THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

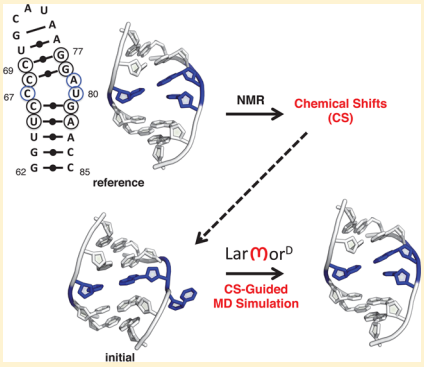
A Simple and Fast Approach for Predicting ^{1}H and ^{13}C Chemical Shifts: Toward Chemical Shift-Guided Simulations of RNA

Aaron T. Frank,* Sean M. Law, and Charles L. Brooks, III*

Department of Chemistry and Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, United States

Supporting Information

ABSTRACT: We introduce a simple and fast approach for predicting RNA chemical shifts from interatomic distances that performs with an accuracy similar to existing predictors and enables the first chemical shift-restrained simulations of RNA to be carried out. Our analysis demonstrates that the applied restraints can effectively guide conformational sampling toward regions of space that are more consistent with chemical shifts than the initial coordinates used for the simulations. As such, our approach should be widely applicable in mapping the conformational landscape of RNAs via chemical shift-guided molecular dynamics simulations. The simplicity and demonstrated sensitivity to three-dimensional structure should also allow our method to be used in chemical shift-based RNA structure prediction, validation, and refinement.



INTRODUCTION

The recent realization of the important role played by ribonucleic acids (RNAs) in regulating cellular processes[1,2] has resulted in significant interest in characterizing the structure of these molecules at atomic resolution. However, RNAs possess significant conformational flexibility, which complicates structure determination via X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. From a biophysical standpoint, this conformational flexibility has significant mechanistic implications on RNA function. For instance, in the context of molecular recognition, a number of RNAs, most notably the HIV-1 trans-activating response (TAR) element RNA, bind ligands via a conformational capture mechanism in which the ligand-free RNA samples a number of distinct "bound-like" states.[3−6] In the case of HIV-1 TAR, the conformations of the RNA in complex with a number of ligands closely resemble some of the conformations sampled in the ligand-free state.[6,7] As such, instead of a static RNA structure, an ensemble representation that captures the entire range of accessible conformations, along with their associated population weights, is needed.

In principle, molecular dynamics (MD) simulations can be used to map the landscape of any biomolecule in a thermodynamically rigorous manner. However, the force fields used to simulate the dynamics of biomolecules in MD simulations are imperfect. This is particularly true for RNA force fields, which, due to the protein-centric view that has been prevalent up to this point, have been the subject of less development than their protein counterparts. The use of experimentally guided simulations has emerged as an alternative to the often tedious and time-consuming reparameterizations of force fields. In these guided simulation approaches, the force field is augmented with a biasing term that ensures that the system being simulated matches some experimental observable(s).[8−12] The use of experimentally guided simulations to accurately map the conformational landscape of proteins is now well-established.[13−20]

Of particular interest are approaches that use NMR chemical shifts to guide MD simulations of biomolecules. Chemical shifts have emerged as an attractive source of structural information due to the fact that, in addition to being readily accessible and being the most *precisely* measured NMR observable, they exhibit exquisite sensitivity to structure.[21−26] Advances in site-specific labeling techniques[27−30] and automated assignments approaches[31−35] means that chemical shifts are, and will continue to become, increasingly accessible, not only for the small RNAs (<40 nt) typically studied using NMR, but for larger and more complex RNAs as well.[30,36]

A prerequisite for the use of chemical shifts-guided simulations to map the conformational landscape of a biomolecule is the availability of structure-based approaches that allow chemical shifts to be predicted from the three-dimensional (3D) coordinates of the molecule. In this regard, empirical methods have been shown to be of great utility. In the case of proteins, a plethora of such empirical structure-based chemical shift predictors have been developed.[37−42] In contrast, only a few such methods exist for RNA. Two of these methods, SHIFTS[43] and NUCHEMICS,[44] predict nonexchangeable ^{1}H chemical shifts, while the recently described RAMSEY[45] is capable of predicting both nonexchangeable ^{1}H and protonated

Received: August 18, 2014
Revised: September 24, 2014
Published: September 25, 2014

^{13}C chemical shifts. In principle, the ability to predict both ^{1}H and ^{13}C chemical shifts makes RAMSEY an ideal predictor to be used in chemical shifts-guided simulations. However, RAMSEY, which was developed using the random forest approach, is a "black box" predictor, which automatically precludes its direct incorporation into MD simulations because no closed-form analytical solution can be obtained.

As a first step toward using chemical shifts to guide MD simulations of RNA, we report on the development of LARMOR^{D} (LARMOR for the Larmor frequency (ω) and D for distance), a simple distance-dependent chemical shift predictor that allows chemical shifts to be *easily* incorporated into molecular simulations of RNA. In what follows we: (i) describe the model and the approach used to parametrize LARMOR^{D}; (ii) assess the accuracy of LARMOR^{D}; (iii) demonstrate the sensitivity of LARMOR^{D} predicted chemical shifts to RNA 3D-structure; and (iv) apply LARMOR^{D} enabled chemical-shift guided MD simulations (CS-MD) to a model RNA system. Collectively, our results indicate that in addition to its simplicity and speed, LARMOR^{D} is accurate, sensitive to RNA 3D structure, and enables effective biasing restraints to be derived from experimental NMR chemical shifts.

METHODS AND MATERIAL

The Chemical Shift-Structure Database. To generate LARMOR^{D}, we compiled a training set consisting of RNAs for which NMR structures were available via the Protein Data Bank (PDB: http://www.pdb.org) and chemical shifts were available via either the Biological Magnetic Resonance Bank (BMRB: http://www.bmrb.wisc.edu/) or the literature (Supporting Information, Table S1). Excluded from the database were RNAs: (i) that were known and verified to contain systematic ^{13}C referencing errors;[45,46] (ii) whose corresponding chemical shifts were assigned at temperatures <290 K; (iii) that were bound to small molecules and/or proteins; and (iv) that contained modified residues. In total, the compiled chemical shift-structure training set contained data from 35 RNAs. In addition to the training set, a testing set was compiled. The testing set consists of chemical shifts and structures for 28 RNAs, 17 of which were known and verified to contain ^{13}C referencing errors (Supporting Information, Table S2). The testing set served two purposes: (i) it allowed us to test whether the LARMOR^{D} predictors, specifically, the ^{13}C predictors, were accurate enough to detect known systematic referencing errors, and (ii) after correcting for identified systematic referencing errors, it allowed us to independently assess the accuracy of *both* ^{1}H and ^{13}C chemical shift predictors. For both data sets, we carried out outlier analysis by examining the distributions of reported chemical shifts for each unique nucleus (i.e., each unique combination of ^{1}H and ^{13}C nuclei and residue types). Outliers were identified as entries that were greater than the median by more than three standard deviations (i.e., the 3σ rule). The outliers comprised <1% of the combined data set and, for completeness, the entries identified as outliers are listed in Supporting Information, Table S3. The final training set (excluding outliers) contained 5505 and 2924 ^{1}H and ^{13}C chemical shifts, respectively, and the testing set (excluding outliers) contained 5520 and 3745 ^{1}H and ^{13}C chemical shifts, respectively.

For each RNA entry in the training and testing sets, the interatomic distances, r_{ij} (see eq 1), that were to be used to predict chemical shifts were extracted from the NMR bundle of the corresponding RNA using the following approach. First, to generate a representative model, the member of the NMR bundle that was closest (i.e., had smallest structural root-mean-square deviation (RMSD)) to the average structure of the bundle was selected and then briefly minimized using the steepest-descent gradient method. For every nucleus for which ^{1}H or ^{13}C chemical shifts were available, the distance between that nucleus and all heavy atoms were measured from the coordinates of the representative model. All distances were determined using the MDAnalysis[47] python module.

Distance-Based Chemical Shift Prediction Model. In our model, the chemical shift, δ_i^{pred}, for a given RNA nucleus i, is expressed as a function of interatomic distances,[40,41] that is:

$$\delta_i^{\mathrm{pred}} = \delta_i^{\mathrm{ref}} + \sum_{j=1}^{N} \alpha_{ij} r_{ij}^{-3}. \quad (1)$$

Here, δ_i^{ref} is the reference chemical shift for nucleus i, N is the total number of heavy atoms in the RNA, r_{ij} is the interatomic distance between atoms i and j, and α_{ij} is a parameter that depends on the atom type of i, and the atom and residue type associated with j (Figure 1). For each nonexchangeable ^{1}H (H1′, H2′, H3′, H4′, H5′, H5″, H2, H5, H6, H8) and protonated ^{13}C (C1′, C2′, C3′, C4′, C5′, C2, C5, C6, C8) nucleus, the set $\{\alpha_{ij}\}$ that minimized the objective function χ^2, which quantifies the error between measured and predicted chemical shifts, was determined using a genetic algorithm (GA) optimization approach. Here,

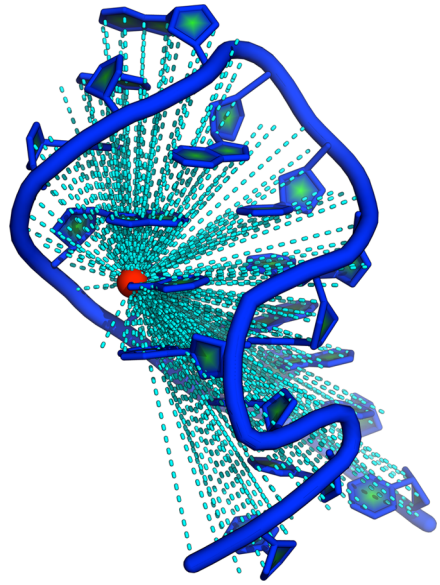


Figure 1. Illustration of the approach used to predict RNA chemical shifts. For a given nucleus (red sphere), chemical shifts are predicted as a function of the distances between that nucleus and all heavy atoms in the RNA (cyan lines; eq 1).

$$\chi^2 = \frac{1}{N_{\mathrm{CS}}} \sum_{i=1}^{N_{\mathrm{CS}}} w_i (\delta_i^{\mathrm{meas}} - \delta_i^{\mathrm{pred}})^2 \quad (2)$$

where δ_i^{meas} and w_i are the measured chemical shift and weighting factor, respectively, and for a given nucleus i. The summation runs over the set of N_{CS} chemical shifts in the training set. All GA optimizations were carried out with a population size of 10, and the number of evolution cycles was

dx.doi.org/10.1021/jp508342x | J. Phys. Chem. B 2014, 118, 12168−12175

set to 4000 using the Pyevolve python module.[48] Each GA optimization was initialized with $\{\alpha_{ij}\} = 0 \forall j$ (i.e., for some nucleus i, $\delta_i^{pred} = \delta_i^{ref}$ (eq 1)) and $w_i = 1 \forall i$ throughout the optimization. Using these GA settings, all optimizations were converged, and the fitted errors ($(\chi^2)^{1/2}$) were near the expected ranges; for protons and carbons, the average $(\chi^2)^{1/2}$ was ∼0.19 and 1.09 ppm, respectively. By comparison RAMSEY has prediction errors of ∼0.16 and 0.90 ppm for protons and carbons, respectively.[45]

Chemical Shift-Guided MD Simulation. To carry out chemical shift-guided MD simulations (CS-MD), LARMOR^D was implemented into the CHARMM macromolecular mechanics package.[49] CS-MD simulations were then carried out using the hybrid energy approach[50] in which the total energy associated with a conformer $\boldsymbol{X}$, $E(\boldsymbol{X})$, is given by

$$E(\boldsymbol{X}) = \frac{N_{CS}}{2\beta}\log \chi^2(\boldsymbol{X}) + E_{CHARMM}(\boldsymbol{X}). \quad (3)$$

Here, $E_{CHARMM}(\boldsymbol{X})$ and β are the CHARMM force-field energy of $\boldsymbol{X}$ and $1/k_BT$ (k_B = Boltzmann constant), respectively, and χ^2 is the expression noted above (see eq 2). In this case, however, the weighting factors (w_i) were used to account for the differential accuracy of the predictors. Specifically,

$$w_i = \frac{1}{MAE_i} \quad (4)$$

where MAE_i is the estimated mean absolute error between measured and LARMOR^D predicted chemical shifts for the nucleus type associated with i. Here the MAE_i were estimated using the testing set. In addition to accounting for the differential accuracy of the predictors, w_i scales the error such that nuclei with different dynamic ranges can contribute similarly to the χ^2 (for example ^{1}H and ^{13}C nuclei). The hybrid energy with a log-harmonic restraint term was introduced by Habeck et al. and is a Bayesian-inspired marginal hybrid energy[51] that has the special feature that it does not include a "force constant" scaling factor for the restraint term—in typical hybrid energy schemes, an ad hoc force constant that scales the contribution of the restraint term relative to the physical energy (here $E_{CHARMM}(\boldsymbol{X})$) must be specified.

As a proof of concept, chemical shift-guided MD (CS-MD) simulations were carried out on the U6 intramolecular stem-loop (ISL) RNA.[52] In solution, the U6 ISL RNA exists in different conformational states at pH 5.7 and at 7.0. At pH 5.7, U80 is extrahelical, whereas at pH 7.0 it is intrahelical (Figure 3B). CS-MD simulations of the U6 ISL were initiated from the pH 5.7 conformation (model 1; PDB: 1SYZ[53]), and measured chemical shifts assigned at pH 7.0 (BMRB: 5371[54]) were used to guide the simulations using eq 3. As a control, conventional unrestrained MD simulations were also carried.

To prepare for simulations, model 1 from PDBID: 1SYZ was solvated in a 67 Å cubic box of TIP3 waters.[55] The system was then charge neutralized with sodium counterions and relaxed with 100 steepest-descent, followed by 1000 adopted basis Newton–Raphson minimization steps. Prior to production runs, the RNA was heated from 250 to 325 K over 20 ps. During this phase the heavy atoms of the RNA were harmonically restrained using a force constant of 2.0 kcal/mol/Å. Production runs were then initiated from the resulting systems. For each independent simulation, the heating phase was initiated with different velocities by supplying distinct random number seeds. For both CS-MD and MD simulations,

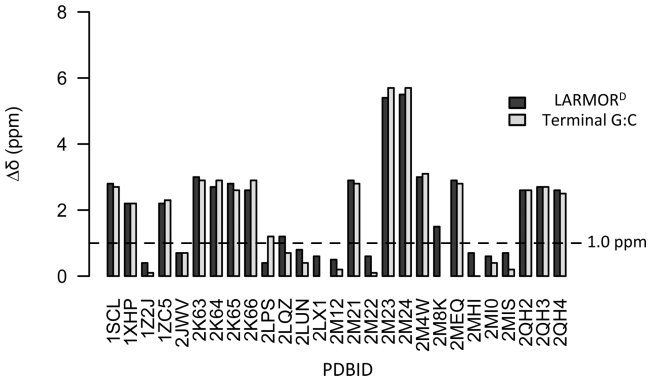


Figure 2. Detecting chemical shift referencing errors using LARMOR^D: comparison between LARMOR^D predicted referencing errors and those estimated using the approach of Aeschbacher et al., in which the chemical shift signature of the terminal G:C base pair is compared to expected values for correctly referenced chemical shifts. Results are shown for RNAs (identified by the PDBIDs) in the testing set.

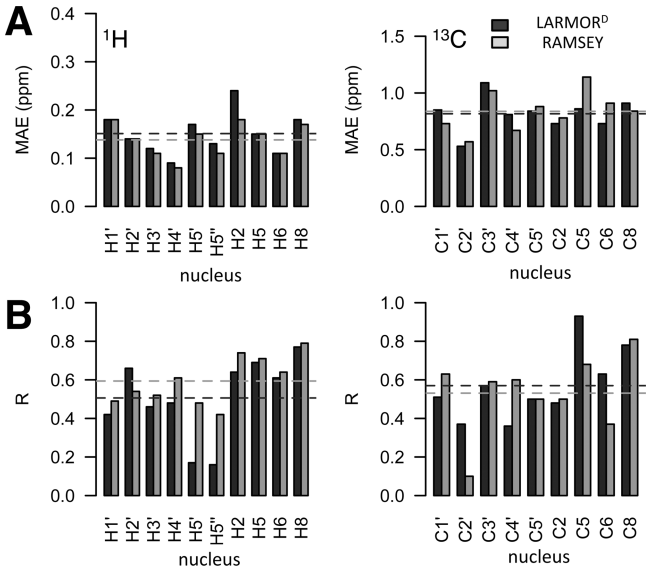


Figure 3. Assessing accuracy of LARMOR^D: (A) Bar plots of the mean absolute error (MAE) between measured and (gray) LARMOR^D and (light gray) RAMSEY predicted ^{1}H (left) and ^{13}C (right) chemical shifts. (B) Bar plots of the Pearson correlation coefficient (R) between measured and LARMOR^D and RAMSEY predicted ^{1}H (left) and ^{13}C (right) chemical shifts. Results are shown for RNAs in the testing set.

10 independent production runs were carried, each 2 ns long and simulated in the NPT ensemble (325 K and 1 atm). All simulations were carried out using the CHARMM36[56] nucleic acid force field. SHAKE was used to constrain hydrogen-containing bonds.[57] The van der Waals potential was truncated using a switching function between 10 and 12 Å. Long-range electrostatics was calculated using particle-mesh Ewald (PME) with a fourth-order B-spline used for interpolation.[58]

RESULTS AND DISCUSSION

For chemical shifts to be incorporated as structural restraints that guide MD simulations, the method used to predict chemical shifts from 3D coordinates must be (i) fast, (ii) accurate, and (iii) sensitive to RNA 3D structure. By design, LARMOR^D's simple dependence on interatomic distances

dx.doi.org/10.1021/jp508342x | J. Phys. Chem. B 2014, 118, 12168–12175

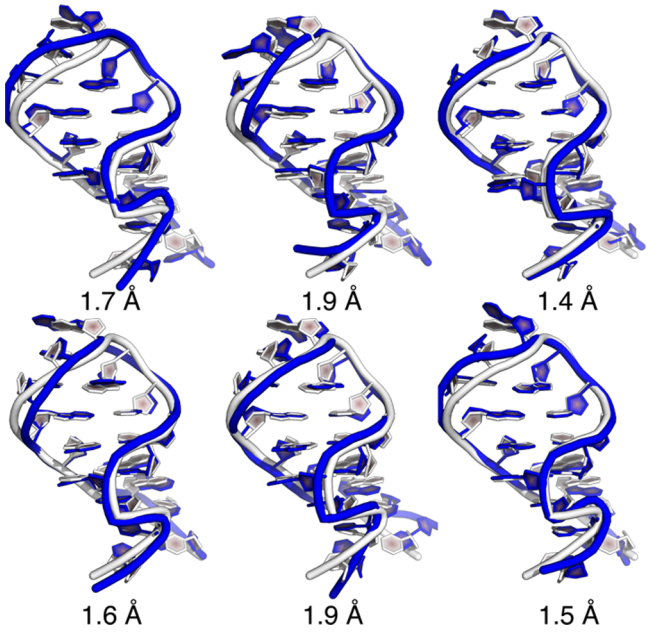


Figure 4. Assessing the ability of LARMOR^D chemicals shifts to resolve 3D structure of the SRL RNA: MCSYM was used to predict the structure of the sarcin−ricin loop (SRL) from sequence. The MCSYM conformational pool contained 8000 models. After using LARMOR^D to predict chemical shifts for each model in the MCSYM conformational pool, models with the lowest chemical shift errors were identified. This figure shows the cartoon overlays comparing the X-ray structure (PDB: 430D; *white*) of SRL RNA with the six MCSYM models that exhibited lowest error between measured and LARMOR^D predicted chemical shifts (*blue*).

guarantees that it can rapidly predict chemical shifts for 3D coordinates, and indeed, profiling of its predictions confirms this (Supporting Information, Table S4). We therefore focus our assessment of LARMOR^D on its accuracy and then its sensitivity to RNA 3D structure.

LARMOR^D Can Detect Referencing Errors in ^{13}C Data Sets. Before assessing the accuracy of individual LARMOR^D ^{1}H and ^{13}C chemical shift predictors, we investigated whether (i) ^{13}C chemical shifts data set with known referencing errors could be detected and (ii) reliable estimates for the magnitude of these errors could be determined using LARMOR^D. Unlike many other experimental observables, NMR chemical shifts are relative measurements. As such, any data set of chemical shifts is susceptible to issues related to inconsistent referencing, which complicates the establishment of reliable structure−chemical shifts relationship.[59]

In the case of proteins it was found that a significant number of data sets deposited in the BMRB contained ^{13}C and ^{15}N referencing errors.[59,60] Wishart and co-workers have addressed this using a structure-based approach in which a predictor, trained on chemical shift data known to be correctly and consistently referenced, is used to predict chemical shifts for a target protein from its solved X-ray or NMR structure. Systematic errors can be identified by comparing the predicted chemical shifts to the measured chemical shifts.[60]

Similar to those of proteins, RNA chemical shifts, in particular ^{13}C chemical shifts, are known to also contain referencing errors. Aeschbacher et al. recently described an approach that utilizes the chemical shift signatures of terminal G-C base pair to detect referencing errors.[46] Application of their approach indicated that a number of the entries in our testing set contained systematic referencing errors ($\Delta\delta_{GC}$).

To test whether LARMOR^D can detect referencing errors, it was used to predict ^{13}C chemical shifts for each RNA in the testing set, and then the absolute median error between measured and predicted chemical shifts was determined and used as an estimate of the systematic error ($\Delta\delta_{pred}$). As shown in Figure 2, 17 of the 27 entries in testing set that contained ^{13}C shifts exhibited large systematic error (i.e., $\Delta\delta_{pred}$ >1.0 ppm). Remarkably, $\Delta\delta_{pred}$ were in excellent agreement with $\Delta\delta_{GC}$ suggesting that LARMOR^D was not only able to identify referencing errors in ^{13}C chemical shift data sets but was also able to provide reliable estimates of the magnitude of these errors. Immediately, LARMOR^D could be incorporated into an automated procedure that allows referencing errors in RNA chemical shifts to be detected, corrected, and then made available to the scientific community via a secondary database, as has been done for proteins.[60]

Despite Its Simplicity, LARMOR^D Accurately Predicts Chemical Shifts. Given the simple model used by LARMOR^D, a critical question is how accurate are the predicted chemical shifts. We answered this question for *individual* ^{1}H and ^{13}C nuclei by calculating the mean absolute error (MAE) and the Pearson correlation coefficient (R) between experimentally measured and LARMOR^D predicted chemical shifts in the testing set. Prior to assessing their accuracy, ^{13}C chemical shifts were re-referenced where necessary. The MAE for ^{1}H nuclei ranged between 0.09 and 0.24 ppm, with a mean of 0.15 ppm. For ^{13}C the range was 0.53 and 1.09 ppm, and the mean was 0.81 ppm (Figure 3B). By comparison, RAMSEY[45] exhibited MAE that ranged between 0.08 and 0.18 ppm (mean of 0.14 ppm) for ^{1}H and between 0.57 and 1.14 ppm (mean of 0.83 ppm) for ^{13}C (Figure 3B). The accuracy for LARMOR^D is therefore similar to RAMSEY. A similar picture emerged when examining R—for ^{1}H and ^{13}C, the mean value was 0.51 and 0.57, respectively, for LARMOR^D, compared to 0.59 and 0.53, respectively, for RAMSEY (Figure 3C).

In general, RAMSEY appears to predict ^{1}H chemical shifts with slightly greater accuracy than LARMOR^D. Because ^{1}H chemical shifts are known to be more highly sensitive to ring-current effects, we investigated whether accounting for ring-current effects when predicting ^{1}H chemical shifts would result in more accurate predictions. For ^{1}H nuclei, we generated a set of new predictors by repeating the parametrization (see eq 1), but this time including an additional ring-current term (calculated using the Johnson−Bovey model[61]) in eq 1. For these predictors, however, we did not observe any noticeable improvements in the accuracy (data not shown). Direct comparison with SHIFTS and NUCHEMICS, two empirical predictors capable of also predicting ^{1}H chemical shifts, revealed LARMOR^D to be more accurate. SHIFTS and NUCHEMICS predict chemical shifts with an MAE of 0.37 and 0.21 ppm, and an R of 0.38 and 0.46, respectively, as compared to 0.15 ppm and 0.51, respectively, for LARMOR^D (Supporting Information, Table S5). Together, these results suggest that without *explicitly* accounting for hydrogen bonding, base−base stacking, ring current, magnetic anisotropies, and bond polarization effects, the simple distance-based approach used here is sufficient to enable LARMOR^D predictions of both ^{1}H and ^{13}C chemical shifts with good accuracy relative to the currently available empirical structure-based approaches.

LARMOR^D Predicted Chemical Shifts Are Sensitive to RNA 3D Structure. For chemical shifts to be used to guide

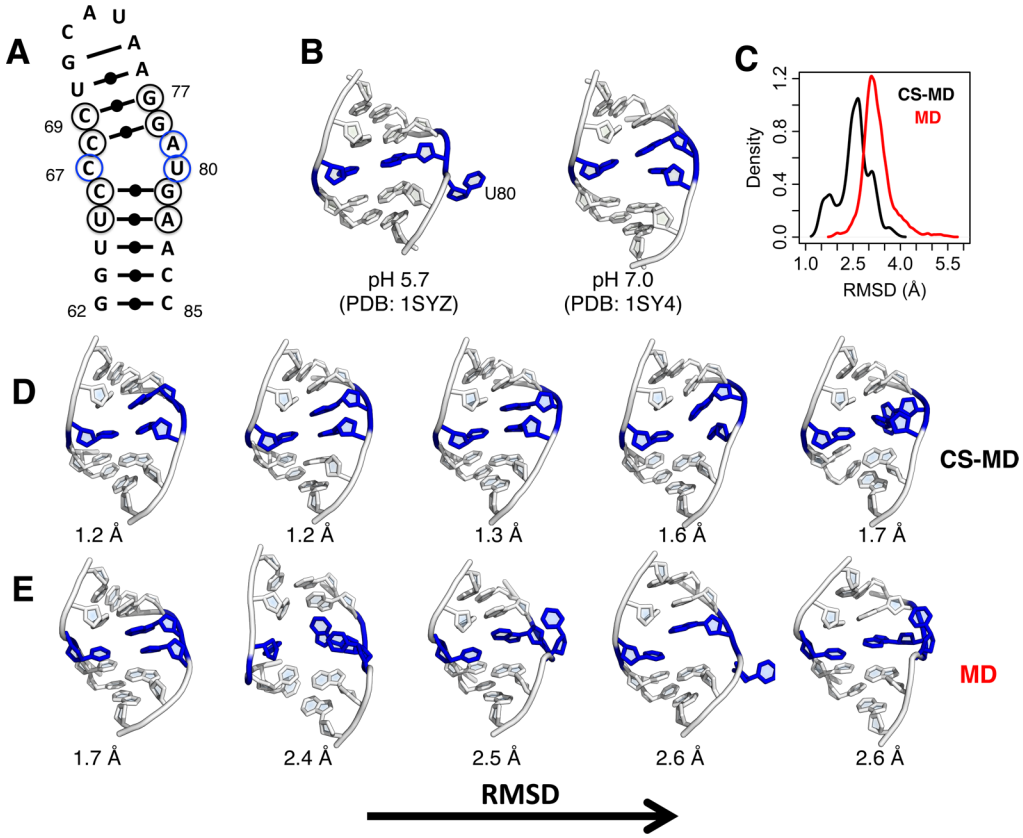


Figure 5. LARMOR^{D} enabled chemical shifts guided simulations of the U6 ISL RNA: (A) Secondary structure of the U6 intramolecular stem-loop (ISL) RNA. Residues C67, A79, and U80 are highlighted in blue. Residues 65−69 and 76−82, which makes up the conformationally active region, are circled. (B) Cartoon representation of the U6 ISL at pH 5.7 (model 1; PDB: 1SYZ) and pH 7.0 (model 1; PDB: 1SY4). Only the conformationally active region is shown in the cartoons. (C) RMSD distribution over the set of 10 CS-MD (*black*) and MD (*red*) trajectories. RMSDs were calculated relative to the pH 7.5. From each of the 10 CS-MD and 10 MD trajectories, the model with the lowest RMSD was extracted; shown are the five models with the lowest RMSD extracted from the (D) CS-MD and (E) MD trajectories.

MD simulations, it is essential that the predictor used to back-predict chemical shifts from atomic models is sensitive to RNA structure. To assess its sensitivity to 3D structure, LARMOR^{D} was used to predict the chemical shifts for each model in a conformational pool that contained 8000 putative models of the sarcin−ricin loop (SRL) RNA.[62] These models were generated from its sequence using the RNA structure prediction software MCSYM[63] and correspond to structures consistent with the nine lowest-energy secondary structures as predicted by the program MC-fold.[49,63] In Figure 4 the six models with the lowest error between measured and LARMOR^{D} predicted chemical shifts are overlaid with the X-ray structure. As can been seen the low-error models are in excellent agreement with the solved structure. Further, we found that 24 of the 30 models with the lowest chemical shift error had an RMSD < 2.0 Å. These results indicate that LARMOR^{D} predicted ^{1}H and ^{13}C chemical shifts are indeed sensitive to RNA 3D structure. LARMOR^{D} should therefore be useful in the context of chemical shift-guided MD simulations. Additionally, the demonstrated sensitivity to RNA 3D structure strongly suggests that LARMOR^{D} can be employed in RNA structure prediction either as a postprocessing structural filter to identify representative models (as was demonstrated for SRL RNA here) or to construct a penalty term to guide conformational exploration in structure prediction approaches à la CS-Rosetta-RNA.[26]

LARMOR^{D} Predicted Chemical Shifts Effectively Guides MD Simulations of RNA. As the focus of this work was the development of a chemical shifts predictor that enabled NMR chemical shifts to be *easily* incorporated into molecular simulations of RNA, we conclude our study by implementing the LARMOR^{D}-based chemical shifts restraint functionality into the CHARMM macromolecular mechanics package[49] and then carrying out chemical shift-restrained MD (CS-MD) simulations on a model system. Specifically, CS-MD simulations were carried out on the U6 intramolecular stem-loop (ISL) RNA[52] (Figure 5A,B). At pH 5.7, the U80 base of U6 ISL is in an extrahelical position, whereas at pH 7.0, U80 is intrahelical (Figure 5B). Base flipping is ubiquitous in RNA structural dynamics and in many cases acts as a switching mechanism between conformational states of an RNA.[64−67] Together with its relatively small size (27 nt), this makes the U6 ISL an excellent benchmark system for our CS-MD simulation approach.

Starting from the pH 5.7 coordinates (PDB: 1SYZ;[53] model 1), we carried out CS-MD simulations using reference chemical shifts assigned at pH 7.0 (BMRB: 5371[54]). If the chemical shifts restraints are effective then, in comparison to the unrestrained simulation, the conformations along the trajectory of the CS-MD simulation should more closely resemble the intrahelically stacked pH 7.0 conformation; this despite the fact that the simulations were initiated from the extrahelical pH 5.0 conformation.

dx.doi.org/10.1021/jp508342x | J. Phys. Chem. B 2014, 118, 12168−12175

We found that in the case of the CS-MD simulations, the distribution of RMSDs relative to the pH 7.0 structure was shifted toward lower RMSD values than in the control MD simulations (Figure 5C; Supporting Information, Figure S1). This indicates that the chemical shifts restraint, constructed using chemical shifts assigned at pH 7.0, was effective in guiding sampling away from the initial pH 5.7 conformation (Figure 5B; left) toward the pH 7.0 (Figure 5B; right). Five out of the 10 CS-MD simulations sampled conformations with RMSDs < 1.8 Å, compared to only one for the control simulations (Figure 5D,E and Supporting Information, Figures S1 and S2). Similarly, three out of the 10 CS-MD simulations were able to sample conformations within ∼1.3 Å of the flipped in and stacked state, while the unrestrained simulations were limited to RMSDs > 1.7 Å (Figure 5D; Supporting Information, Figure S1). These results demonstrate that the LARMOR^{D} chemical shift restraints are able to effectively guide conformational sampling of U6 ISL, allowing chemical shifts to be used within the context minimally biased[12] and/or maximum entropy-based[11] approaches to map the conformational landscape of RNA.

In addition to opening up the possibility of using chemical shifts to map the conformational landscape of RNAs,[68] chemical shift-guided simulations may be of utility in carrying explicit solvent MD-based refinement of NMR structures. Conventional NMR structure determination is typically carried out without explicitly accounting for solvent effects (i.e., carried out *in vacuo*) and using *ad hoc* treatment of nonbonded interactions. Indeed, previous work has shown that structure refinement using an accurate force field, while explicitly accounting for solvent, resulted in structures that are sometimes in better agreement with the original NMR data than the structures solved using conventional *in vacuo* approaches.[69]

To facilitate the use of LARMOR^{D}, the parameters and the source code implementing the predictor is made freely available to academics (see the link shown in the Supporting Information byline). In addition, the LARMOR^{D}-based chemical shifts restraint module will be available in CHARMM.

Though RNA was the focus of the current work, the implementation of LARMOR^{D} in the stand-alone predictor and CHARMM follows a general design approach, so that in principle they can be used for any molecular assembly, for example, proteins, nucleic acids (NA), protein−protein, protein−NA, and NA−NA complexes, provided, of course, that the appropriate parameters are available, which is a component of ongoing work.

CONCLUSION

In summary, we have developed LARMOR^{D}, a simple, fast, and accurate method for predicting nonexchangeable ^{1}H and protonated ^{13}C RNA chemical shifts based only on interatomic distances. We showed that LARMOR^{D} was capable of resolving RNA 3D structure, and more importantly for this study, we demonstrated that LARMOR^{D}-based chemical shifts restraints were effective in guiding conformational sampling during CS-MD simulations. Future work will focus on developing and validating robust CS-MD simulations protocols that will allow chemical shifts to be used to map the conformational landscape of RNAs, as well as protocols for using chemical shifts to refine NMR structures of RNAs. In addition, we will also explore more fully combining LARMOR^{D} with current state-of-the-art prediction approaches to aid in RNA structure prediction. This application would be particularly useful in cases where chemical shifts are the only high-quality structural data available−as is the case for transiently populated "invisible" states of RNAs that can now be detected using chemical shifts relaxation dispersion experiments.[70,71]

ASSOCIATED CONTENT

Supporting Information

Tables S1−S8: (S1) training set; (S2) testing set; (S3) chemical shift entries identified as outliers; (S4) LARMOR^{D} execution times. (S5) Comparison between the accuracy of LARMOR^{D} ^{1}H predictors to SHIFTS and NUCHEMICS; (S6) Reference chemical shifts used in eq 1. (S7) Accuracy of preliminary imino ^{1}H and ^{15}N chemical shift predictors; and (S8) Comparison between measured and predicted chemical shifts for U6 ISL RNA. Figures S1 and S2: (S1) RMSD time series for CS-MD and MD simulations of the U6 ISL RNA; and (S2) Effective chemical shifts error time series for CS-MD and MD simulations of the U6 ISL RNA. Movie 1: (S1) Movie of a CS-guided MD trajectory of U6 ISL RNA. Data S1: (S1) The optimized LARMOR^{D} parameters. This material is available free of charge via the Internet at http://pubs.acs.org. The source for LARMOR^{D} is available at: http://brooks.chem.lsa.umich.edu/software. The LARMOR^{D} module implementing the predictor and the chemical shift biasing restraint will also be made available in CHARMM.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: afrankz@umich.edu. (A.T.F.)
*E-mail: brookscl@umich.edu. (C.L.B.)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge valuable scientific discussions with G. Goh. A.T.F. was supported through the Univ. of Michigan President's Postdoctoral Fellowship. The authors acknowledge support for this work from the National Institutes of Health (GM103695) and the National Science Foundation through the Center for Theoretical Biological Physics (PHY0216576).

REFERENCES

(1) Gestland, R. F.; Atkins, J. F. *The RNA World. the Nature of Modern RNA Suggests a Prebiotic RNA World;* Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006.

(2) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777−793.

(3) Al-Hashimi, H. M.; Gosser, Y.; Gorin, A.; Hu, W.; Majumdar, A.; Patel, D. J. Concerted Motions in HIV-1 TAR RNA May Allow Access to Bound State Conformations: RNA Dynamics From NMR Residual Dipolar Couplings. *J. Mol. Biol.* **2002**, *315*, 95−102.

(4) Wunnicke, D.; Strohbach, D.; Weigand, J. E.; Appel, B.; Feresin, E.; Suess, B.; Muller, S.; Steinhoff, H. J. Ligand-Induced Conformational Capture of a Synthetic Tetracycline Riboswitch Revealed by Pulse EPR. *RNA* **2010**, *17*, 182−188.

(5) Haller, A.; Rieder, U.; Aigner, M.; Blanchard, S.; Micura, R. Conformational Capture of the SAM-II Riboswitch. *Nat. Chem. Biol.* **2011**, *7*, 393−400.

(6) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. Visualizing Spatially Correlated Dynamics That Directs RNA Conformational Transitions. *Nature* **2007**, *450*, 1263−1267.

(7) 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*, 3670−3679.
(8) Pitera, J. W.; Chodera, J. D. On the Use of Experimental Observations to Bias Simulated Ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445−3451.
(9) Roux, B.; Weare, J. On the Statistical Equivalence of Restrained-Ensemble Simulations with the Maximum Entropy Method. *J. Chem. Phys.* **2013**, *138*, 084107.
(10) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular Dynamics Simulations with Replica-Averaged Structural Restraints Generate Structural Ensembles According to the Maximum Entropy Principle. *J. Chem. Phys.* **2013**, *138*, 094112.
(11) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10*, e1003406.
(12) White, A. D.; Voth, G. A. Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J. Chem. Theory Comput.* **2014**, *10*, 3023−3030.
(13) Różycki, B.; Kim, Y. C.; Hummer, G. SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions. *Structure* **2011**, *19*, 109−116.
(14) Islam, S. M.; Stein, R. A.; McHaourab, H. S.; Roux, B. Structural Refinement From Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme. *J. Phys. Chem. B* **2013**, *117*, 4740−4754.
(15) De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M. Characterization of the Lnterdomain Motions in Hen Lysozyme Using Residual Dipolar Couplings as Replica-Averaged Structural Restraints in Molecular Dynamics Simulations. *Biochemistry* **2013**, *52*, 6480−6486.
(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*, 128−132.
(17) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. Recognition Dynamics Up to Microseconds Revealed From an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **2008**, *320*, 1471−1475.
(18) Camilloni, C. C.; Robustelli, P. P.; De Simone, A. A.; Cavalli, A. A.; Vendruscolo, M. M. Characterization of the Conformational Equilibrium Between the Two Major Substates of RNase a Using NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134*, 3968−3971.
(19) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the Free-Energy Landscapes of Proteins by NMR-Guided Metadynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6817−6822.
(20) Camilloni, C.; Vendruscolo, M. Statistical Mechanics of the Denatured State of a Protein Using Replica-Averaged Metadynamics. *J. Am. Chem. Soc.* **2014**, *136*, 8982−8991.
(21) Giessner-Prettre, C.; Pullman, B. Quantum Mechanical Calculations of NMR Chemical Shifts in Nucleic Acids. *Q. Rev. Biophys.* **1987**, *20*, 113−172.
(22) Ghose, R.; Marino, J.; Wiberg, K.; Prestegard, J. Dependence of ^{13}C Chemical Shifts on Glycosidic Torsional Angles in Ribonucleic Acids. *J. Am. Chem. Soc.* **1994**, *116*, 8827−8828.
(23) Farès, C.; Amata, I.; Carlomagno, T. 13C-Detection in RNA Bases: Revealing Structure−Chemical Shift Relationships. *J. Am. Chem. Soc.* **2007**, *129*, 15814−15823.
(24) Rossi, P.; Harbison, G. S. Calculation of 13C Chemical Shifts in Rna Nucleosides: Structure-13C Chemical Shift Relationships. *J. Magn. Reson.* **2001**, *151*, 1−8.
(25) Ebrahimi, M.; Rossi, P.; Rogers, C.; Harbison, G. S. Dependence of 13C NMR Chemical Shifts on Conformations of RNA Nucleosides and Nucleotides. *J. Magn. Reson.* **2001**, *150*, 1−9.
(26) Sripakdeevong, P.; Cevec, M.; Chang, A. T.; Erat, M. C.; Ziegeler, M.; Zhao, Q.; Fox, G. E.; Gao, X.; Kennedy, S. D.; Kierzek, R.; et al. Structure Determination of Noncanonical RNA Motifs Guided by 1H NMR Chemical Shifts. *Nat. Meth* **2014**, *11*, 413−416.
(27) Lu, K.; Miyazaki, Y.; Summers, M. F. Isotope Labeling Strategies for NMR Studies of RNA. *J. Biomol. NMR* **2009**, *46*, 113−125.
(28) Dayie, T. K.; Thakur, C. S. Site-Specific Labeling of Nucleotides for Making RNA for High Resolution NMR Studies Using an E. Coli Strain Disabled in the Oxidative Pentose Phosphate Pathway. *J. Biomol. NMR* **2010**, *47*, 19−31.
(29) Duss, O.; Maris, C.; Schroetter, von C.; Allain, F. H. T. A Fast, Efficient and Sequence-Independent Method for Flexible Multiple Segmental Isotope Labeling of RNA Using Ribozyme and RNase H Cleavage. *Nucleic Acids Res.* **2010**, *38*, e188.
(30) Alvarado, L. J.; LeBlanc, R. M.; Longhini, A. P.; Keane, S. C.; Jain, N.; Yildiz, Z. F.; Tolbert, B. S.; D'Souza, V. M.; Summers, M. F.; Kreutz, C.; et al. Regio-Selective Chemical-Enzymatic Synthesis of Pyrimidine Nucleotides Facilitates RNA Structure and Dynamics Studies. *ChemBioChem* **2014**, *15*, 1573−1577.
(31) Barton, S.; Heng, X.; Johnson, B. A.; Summers, M. F. Database Proton NMR Chemical Shifts for RNA Signal Assignment and Validation. *J. Biomol. NMR* **2012**, *55*, 33−46.
(32) Bahrami, A.; Clos, L. J.; Markley, J. L.; Butcher, S. E.; Eghbalnia, H. R. RNA-PAIRS: RNA Probabilistic Assignment of Imino Resonance Shifts. *J. Biomol. NMR* **2012**, *52*, 289−302.
(33) Aeschbacher, T.; Schmidt, E.; Blatter, M.; Maris, C.; Duss, O.; Allain, F. H. T.; Guntert, P.; Schubert, M. Automated and Assisted RNA Resonance Assignment Using NMR Chemical Shift Statistics. *Nucleic Acids Res.* **2013**, *41*, e172.
(34) Krähenbühl, B.; Bakkali, El. I.; Schmidt, E.; Güntert, P.; Wider, G. Automated NMR Resonance Assignment Strategy for RNA via the Phosphodiester Backbone Based on High-Dimensional Through-Bond APSY Experiments. *J. Biomol. NMR* **2014**, *59*, 87−93.
(35) Krähenbühl, B.; Lukavsky, P.; Wider, G. Strategy for Automated NMR Resonance Assignment of RNA: Application to 48-Nucleotide K10. *J. Biomol. NMR* **2014**, *59*, 231−240.
(36) D'Souza, V.; Dey, A.; Habib, D.; Summers, M. F. NMR Structure of the 101-Nucleotide Core Encapsidation Signal of the Moloney Murine Leukemia Virus. *J. Mol. Biol.* **2004**, *337*, 427−442.
(37) Meiler, J. J. PROSHIFT: Protein Chemical Shift Prediction Using Artificial Neural Networks. *J. Biomol. NMR* **2003**, *26*, 25−37.
(38) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50*, 43−57.
(39) Xu, X. P.; Case, D. A. Automated Prediction of 15N, 13Calpha, 13Cbeta and 13C' Chemical Shifts in Proteins Using a Density Functional Database. *J. Biomol. NMR* **2001**, *21*, 321−333.
(40) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and Accurate Predictions of Protein NMR Chemical Shifts From Interatomic Distances. *J. Am. Chem. Soc.* **2009**, *131*, 13894−13895.
(41) Atieh, Z.; Aubert-Frécon, M.; Allouche, A.-R. Rapid, Accurate and Simple Model to Predict NMR Chemical Shifts for Biological Molecules. *J. Phys. Chem. B* **2010**, *114*, 16388−16392.
(42) Shen, Y.; Bax, A. SPARTA+: a Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48*, 13−22.
(43) Dejaegere, A.; Bryce, R. A.; Case, D. A. An Empirical Analysis of Proton Chemical Shifts in Nucleic Acids. *ACS Symp. Ser.* **1999**, *732*, 194−206.
(44) Cromsigt, J. A.; Hilbers, C. W.; Wijmenga, S. S. Prediction of Proton Chemical Shifts in RNA. Their Use in Structure Refinement and Validation. *J. Biomol. NMR* **2001**, *21*, 11−29.
(45) Frank, A. T.; Bae, S.-H.; Stelzer, A. C. Prediction of RNA 1H and 13C Chemical Shifts: a Structure Based Approach. *J. Phys. Chem. B* **2013**, *117*, 13497−13506.
(46) Aeschbacher, T.; Schubert, M.; Allain, F. H.-T. A Procedure to Validate and Correct the 13C Chemical Shift Calibration of RNA Datasets. *J. Biomol. NMR* **2012**, *52*, 179−190.
(47) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.

(48) Perone, C. S. Pyevolve: a Python Open-Source Framework for Genetic Algorithms. *ACM SIGEVOlution* **2009**, *4*, 12−20.
(49) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics. *J. Comput. Chem.* **1983**, *4*, 187−217.
(50) Jack, A.; Levitt, M. Refinement of Large Structures by Simultaneous Minimization of Energy and R-Factor. *Acta Crystallogr., Sect. A* **1978**, *34*, 931−935.
(51) Habeck, M.; Rieping, W.; Nilges, M. Weighting of Experimental Evidence in Macromolecular Structure Determination. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1756−1761.
(52) Nilsen, T. W. RNA-RNA Interactions in Nuclear Pre-mRNA Splicing. *Cold Spring Harbor Monogr. Arch.* **1998**, *35*, 279−307.
(53) Reiter, N. J.; Blad, H.; Abildgaard, F.; Butcher, S. E. Dynamics in the U6 RNA Intramolecular Stem−Loop: a Base Flipping Conformational Change. *Biochemistry* **2004**, *43*, 13739−13747.
(54) Huppler, A.; Nikstad, L. J.; Allmann, A. M.; Brow, D. A.; Butcher, S. E. Metal Binding and Base Ionization in the U6 RNA Intramolecular Stem-Loop Structure. *Nat. Struct. Biol.* **2002**, *9*, 431−435.
(55) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(56) Denning, E.; Priyakumar, U.; Nilsson, L.; MacKerell, A. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(57) Barth, E.; Kuczera, K.; Leimkuhler, B.; Skeel, R. D. Algorithms for Constrained Molecular-Dynamics. *J. Comput. Chem.* **1995**, *16*, 1192−1209.
(58) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.
(59) Wishart, D. S.; Case, D. A. Use of Chemical Shifts in Macromolecular Structure Determination. *Methods Enzymol.* **2001**, *338*, 3−34.
(60) Zhang, H.; Neal, S.; Wishart, D. S. RefDB: a Database of Uniformly Referenced Protein Chemical Shifts. *J. Biomol. NMR* **2003**, *25*, 173−195.
(61) Johnson, C. E.; Bovey, F. A. Calculation of Nuclear Magnetic Resonance Spectra of Aromatic Hydrocarbons. *J. Chem. Phys.* **1958**, *29*, 1012.
(62) Endo, Y.; Wool, I. G. The Site of Action of Alpha-Sarcin on Eukaryotic Ribosomes. the Sequence at the Alpha-Sarcin Cleavage Site in 28 S Ribosomal Ribonucleic Acid. *J. Biol. Chem.* **1982**, *257*, 9054−9060.
(63) Parisien, M.; Major, F. The MC-Fold and MC-Sym Pipeline Infers RNA Structure From Sequence Data. *Nature* **2008**, *452*, 51−55.
(64) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional Insights From the Structure of the 30S Ribosomal Subunit and Its Interactions with Antibiotics. *Nature* **2000**, *407*, 340−348.
(65) Rupert, P.; Ferre-D'Amare, A. Crystal Structure of a Hairpin Ribozyme-Inhibitor Complex with Implications for Catalysis. *Nature* **2001**, *410*, 780−786.
(66) Hoang, C.; Ferre-D'Amare, A. R. Cocrystal Structure of a tRNA Psi 55 Pseudouridine Synthase: Nucleotide Flipping by an RNA-Modifying Enzyme. *Cell* **2001**, *107*, 929−939.
(67) Yang, X.; Gérczei, T.; Glover, L. T.; Correll, C. C. Crystal Structures of Restrictocin-Inhibitor Complexes with Implications for RNA Recognition and Base Flipping. *Nat. Struct. Biol.* **2001**, *8*, 968−973.
(68) Rinnenthal, J.; Buck, J.; Ferner, J.; Wacker, A.; Fürtig, B.; Schwalbe, H. Mapping the Landscape of RNA Dynamics with NMR Spectroscopy. *Acc. Chem. Res.* **2011**, *44*, 1292−1301.
(69) Henriksen, N. M.; Davis, D. R.; Cheatham, T. E. Molecular Dynamics Re-Refinement of Two Different Small RNA Loop Structures Using the Original NMR Data Suggest a Common Structure. *J. Biomol. NMR* **2012**, *53*, 321−339.
(70) Dethoff, E. A.; Petzold, K.; Chugh, J.; Casiano-Negroni, A.; Al-Hashimi, H. M. Visualizing Transient Low-Populated Structures of RNA. *Nature* **2012**, 1−7.
(71) Lee, J.; Dethoff, E. A.; Al-Hashimi, H. M. Invisible RNA State Dynamically Couples Distant Motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9485−9490.