and experimental T_2 values for U23 deviate by about 16%; for C24, the relative deviation is even greater. This is likely due to the fact that the T_2 relaxation time has a spectral density component (the $J(0)$ term) that makes this observable sensitive to motions much slower than the Larmor frequencies of carbon-13 and protons (which are on the order of nanoseconds). The fact that we have been unable to capture the T_2 values may indicate that there are additional slower motions of these relatively underconstrained residues that are missing from the conformer set we have selected.

Although we have been able to successfully match most of the solution relaxation times to almost within the statistical error bars using the RDC-filtered conformer set, we must address the basic question, are the relaxation times sensitive to motions occurring at rates on the order of microseconds or slower? Though the T_1 time has no spectral density dependence slower than a time scale of a nanosecond, the T_2 are determined by slower motions as well and their expressions contain a dependence on the $J(0)$ spectral density. More importantly, the slow exchange and general rate methods depend on the fact that the time scales of rotational diffusion of

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

many molecules (including the TAR RNA considered here) overlap with the time scales to which both T_1 and T_2 are sensitive. Thus, two different conformers, with slightly different diffusion tensors will have different characteristic relaxation times when calculated separately. Even the slow exchange averaging process will result in a unique linear combination of relaxation rates that becomes discernible when enough data points are compared. The general rate theory loses some sensitivity to dynamics for rates much slower than a microsecond, but our least-squares simulations for relaxation times have shown that there is still discernible information to be gained at these time scales.

4. DISCUSSION

In this manuscript, we introduce a methodology based on energy-minimized structures that ties together structural and dynamic data, as well as solid-state and solution NMR, to build a dynamic trajectory for the HIV-1 TAR RNA to atomic-level detail. The protocol uses (a) solid-state NMR data to acquire information about local motions of the bases, (b) solution NMR RDC data to identify conformational states and their relative populations, (c) PCA analyses to identify degrees of freedom relevant to the overall reconfiguration of the molecule, as well as to corroborate the clustering of structures and choose parameters for the dynamics analysis, and (d) solution NMR relaxation time simulation techniques previously developed to simulate experimental data and fit the jump rates between the molecular conformers. The experimental data utilized covers a wide spectrum of motional time scales, from the picosecond scale for solution relaxation times and micro- to nanosecond times derived from solid-state NMR line shapes to the submillisecond time scale investigated with RDCs. The method reproduces relaxation data at multiple helical residues within the molecule using only five structures out of a set of just 500 possible conformers.

The advantage of this approach lies in the coverage of multiple time scales, including long time scales that are difficult to sample with MD, and the ease with which energy-minimized conformers may be obtained for small-to-midsized molecules using public-access software like the Rosetta suite of programs.[65] The use of these structures inverts the problem of dynamics relative to MD methods: instead of starting with an initial structure and running the clock forward from $t = 0$, we filter out long-lived structures using experimental data and interconnect them in a stochastic trajectory. Thus, the results from such a technique would prove valuable to corroborate MD-based simulations, and the protocol could provide a less computationally intensive alternative to extended molecular dynamic simulations.

We have been able to simulate relaxation times of most of the helical residues in the molecule with limited conformational sampling from an already small set of structures. The success in matching the experimental data indicates that, to gain an understanding of the gross motional properties of RNA, it is sufficient to sample the limited phase space of the particular structural motifs that constitute the molecule. This protocol was designed to derive a geometric picture of interhelical reorientation based on the limited conformational space available to the torsion angles in and around the bulge and hinge regions. The assumption that the helices are rigid and the requirement to close the loop formed by the single stranded bulge, adjacent helical base-pairs and the hinge backbone should restrict conformational possibilities for the entire molecule. Steric hindrances and limitations on the "stretchability" of the single-stranded region would further impose constraints on molecular reorientation.[45,63,64] In practice, we found that parameter space for the reorientation of one helix relative to the other was expanded by the possibility of one of the bulge adjacent base pairs opening up. Among the energy-minimized structures, a significant number had a missing A22-U40 base pair. The U40 base often forms a stabilizing interaction with U25 instead and, among the full set of 500 structures, sometimes with U23 and C24 as well. We allowed these possibilities to occur in our sample set to reflect fluctuations in the residue orientations, as well as the impacts of these fluctuations on overall molecular conformation. Thus, we believe that models generated using well-vetted potential energy functions can identify sites where new intramolecular bonding and conformational variability might occur.

Furthermore, we made a conscious choice to characterize the structural bins by their Euler angles. This choice of parameters has been made previously[33,66] to enhance reproducibility and comparability to other analyses of the molecule. Such a parametrization represents the core ingredient of most molecular analyses: reduction of the dimensions of the problem to render it tractable. Molecular studies often aim to distill out a few degrees of freedom that are implicated in determining either the structure or dynamics of the system, and many different techniques (Ramachandran plots,[67] phenomenological models,[4,5] PCA analyses[12,42]) are directed toward identifying a minimal set of relevant coordinates.

A. Assumptions in the Protocol. The methodology relies on the assumption that energy-minimized structures sufficiently populate the available conformational space, i.e., on the assumption of ergodicity. If ergodic behavior holds, then a sufficiently representative characterization of the energy landscape of the phase space will allow a calculation of the requisite time averages of observables. In the case of the TAR RNA, the structural motif (helix−bulge−helix) is fairly simple, and it is possible to cover a large region of the interhelical orientation space with a relatively small number of structures. For more complex structural motifs, it would be necessary to generate a sample set that covers both the space of molecular reorientations and the range of conformations of local residues relative to a fixed large-scale molecular orientation. Even in the current work, it is possible that we have under-sampled the full range of conformations available to the bulged loop. A richer sample of both interhelical orientations and orientations of bulge residues relative to particular interhelical orientations may improve the RDC fit.

A second assumption is that the slow exchange (SE) and general rate (GR) methods assume coincidence of the diffusion tensors at the moment of exchange. This is not a significant problem for conformers that are not significantly different, but it could pose problems for conformers that are widely separated in conformational space. We have not attempted to quantify the actual deviation in relaxation times on account of this assumption. This issue could be addressed in combination with molecular dynamics simulations.

Finally, it bears mentioning that there is a 2-fold degeneracy in the choice of the unit eigenvectors of the rotational diffusion tensors, with the negative of a given choice of unit vector being acceptable as the eigenvector as well. The choice of eigenvectors does not change the results for the slow exchange formalism (the expressions are invariant to such changes) but does impact the general rate theory expressions. For example,

keeping the z-axis the same, the two choices of a right-handed coordinate system vary by 180° and would artificially introduce such an extra jump into the calculations. The means of consistently dealing with such a jump is to track the diffusion tensors as a function of changes in the shape of the molecule, either visually or geometrically and ensure that there is no additional change in the diffusion tensor orientations due to axis inversions. In our particular case, we tested our calculations by artificially inverting the orientation of the diffusion tensors of some of the structures and found very small changes in the relaxation times as a result: a change of at most ∼0.25 ms in T_1 and at most ∼0.02 ms in T_2. These changes would not significantly impact the conclusions of this manuscript and a detailed analysis is omitted here.

B. Comparison with Previous Results. We examined the five structures selected from the RDC-filter in light of unbound TAR RNA structures generated using NOE constraints (but not RDC constraints) by one of the authors.[47] These 20 energy-minimized structures, termed the TAR2013 series, have β angles in the range 38−59°, but the five lowest energy structures cover a smaller range of 45−53°. This range corresponds well to the bend angle of highest population structure we have obtained (i.e., β = 45°). In fact, the lowest energy TAR2013 structure has an α value of 202° and a γ value of 225°, similar to the "45° structure" in our current analysis.

We previously published two studies where the new methods developed to simulate relaxation times[6,7,55] were applied to U38 relaxation data and used to select regions of interhelical motional parameter space that fit the data. The models consisted of two-site jumps between the lowest energy structure of 1ANR and structures artificially modified from that structure to reflect changes in interhelical orientation. The closest approximation to a two-site jump in this manuscript is found by considering only exchanges between pairs of conformers within the three most populated structures (the 45° structure, the 61° structure, and the 115° structure, which are almost equal in population). The exchange between the 45° and the 61° structures involves a bend angle modification of 16°, and a twist angle about the upper helix of 2°. Cross-checking this parameter set against the results of applying the general rate theory,[7,55] we find that the U38 data was fit by a two-site jump model with a twist of 0° and bend angle between 5° and 12° (among other possible models). Thus, the 45° and 61° structures fit the profiles of two structures selected previously on the basis of the U38 data alone. Exchanges between either of these structures and the 115° structure, however, are of a magnitude not simulated in the previous studies.

We can also make a few basic comparisons to the results of Dayie et al.,[43] even though the authors consider the HIV-2 TAR RNA, as both molecules consist of a helix−bulge−helix motif. Given our focus on the C6 and C8 atoms in the current work, we first observe a similarity in the fact that in the relaxation data used in this work (from Bardaro et al.[30]) the helical residues seem to have similar ^{13}C (C6 and C8) T_1 relaxation times, a fact observed in the work of Dayie et al. as well. However, the U23 ^{13}C $T_{1\rho}$ is nearly half the value of the corresponding U25 time in Dayie et al., whereas Bardaro et al. report values much closer in magnitude. Also, the magnitudes of the T_1 and $T_{1\rho}$ relaxation times in the two papers are different. We mention these facts to bring up three relevant considerations in interpreting the results of the two different sets of experiments: (a) the obvious difference made by the presence of only two residues in the bulge in HIV-2 RNA versus three in the HIV-1 RNA; (b) the fact that the relaxation times of Dayie et al. include a component from the C5−C6 dipolar coupling for pyrimidines, whereas this coupling is explicitly suppressed in Bardaro et al. (see Shajani and Varani[68]); (c) the use of a model-free analysis in Dayie et al. versus the SE and GR methods used herein. Notwithstanding these caveats, the common results we can extract are that both papers observe significant flexibility in the U23 and U25 residues and rigid, slow motions in the helices.

C. Comparison with Extended MD Simulations. To examine the extent to which our approach matches results obtained by the MD approach, we compare our results to those of Salmon et al.,[18] where the authors describe the selection of conformers from 8.2 μs MD simulations of the TAR RNA using Pf1 phage-aligned RDC data sets. It was reported that the best fit to RDC data is obtained with a set of 20 conformations selected from the full MD ensemble. Though it is not possible to compare the absolute values of the interhelical bend and twist angles due to differing methods of characterizing the helices and their relative orientations, we can compare the spans of the angles reported for their ensemble to those in ours. The bend angles in the Salmon et al. ensemble of 20 span 88° (from 3° to 91°) whereas those in our ensemble span 87°, the rotations of the upper helix about the lower helical axis span 191° in their ensemble whereas those in our set span 215°, and the rotations of the upper helix about its own symmetry axis span 224° in their ensemble and 210° in ours. Thus the span of angles obtained by the two approaches are in excellent agreement. Moreover, the full, prefilter ensembles in both papers show correlations between the α and γ twist angles. A more fine-grained comparison relates to the behavior of individual residues. We have already mentioned that the A22-U40 base pair is often found to be open among the full set of 500 structures. We also find that, among our five RDC-filtered structures, four (the 45°, 76°, 115°, and 132° structures) lack a A22-U40 base pair. However, the G26-C39 base pair is maintained in all five of these structures. Salmon et al. find a similar asymmetry between the A22-U40 base pair and the G26-C39 base pairs in their RDC-selected ensemble, with the former adopting a broader conformational distribution and the latter being in an A-form helix-like conformation.

Differences are nonetheless observed with regards to the bulge conformation. Salmon et al. report the occurrence of three clusters within their RDC-selected ensemble: a 66% population cluster with A22 stacked on U23, a 19% population cluster with U23 flipped out, and a 15% cluster with paired U25-U40 and unpaired U23 and C24. With regard to the third cluster, nearly 30% of the 500 energy-minimized structures used in our analysis show a U25-U40 pair, with three of the RDC-filtered structures (the 45°, 76°, and 132° structures) included in this list as well. The 115° structure simply lacks the A22-U40 pair and does not have any alternative pairings of either residue. Salmon et al. stated that the U25-U40 pair is predicted to be the second most energetically favorable bulge conformation in MC-fold. A visual inspection of our structures shows the following behavior for the bulge:

(1) The 45° structure has U23 flipped into the interhelical space but not stacked, C24 is flipped in but not stacked and U25 is paired with U40.
(2) The 61° structure has U23 stacked on A22 and has C24 and U25 flipped out.

dx.doi.org/10.1021/jp409386t | J. Phys. Chem. B 2014, 118, 1726−1742

Table 4. Comparison of Rotational Diffusion Tensor Eigenvalues As Calculated by HYDRONMR and BEST

structure	D_{r1} (BEST) (10^7 s^{-1})	D_{r1} (HYDRONMR) (10^7 s^{-1})	D_{r2} (BEST) (10^7 s^{-1})	D_{r2} (HYDRONMR) (10^7 s^{-1})	D_{r3} (BEST) (10^7 s^{-1})	D_{r3} (HYDRONMR) (10^7 s^{-1})
45°	1.764	1.826	1.807	1.873	3.523	3.805
61°	1.549	1.617	1.591	1.658	3.176	3.428
76°	2.460	2.503	2.506	2.561	3.323	3.426
115°	2.119	2.152	2.134	2.175	3.415	3.553
132°	1.783	1.789	1.807	1.809	3.310	3.407

(3) The 76° structure has U23 flipped out, C24 stacked with U25, and U25 paired with U40.
(4) The 115° structure has U23 flipped out, C24 flipped in and stacked close to U25, and U25 flipped in and stacked only with C24.
(5) The 132° structure has U23 and C24 flipped in but not stacked and has U25 paired with U40.

Thus, only one of the structures in our ensemble (the 61° structure with a 27% population) has a significant A22-U23 stacking interaction and two have U23 flipped out (39% total population), a clear deviation from the results of Salmon et al., suggesting that the conformational variability of this region is more than can be captured by a small number of sampled structures. A solution to this problem is to generate energy-minimized structures where the interhelical orientation is fixed (or nearly so) and the bulge flexibility is evaluated under the constraint of fixed end points. Such an analysis would establish the inherent conformational flexibility of the bulge.

D. Confirming Hydrodynamics Calculations. To cross-check the results of the program HYDRONMR, we recalculated the diffusion tensors using the program BEST,[69] which tessellates the solvent-accessible surface of the molecule and calculates the various diffusion properties using a finite element analysis. The molecule was uniformly hydrated to a hydration shell thickness of 1.1 Å.[70] The eigenvalues of the rotational diffusion tensors (in ascending order) of the two programs are compared in Table 4.

We find that the diffusion tensor eigenvalues as found by BEST were uniformly smaller than those found by HYDRONMR, indicating that HYDRONMR underestimated the hydration effect relative to BEST. The unit eigenvectors were very similar between the two programs. We subsequently calculated the relaxation times using the slow exchange formalism and found a T_1 shift of at most 18 ms and a T_2 shift of at most 1.2 ms, corresponding to a shift of about 4% of the experimental values for both relaxation times. This may modify the choice of parameters described above, but we believe that the impact will not be substantial.

5. CONCLUSIONS

We have carried out a characterization of the essential dynamics of the TAR RNA molecule using techniques with time scale sensitivities ranging from subnanosecond (solid-state and solution relaxation times) to millisecond (RDCs). We have been able to capture the long-time scale behavior of the conformational exchange processes that characterize this molecule and fit experimental relaxation times very well, with exchanges between discrete conformers occurring at time scales longer than 1 μs. The similarities of results of this method with those of extended MD simulations provide independent corroboration of our conformational analysis. Further computational explorations and sample-size increases will enhance the results obtained by this methodology.

AUTHOR INFORMATION

Corresponding Author

*G. Drobny: e-mail, drobny@chem.washington.edu; tel, 1-206 685 2051; fax, 1-206 685 8665.

Present Address

[‖]Department of Chemistry, Arizona State University, Box 871604, Tempe, AZ 85287-1604.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (RO1-GM109417) and NSF (MCB-0642253) to G.P.D. and G.V. We thank Professor Rhiju Das and Dr. Parin Sripakdeevong of the Department of Biophysics at Stanford University for providing the FARFAR structures and consistently offering helpful advice.

REFERENCES

(1) Korzhnev, D. M.; Kay, L. E. Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Acc. Chem. Res.* **2008**, *41* (3), 442−451.
(2) Havlin, R. H.; Tycko, R. Probing Site-Specific Conformational Distributions in Protein Folding with Solid-State NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (9), 3284−3289.
(3) Hu, K. N.; Yau, W. M.; Tycko, R. Detection of a Transient Intermediate in a Rapid Protein Folding Process by Solid-State Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **2010**, *132* (1), 24−25.
(4) Olsen, G. L.; Bardaro, M. F. J.; Echodu, D. C.; Drobny, G. P.; Varani, G. Intermediate Rate Atomic Trajectories of RNA by Solid-State NMR. *J. Am. Chem. Soc.* **2010**, *132* (1), 303−308.
(5) Echodu, D.; Goobes, G.; Shajani, Z.; Pederson, K.; Meints, G.; Varani, G.; Drobny, G. Furanose Dynamics in the HhaI Methyltransferase Target DNA Studied by Solution and Solid-State NMR Relaxation. *J. Phys. Chem. B* **2008**, *112* (44), 13934−13944.
(6) Emani, P. S.; Olsen, G. L.; Echodu, D. C.; Varani, G.; Drobny, G. P. Slow Exchange Model of Nonrigid Rotational Motion in RNA for Combined Solid-State and Solution NMR Studies. *J. Phys. Chem. B* **2010**, *114* (48), 15991−16002.
(7) Emani, P. S.; Olsen, G. L.; Varani, G.; Drobny, G. P. Theory of Non-Rigid Rotational Motion Applied to NMR Relaxation in RNA. *J. Phys. Chem. A* **2011**, *115* (44), 12055−12069.
(8) Amber Home Page. http://ambermd.org (accessed May 5, 2013).
(9) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.
(10) Cheatham, T. E.; Young, M. A. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers* **2001**, *56* (4), 232−256.
(11) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Protein Simulations* **2003**, *66*, 27−85.
(12) Riccardi, L.; Nguyen, P. H.; Stock, G. Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis. *J. Phys. Chem. B* **2009**, *113* (52), 16660−16668.

(13) CHARMM. www.charmm.org (accessed May 5, 2013).
(14) 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* (2), 187−217.
(15) Horton, J. R.; Ratner, G.; Banavali, N. K.; Huang, N.; Choi, Y.; Maier, M. A.; Marquez, V. E.; MacKerell, A. D., Jr.; Cheng, X. Caught in the Act: Visualization of an Intermediate in the DNA Base-Flipping Pathway Induced by HhaI Methyltransferase. *Nucleic Acids Res.* **2004**, *32* (13), 3877−3886.
(16) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics. *Nature* **2005**, *433* (7022), 128−132.
(17) Do, T. N.; Ippoliti, E.; Carloni, P.; Varani, G.; Parrinello, M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. *J. Chem. Theory Comput.* **2012**, *8* (2), 688−694.
(18) Salmon, L.; Bascom, G.; Andricioaei, I.; Al-Hashimi, H. M. A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. *J. Am. Chem. Soc.* **2013**, *135* (14), 5457−5466.
(19) Li, Z.; Scheraga, H. A. Monte Carlo-Minimization Approach to the Multiple-Minima Problem in Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (19), 6611−6615.
(20) Bradley, P.; Baker, D. Improved Beta-Protein Structure Prediction by Multilevel Optimization of NonLocal Strand Pairings and Local Backbone Conformation. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (4), 922−929.
(21) Das, R.; Baker, D. Automated De Novo Prediction of Native-like RNA Tertiary Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (37), 14664−14669.
(22) Das, R.; Karanicolas, J.; Baker, D. Atomic Accuracy in Predicting and Designing Noncanonical RNA Structure. *Nat. Methods* **2010**, *7* (4), 291−294.
(23) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G. H.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; et al. NMR Structure Determination for Larger Proteins Using Backbone-Only Data. *Science* **2010**, *327* (5968), 1014−1018.
(24) Lyskov, S.; Chou, F.-C.; Conchuir, S. O.; Der, B. S.; Drew, K.; Kuroda, D.; Xu, J.; Weitzner, B. D.; Renfrew, P. D.; Sripakdeevong, P.; et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE* **2013**, *8* (5), e63906.
(25) Parisien, M.; Major, F. The MC-Fold and MC-Sym Pipeline Infers RNA Structure from Sequence Data. *Nature* **2008**, *452* (7183), 51−55.
(26) Aboul-ela, F.; Karn, J.; Varani, G. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **1996**, *24* (20), 3974−3981.
(27) Aboul-ela, F.; Karn, J.; Varani, G. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* **1995**, *253* (2), 313−332.
(28) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* **1992**, *257* (5066), 76−80.
(29) Long, K. S.; Crothers, D. M. Characterization of the Solution Conformations of Unbound and Tat Peptide-Bound Forms of HIV-1 TAR RNA. *Biochemistry* **1999**, *38* (31), 10059−10069.
(30) Bardaro, M. F. J.; Shajani, Z.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. How Binding of Small Molecule and Peptide Ligands to HIV-1 TAR Alters the RNA Motional Landscape. *Nucleic Acids Res.* **2009**, *37* (5), 1529−1540.
(31) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. Resolving the Motional ModesThat Code for RNA Adaptation. *Science* **2006**, *311* (5761), 653−656.
(32) Hansen, A. L.; Al-Hashimi, H. M. Dynamics of Large Elongated RNA by NMR Carbon Relaxation. *J. Am. Chem. Soc.* **2007**, *129* (51), 16072−16082.
(33) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. Visualizing Spatially Correlated Dynamics that Directs RNA Conformational Transitions. *Nature* **2007**, *450* (7173), 1263−1267.
(34) 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* (27), 7947−7956.
(35) Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T. Structure of the K-turn U4 RNA: a Combined NMR and SANS Study. *Nucleic Acids Res.* **2010**, *38* (18), 6274−6285.
(36) Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. Alpha Helix-RNA Major Groove Recognition in an HIV-1 REV Peptide-RRE RNA Complex. *Science* **1996**, *273* (5281), 1547−1551.
(37) Zemmel, R. W.; Kelley, A. C.; Karn, J.; Butler, P. J. Flexible Regions of RNA Structure Facilitate Co-Operative Rev Assembly on the Rev-Response Element. *J. Mol. Biol.* **1996**, *258* (5), 763−777.
(38) Frank, A. T.; Stelzer, A. C.; Al-Hashimi, H. M.; Andricioaei, I. Constructing RNA Dynamical Ensembles by Combining MD and Motionally Decoupled NMR RDCs: New Insights into RNA Dynamics and Adaptive Ligand Recognition. *Nucleic Acids Res.* **2009**, *37* (11), 3670−3679.
(39) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **2008**, *320* (5882), 1471−1475.
(40) Al-Hashimi, H. M. NMR Studies of Nucleic Acid Dynamics. *J. Magn. Reson.* **2013**, *237*, 191−204.
(41) Altis, A.; Otten, M.; Nguyen, P. H.; Hegger, R.; Stock, G. Construction of the Free Energy Landscape of Biomolecules via Dihedral Angle Principal Component Analysis. *J. Chem. Phys.* **2008**, *128* (24), 245102.
(42) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins* **2005**, *58* (1), 45−52.
(43) Dayie, K. T.; Brodsky, A. S.; Williamson, J. R. Base Flexibility in HIV-2 TAR RNA Mapped by Solution 15N, 13C NMR Relaxation. *J. Mol. Biol.* **2002**, *317* (2), 263−278.
(44) Olsen, G. L.; Bardaro, M. F. J.; Echodu, D. C.; Drobny, G. P. Hydration Dependent Dynamics in RNA. *J. Biomol. NMR* **2009**, *45* (1−2), 133−142.
(45) Bailor, M. H.; Mustoe, A. M.; Brooks, C. L., 3rd; Al-Hashimi, H. M. Topological Constraints: Using RNA Secondary Structure to Model 3D Conformation, Folding Pathways, and Dynamic Adaptation. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 296−305.
(46) Lu, X. J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108−5121.
(47) Bardaro, M. F., Jr.; Varani, G. Independent Alignment of RNA for Dynamic Studies Using Residual Dipolar Couplings. *J. Biomol. NMR* **2012**, *54* (1), 69−80.
(48) Zweckstetter, M.; Bax, A. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000**, *122* (15), 3791−3792.
(49) Ruckert, M.; Otting, G. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* **2000**, *122* (32), 7793−7797.
(50) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. Precision Neutron-Diffraction Structure Determination of Protein and Nucleic-Acid Components 0.12. Study of Hydrogen-Bonding in Purine-Pyrimidine Base Pair 9-Methyladenine . 1-Methylthymine. *J. Chem. Phys.* **1973**, *59* (2), 915−924.
(51) Ying, J. F.; Grishaev, A.; Bryce, D. L.; Bax, A. Chemical Shift Tensors of Protonated Base Carbons in Helical RNA and DNA from NMR Relaxation and Liquid Crystal Measurements. *J. Am. Chem. Soc.* **2006**, *128* (35), 11443−11454.

(52) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. Tables of Bond Lengths Determined by X-ray and Neutron Diffraction. Part 1. Bond Lengths in Organic Compounds. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1−S19.
(53) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(54) Murray, L. J.; Arendall, W. B., 3rd; Richardson, D. C.; Richardson, J. S. RNA Backbone is Rotameric. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13904−13909.
(55) Emani, P. S.; Olsen, G. L.; Varani, G.; Drobny, G. P. Correction to "Theory of Nonrigid Rotational Motion Applied to NMR Relaxation in RNA". *J. Phys. Chem. A* **2012**, *116* (26), 7253−7260.
(56) Ryabov, Y.; Clore, G. M.; Schwieters, C. D. Coupling between Internal Dynamics and Rotational Diffusion in the Presence of Exchange between Discrete Molecular Conformations. *J. Chem. Phys.* **2012**, *136* (3), 034108.
(57) Peng, J. W.; Thanabal, V.; Wagner, G. 2d Heteronuclear NMR Measurements of Spin-Lattice Relaxation-Times in the Rotating Frame of X Nuclei in Heteronuclear HX Spin Systems. *J. Magn. Reson.* **1991**, *94* (1), 82−100.
(58) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magn. Reson.* **2000**, *147* (1), 138−146.
(59) Fernandes, M. X.; Ortega, A.; Martinez, M. C. L.; de la Torre, J. G. Calculation of Hydrodynamic Properties of Small Nucleic Acids from their Atomic Structure. *Nucleic Acids Res.* **2002**, *30* (8), 1782−1788.
(60) Millero, F. J.; Dexter, R.; Hoff, E. Density and Viscosity of Deuterium Oxide Solutions from 5−70°C. *J. Chem. Eng. Data* **1971**, *16* (1), 85−87.
(61) Yuwen, T.; Post, C. B.; Skrynnikov, N. R. Domain Cooperativity in Multidomain Proteins: What Can We Learn from Molecular Alignment in Anisotropic Media? *J. Biomol. NMR* **2011**, *51* (1−2), 131−150.
(62) Hansen, M. R.; Mueller, L.; Pardi, A. Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions. *Nat. Struct. Biol.* **1998**, *5* (12), 1065−1074.
(63) Bailor, M. H.; Sun, X.; Al-Hashimi, H. M. Topology Links RNA Secondary Structure with Global Conformation, Dynamics, and Adaptation. *Science* **2010**, *327* (5962), 202−206.
(64) Mustoe, A. M.; Bailor, M. H.; Teixeira, R. M.; Brooks, C. L. 3rd; Al-Hashimi, H. M. New Insights into the Fundamental Role of Topological Constraints as a Determinant of Two-Way Junction Conformation. *Nucleic Acids Res.* **2012**, *40* (2), 892−904.
(65) RosettaCommons. www.rosettacommons.org (accessed May 7, 2013).
(66) Zhang, Q.; Al-Hashimi, H. M. Domain-Elongation NMR Spectroscopy Yields New Insights into RNA Dynamics and Adaptive Recognition. *RNA* **2009**, *15* (11), 1941−1948.
(67) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95−99.
(68) Shajani, Z.; Varani, G. NMR Studies of Dynamics in RNA and DNA by 13C Relaxation. *Biopolymers* **2007**, *86* (5−6), 348−359.
(69) Aragon, S. A Precise Boundary Element Method for Macromolecular Transport Properties. *J. Comput. Chem.* **2004**, *25* (9), 1191−1205.
(70) Aragon, S.; Hahn, D. K. Precise Boundary Element Computation of Protein Transport Properties: Diffusion Tensors, Specific Volume, and Hydration. *Biophys. J.* **2006**, *91* (5), 1591−1603.