Published in final edited form as:

J Phys Chem B. 2013 February 21; 117(7): 2045–2052. doi:10.1021/jp310863c.

Utility of ^1H NMR Chemical Shifts in Determining RNA Structure and Dynamics

Aaron T. Frank^{1,3}, Scott Horowitz², Ioan Andricioaei^{1,*}, and Hashim M. Al-Hashimi^{1,2,*}

¹Department of Chemistry, University of California Irvine, 1102 Natural Sciences 2, Irvine, California, 92697, USA

²Department of Chemistry & Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan, 48109, USA

Abstract

The development of methods for predicting NMR chemical shifts with high accuracy and speed is increasingly allowing use of these abundant, readily accessible measurements in determining the structure and dynamics of proteins. For nucleic acids, however, despite the availability of semi-empirical methods for predicting ^1H chemical shifts, their use in determining the structure and dynamics has not yet been examined. Here, we show that ^1H chemical shifts offer powerful restraints for RNA structure determination, allowing discrimination of native structure from non-native states to within 2–4 Å, and <3 Å when ignoring highly flexible residues. Theoretical simulations shows that while ^1H chemical shifts can provide valuable information for constructing RNA dynamic ensembles, large uncertainties in the chemical shift predictions and inherent degeneracies lead to higher uncertainties as compared to residual dipolar couplings.

There has been a long-standing effort directed towards extracting the rich structural and dynamic information contained within NMR isotropic chemical shifts^{1–6}. Recent advances in semi-empirical methods for predicting chemical shifts based on protein structure with high accuracy and speed are allowing protein structure determination with chemical shifts as the sole input experimental restraint^{7–10}. Isotropic chemical shifts are also time-averaged over all motions occurring at timescales faster than milliseconds and therefore provide a sensitive probe of dynamics occurring over a wide range of timescales^{11,12}. They have successfully been used to predict protein S^2 NMR order parameters^{13,14}, in constructing dynamic ensembles of proteins¹⁵ including intrinsically disordered proteins^{16–19}, and in evaluating molecular dynamics (MD) simulations of proteins^{20–22}.

Chemical shifts have long been recognized to be sensitive reporters of nucleic acid structure, but are seldom used in structure determination or dynamics characterization²³. This is despite the availability of semi-empirical methods for predicting ^1H chemical shifts based on nucleic acid structure with high speed and accuracy. In particular, the program SHIFTS²⁴ developed by Case and co-workers, and NUCHEMICS²⁵ developed by Wijmenga and co-workers, have been shown to predict sugar (H1', H2', H3', H4', H5' and H5'') and nucleobase (H2, H5, H6 and H8) ^1H chemical shifts for a panel of diverse nucleic acid structures with an overall root-mean-square deviation (RMSD) ranging between 0.16–0.28 ppm. These predictions compare favorably with protein ^1H chemical shift predictions

*hashimi@umich.edu or andricio@uci.edu.

³Current address Nymirum Inc., 3510 West Liberty Road, Ann Arbor, Michigan, 48103, USA

Supporting Information: Agreement between measured and predicted chemical shifts for individual proton types are reported in Fig S1 and Table S1. This information is available free of charge via the Internet at <http://pubs.acs.org>.

(RMSD ranging between 0.15–0.6 ppm)^{26–28} and suggest that ¹H chemical shifts may be used to define the structure and dynamic properties of nucleic acids in a manner analogous to what has been done with proteins.

The use of chemical shift as new abundant source of structure and dynamic information is arguably more important for nucleic acids as compared to proteins. NMR structure determination of nucleic acids traditionally suffers from a shortage in accessible inter-proton NOE-derived distance restraints that can be applied towards structure characterization. This problem is compounded by a high degree of flexibility, particularly in RNA, which can complicate the interpretation of NOE-derived distances. Here, we examine the utility of ¹H chemical shifts in determining the structure and dynamic ensembles of RNA. Our results suggest that ¹H chemical shifts can immediately be used in RNA structure validation and in evaluating the quality of RNA dynamic ensembles. However, our results suggest that additional improvements in chemical shift predictions are required in order to allow their use as primary data in generating ensembles.

METHODS AND MATERIALS

Predicting RNA ¹H Chemical Shifts

We used a panel of 18 RNA structures (1IDV, 1JU7, 1L1W, 1N8X, 1NC0, 1OW9, 1XHP, 1Z2J, 1ZC5, 2FDT, 2JTP, 2JYM, 2KOC, 2KYD, 2L1V, 2L3E, 2L5Z and 2QH2) to evaluate ¹H chemical shift predictions using SHIFTS and NUCHEMICS. This panel represents RNA structures that have been determined by NMR after SHIFTS and NUCHEMICS (2002–2011) were developed and parameterized and for which ¹H chemical shift assignments were also available in the Biological Magnetic Resonance Bank (BMRB)²⁹ (<http://www.bmrwisc.edu/>). Four additional structures (2QH3, 2QH4, 1YMO, 2JWV) were not included due to undocumented or incomplete chemical shift referencing. However, including these structures had little to no impact on the results presented here.

Molecular dynamics simulations

MD simulations of an RNA duplex (PDBID:2KYD)³⁰, UUCG tetra-loop (PDBID:2KOC)³¹, asymmetrical internal loop (PDBID:2L3E)³² and pre-quenosine-1 (preQ1) riboswitch (PDBID:2L1V)^{33,34} were performed at 300 and 500 K using GROMACS 4.5.1³⁵ and the AMBER94³⁶ nucleic acid forcefield. Structures were subjected to 100 steps of steepest descent minimization and subsequently solvated with TIP3 water³⁷ in a triclinic box and charge neutralized using sodium counterions. Harmonic restraints with a force constant of 1000 kJ mol⁻¹ nm⁻² were placed on the heavy atoms and simulated at 300 K for 1.4 ns. The harmonic restraints were then gradually released over 200 ps. Starting from the equilibrated coordinates, two 4 ns trajectories were generated at 300 and 500 K, respectively. Coordinates were saved every 2 ps.

Replica-exchange molecular dynamics (REMD) simulations³⁸ were used to generate a broad conformational pool for the HIV-1 TAR apical loop from which sets of non-overlapping reference ensembles could be constructed (see below). Initial coordinates were obtained using Rosetta FARNA³⁹, a *de novo* structure determination program for nucleic acids. Starting from the primary sequence, UAUCGAGCCUGGGAGCUCGAUA, 1000 candidate structures were generated with base pairing restraints between residues U1 and A22, A2 and U21, U3 and A20, C4 and G19, G5 and C18, A6 and U17, G7 and C16, and C8 and G15. The conformation with the lowest energy was used as the initial coordinates for the REMD simulations. The initial structure was subjected to 100 steps of steepest descent minimization and then solvated with TIP3 water³⁷ in an octahedron box and charge neutralized using sodium counterions. Harmonic restraints with a force constant of 1000 kJ mol⁻¹ nm⁻² were

placed on the heavy atoms and simulated at 300 K for 1.4 ns. The harmonic restraints were then gradually released over 200 ps. Starting from the equilibrated coordinates at 300 K, 15 additional apical loop replicas were prepared by slowly heating the system to 303, 306, 309, 312, 315, 319, 322, 325, 329, 332, 335, 339, 342, 346 and 350 K. REMD simulations were then initiated from these 16 replicas. Exchanges were attempted every 2 ps and coordinates were saved every 2 ps. Production trajectories 90 ns in length were generated. The 45,000 conformations were used as the representative conformational pool for the TAR apical loop.

Weighted chemical shift RMSD

To investigate the impact of the accuracy of chemical shifts predictions on the ability of ^1H chemical shifts to resolve the difference between related RNA conformations, we defined a weighted chemical shift RMSD (CS^{RMSD}) that takes into account the correlation between measured and predicted chemical shifts calculated from the NMR structure. The weighted CS^{RMSD} is calculated using:

$$\frac{1}{L_{\text{CS}}} \sum_{i=1}^{L_{\text{CS}}} \sum_j^N R_j^2 (\delta_{\text{meas}} - \delta_{\text{pred}})^2 \quad (1)$$

where R_j^2 is the square of the Pearson correlation coefficient for a given proton type j and calculated from the average NMR structure. Using this equation, proton types that exhibit higher correlations between measured and predicted chemical shifts in the average NMR structure contribute more to the CS^{RMSD} and *vice versa*.

Selection Algorithm for Generating Ensembles

Ensembles were constructed using chemical shifts, residual dipolar coupling (RDC) and chemical shift+RDC data using the sample and select (SAS) approach as described previously^{40,41}. The ensembles were selected by minimizing, using Monte Carlo procedures (see below), the cost function,

$$\chi^2 = K_{\text{CS}} \chi_{\text{CS}}^2 + K_{\text{RDC}} \chi_{\text{RDC}}^2 \quad (2)$$

where

$$\chi_{\text{CS}}^2 = \frac{1}{L_{\text{CS}}} \sum_{i=1}^{L_{\text{CS}}} (\delta_i^{\text{pred}} - \delta_i^{\text{meas}})^2$$

and

$$\chi_{\text{RDC}}^2 = \frac{1}{L_{\text{RDC}}} \sum_{k=1}^{L_{\text{RDC}}} (D_k^{\text{pred}} - D_k^{\text{meas}})^2$$

Here, χ^2 is the total cost function to be minimized; χ_{CS}^2 and χ_{RDC}^2 are the chemical shift and RDC components of χ^2 , respectively; K_{CS} and K_{RDC} are coefficients that determine the contribution of each component to χ^2 ; δ_i^{pred} and δ_i^{meas} are the predicted and measured chemical shifts for the i^{th} proton averaged over the N structures in an ensemble (see below), while D_k^{pred} and D_k^{meas} are the predicted and measured RDCs averaged over the N members of the ensemble, with k an index that runs over the bond vectors; L_{CS} and L_{RDC} are the total number of chemical shifts and RDCs, respectively. For selections using

chemical shifts only, $K_{CS}=1$ and $K_{RDC}=0$. For selections using RDCs only, $K_{CS}=0$ and $K_{RDC}=1$. For selections carried out using a combination of chemical shifts and RDCs, K_{CS} was varied until χ^2_{CS} and χ^2_{RDC} were near specified thresholds (see below) while $K_{RDC}=1$. Each selection cycle was initiated from N randomly selected conformers. A Monte Carlo (MC) simulated annealing scheme was then used to minimize the cost function χ^2 in Eq. (2) over the set of N -member ensembles. This is done by selecting at random one of the N conformers in the ensemble, suggesting a replacement for it from the pool of all generated conformers, and accepting or rejecting based on the regular Metropolis MC probability⁴². Simulations were initiated at a high “temperature” (a parametric, effective temperature), where the MC acceptance probability was high (0.99), and slowly decreased until the MC acceptance probability reached 10^{-5} . At a given effective temperature, 10^5 MC steps were carried out. The effective temperature was then decreased according to the exponential schedule $T_{n+1}=0.92 T_n$.

The next two paragraphs describe procedures to generate two types of ensembles central to our study: the dynamical ensemble and the reference ensemble. The “dynamical ensemble” of structures is a set of structures that was selected using available chemical shift and/or RDC data. The “reference ensemble” addresses the problem of checking that the SAS-generated structures in the *dynamical* ensemble are drawn from the same distribution as that from which the chemical shift data was measured.

Generating TAR apical loop *dynamical* ensembles

Five ensembles were constructed with $N=1, 2, 4, 6$ and 8 members. At each N value, a number of M independent selection cycles were carried out with the simulated annealing scheme described above, initiated each of the M times with N random structures. Subsequently, all the $N*M$ conformers were combined to form a super—ensemble, the “dynamical” ensemble. For $N=2, 4, 6$ and 8 members, $M=80, 40, 26$ and 20 selection cycles were carried out, respectively, so as to ensure that the total number of conformers selected were approximately equal. For $N=1$, the algorithm selects each of the M times, a unique structure that minimizes the cost function.

Generating TAR apical loop *reference* ensembles

The reference ensemble approach is used to test the ability of our chemical shift based selection algorithm to generate the *dynamical* ensembles above. This is important because, in the absence of single-molecule data, experimentally it is difficult to gauge the distribution of variables from bulk that typically report averages only. In this approach a known ensemble is computationally constructed and then synthetic experimental data back-calculated from the ensemble. The ‘experimental’ data is then used as an input to construct *dynamical* ensembles. The *dynamical* ensembles are then compared with the *reference* ensemble to assess their similarity. To generate the *reference* ensembles a random reference conformer is selected from the conformational pool and then a set of conformers that are within a given RMSD cutoff are randomly selected and pooled together in the *reference* ensemble with the reference structure at its center. For the TAR apical loop, we generated three sets of *reference* ensembles. The three sets of ensembles differ in the RMSD cutoff used to generate them. Specifically, ensembles with RMSD cutoff of 2, 3 and 4 Å were generated. At each cutoff value, 7 independent *reference* ensembles were generated from 7 independent reference structures. Thus, in total, 21 *reference* ensembles were generated for the TAR apical loop and each ensemble consisted of 100 conformers. To generate synthetic ‘experimental’ datasets, ^1H chemical shifts were then calculated by averaging over each of the 21 reference ensembles using SHIFTS. To simulate the presence of errors in the dataset when carrying out chemical shift based selections, ^1H chemical shifts were calculated for pool conformers using NUCHEMICS; for the set of 18 benchmark RNAs studied here the

RMSD between SHIFTS and NUCHEMICS chemical shifts ~ 0.24 ppm, which is comparable to the uncertainty in NUCHEMICS predictions (~ 0.30 ppm; see below). Using SHIFTS chemical shifts to generate the reference datasets and then NUCHEMICS chemical shifts to select ensembles therefore effectively simulates the presence ~ 0.24 ppm error in predictions. This approach to simulate errors in numerical data is similar to that used by Vendruscolo and coworkers in their study validating the use of chemical shifts to characterize the *dynamical* ensemble of the protein RNase A¹⁵.

Comparing Ensembles

To examine how well the generated *dynamical* ensembles reproduce the target, i.e., the *reference* ensembles, we employed the *S*-matrix method⁴³. In this approach one directly compares the distributions of the two ensembles. Specifically, we defined the elements of matrix $S = \{s_{ij}\}$ as a distribution overlap measure by

$$s_{ij} = |\rho_r^{ij} - \rho_d^{ij}|,$$

where ρ_r^{ij} and ρ_d^{ij} are the normalized distribution of the inter-atomic distance between atoms i and j in the *reference* and *dynamical* ensemble, respectively. s_{ij} ranges between 0 and 2 and is 0 if and only if $\rho_r^{ij} = \rho_d^{ij}$. We constructed *S*-matrices using the C1' atoms and utilized a bin-size of 0.5 Å to discretize ρ^{ij} . Ensembles were compared on the basis of the average $s_{ij} = \langle s_{ij} \rangle_A$.

RESULTS AND DISCUSSION

Accuracy of ¹H RNA chemical shift predictions

We first examined the accuracy with which RNA ¹H chemical shifts can be predicted using SHIFTS and NUCHEMICS based on an RNA structure. We note that, to our knowledge, SHIFTS ¹H chemical shift predictions have never been evaluated for RNA. For these benchmark studies, we used a panel of 18 RNA structures determined by NMR for which ¹H chemical shift assignments (H1', H2, H5, H6 and H8) are available at the Biological Magnetic Resonance Bank (<http://www.bmr.b.wisc.edu/>). This data set represents RNAs for which ¹H chemical shifts and NMR structures were deposited in the BMRB and PDB respectively following the introduction of SHIFTS and NUCHEMICS. Thus, they were not used in the development of SHIFTS and NUCHEMICS. In all cases, the ¹H chemical shifts were not used as restraints in RNA structure determination. Four additional data sets were excluded due to undocumented or incomplete chemical shift referencing (note however that including those data sets had little impact on the overall results but generally led to a deterioration in the chemical shift predictions). RNAs with modified bases were excluded because they cannot be handled by either SHIFTS or NUCHEMICS.

We used SHIFTS and NUCHEMICS to compute ¹H chemical shifts based on the NMR structure for our panel of 18 RNA structures. These structures are mainly stem-loop RNAs containing a diverse set of apical loops, ranging from four to twelve bases in size, and internal bulges of varying sequence and type. Most structures contain either single or multiple non-canonical base-pairs, and the set contains one pseudoknot riboswitch structure.

The ¹H chemical shifts were computed for every conformer in the NMR bundle. We then computed the root-mean-square-difference (RMSD) between the measured and predicted chemical shifts (CS^{RMSD}) for each conformer. Shown in Figure 1 are the lowest CS^{RMSD} values obtained over the bundle of NMR conformers for each RNA structure examined. The

overall SHIFT and NUCHEMICS CS^{RMSD} ranged between 0.26–0.54 ppm and 0.22–0.64 ppm with mean values of 0.35 and 0.34 ppm, respectively (Fig. 1). Similar CS^{RMSD} were observed when examining the agreement for individual proton types, and ranged between 0.32–0.37 ppm and 0.29–0.41 ppm with mean values of 0.32 and 0.29 ppm, respectively (Fig. S1 and Table S1). The overall and individual CS^{RMSD} compare well with the agreement originally reported for SHIFTS (0.28 ppm)²⁴. For NUCHEMICS, both the overall and individual CS^{RMSD} reported here are larger than those reported originally (0.16 ppm)²⁵. This is mostly likely due to the fact that the reported CS^{RMSD} was calculated from the same dataset used to fit the chemical shift prediction model. The Pearson correlation coefficient for individual proton types ranged between 0.47 and 0.76 and 0.61–0.80, respectively, with notably weaker correlations for H1' (Fig. S1 and Table S1). These R values are similar to those reported originally for SHIFTS²⁴ (no R values for individual protons were not reported for NUCHEMICS²⁵).

Despite the small magnitude of the CS^{RMSD} observed for individual proton types, on average they span ~ 18 % of the total chemical shift range for the H1', H2, H5, H6 and H8 protons²⁵ (Table S1). Note that this uncertainty not only include intrinsic errors in chemical shift predictions, it also includes any errors in the NMR structures, and uncertainty due to internal motional averaging, which is ignored in our analysis. While the three structures (PDBID: 2KOC, 2FDT, 1XHP) that yield the best agreement using NUCHEMICS (CS^{RMSD} = 0.19, 0.19 and 0.21 ppm respectively) also have the largest numbers of RDCs constraints per residue (~2.2 as compared to ~0.91 across all structures), a similar trend is not observed for SHIFTS. However, the overall CS^{RMSD} did decrease from 0.35 to 0.32 ppm and from 0.34 to 0.27 ppm for SHIFTS and NUCHEMICS respectively when subjecting the NMR structures to energy minimization prior to chemical shift prediction using the Generalized Born Surface Area (GBSA) implicit solvent model⁴⁰. This improvement is observed across all RNA structures and suggests that some uncertainty in the NMR structure does contribute to the observed CS^{RMSD} .

The agreement between measured and predicted 1H chemical shifts is likely also affected by motional averaging, which is not accounted for during the calculation of 1H chemical shifts. For example, for pseudoknot preQ RNA, the poor CS^{RMSD} value (0.64 ppm) improves when using the X-ray structure (0.36/0.38 ppm when using SHIFTS/NUCHEMICS respectively) or when excluding highly flexible residues (0.32 ppm when residues with a root-mean-square-fluctuation (RMSF) > 2.0 Å are excluded). However, we did not observe improved agreement when averaging the predicted CS data over the entire NMR bundle of structures (CS^{RMSD} = 0.37 ppm and 0.35 ppm for SHIFTS and NUCHEMICS respectively).

Resolving power of 1H chemical shifts

Next, we examined how well 1H chemical shifts can be used to resolve differences between competing RNA conformations. In particular, we attempted to assess whether predicted 1H chemical shifts, despite the demonstrated limitations of the chemical shifts predictors (see above), possessed sufficient resolving power to distinguish native-like from non-native RNA structure. For these studies, we used experimental 1H chemical shifts for four RNAs in our panel that contain representative RNA motifs and whose structure was determined with the use of RDCs. These include (i) a 32-nt RNA duplex structure (“duplex”) containing a canonical A-form helix determined with a large number of RDC and residual chemical shift anisotropy (RCSA) data³⁰, (ii) a 14-nt hairpin containing a UUCG tetraloop (“tetraloop”) for which a high resolution NMR structure has recently been reported based on an unprecedented amount of NMR input experimental data: nuclear Overhauser effect (NOE) derived-distances, torsion-angle dependent homonuclear and heteronuclear scalar coupling constants, cross-correlated relaxation rates and RDCs³¹, (iii) a 35-nt RNA containing an asymmetrical internal loop and flanked by two helices (“internal loop”) and determined

using NOE derived-distances and RDCs, and (iv) a 36-nt preQ₁ riboswitch RNA structure determined with the aid of RDCs that contains a pseudoknot motif (“pseudoknot”)^{33,34}. These structures fit the ¹H chemical shifts with variable agreement (the best CS^{RMSD} = 0.30/0.28, 0.28/0.21, 0.33/0.31 and 0.64/0.56 ppm for duplex, tetraloop, internal loop and pseudoknot). The four RNAs have a similar density of ¹H experimental chemical shifts (~2.8, 2.6, 2.6 and 2.8 chemical shifts per residue for duplex, 14-mer, internal loop and pseudoknot respectively).

We examined how well the agreement between the measured and predicted ¹H chemical shifts can be used to distinguish between related RNA conformations. For each of the four RNA structures, we generated a broad distribution of 8,000 conformations spanning native and denatured conformations by carrying out high temperature MD simulations (see Methods). This pool of conformations superimposes with the native structure with an average heavy atom RMSD of 6.0 ± 4.2, 3.5 ± 2.6, 7.7 ± 4.6 and 5.6 ± 3.0 Å for the duplex, tetraloop, bugle-loop and pseudoknot, respectively. ¹H chemical shifts were then calculated for each conformer within each pool using SHIFTS and NUCHEMICS. For each conformer the CS^{RMSD} value was then computed as the average of the CS^{RMSD} for each proton type and then compared to the heavy atom root-mean-square deviation between the conformer and the native, i.e. average, NMR conformation (structure^{RMSD}).

The value of CS^{RMSD} generally decreases with decreasing structure^{RMSD} particularly for structure^{RMSD} > 4 Å (Fig. 2A). These data suggest that the CS data can resolve RNA structures to within 4 Å. The continued decrease of CS^{RMSD} for structure^{RMSD} < 4 Å for UUCG suggests an even stronger structure resolving power. This is likely due to the compact nature and well-known high stability of the UUCG structure³¹ in which fluctuations away from the native structure tend to involve coordinated movements of several bases that can lead to large changes in ring current effects and therefore the predicted chemical shifts. By contrast, motions in duplex, internal loop and pseudoknot may preserve aspects of stacking interactions and therefore affect the predicted chemical shifts to a lesser extent.

Further analysis suggests that ¹H chemical shifts can resolve RNA structure to < 4 Å resolution. Out of the broad conformational pool that was generated for our four target RNAs, the conformation that best satisfies the measured ¹H chemical shifts according to SHIFTS/NUCHEMICS (i.e. conformation that yields the lowest CS^{RMSD}) superimposes with the native structures with structure^{RMSD} of 2.3/1.9, 1.4/1.4, 3.3/3.7 and 2.9/3.7 Å for duplex, tetraloop, internal loop and pseudoknot, respectively (Fig. 2B and C). Although less agreement is observed for internal loop and pseudoknot, the structure^{RMSD} improves significantly when excluding highly flexible residues (SHIFTS/NUCHEMICS structure^{RMSD} reduces to 2.9/3.0 Å and 2.3/2.6 Å, respectively). Similar results were obtained when using a weighted CS^{RMSD} in selecting conformations which weighs more heavily proton types data that exhibits stronger correlations between measured and predicted chemical shifts (Fig. 2C). Therefore, even when accounting for the inherent error in the chemical shifts predictions, we were able to resolve the RNA structure to 2–4 Å and this improves to <3 Å when neglecting highly flexible residues. Taken together, our results strongly suggest that ¹H chemical shifts can already be implemented as powerful restraints in RNA structure determination. This is despite the demonstrated limitations of the empirical methods employed by SHIFTS and NUCHEMICS to predict chemical shifts.

Use of ¹H chemical shifts in constructing RNA dynamical ensembles

In solution, chemical shifts are time-averaged over all conformations that are sampled at timescales faster than milliseconds. Studies on protein systems have established the ability to extract this dynamics information from measured chemical shift data. We therefore

examined whether ^1H chemical shifts can facilitate the determination of dynamic ensembles (see Methods) of RNA using the SAS approach⁴⁰, which we previously used to construct ensembles of RNA with the use of RDCs only⁴¹. Here, upon increasing N , we seek to find those dynamic ensembles that satisfy the time-averaged ^1H chemical shifts with the minimum value of N . This is because, while increasing N improves the minimization of the cost function, therefore seemingly increasing the accuracy, we do not wish to go beyond the inherent finite accuracy of chemical shift predictors. This therefore gives us the smallest N that produces accuracy within the threshold of 0.24 ppm (see below). In this approach, N conformers are randomly selected from a pool typically generated using MD simulations, and the agreement between measured and predicted ^1H chemical shift data is computed. Next, one conformer is randomly replaced with another conformer from the pool, and the agreement with measured ^1H chemical shift data is re-examined and the newly selected conformer is either accepted or rejected based on the Metropolis criteria. Using such a MC based approach, several iterations are carried out until convergence is reached, defined as achieving agreement between measured and calculated data to within the specified error (see below).

We examined the utility of ^1H chemical shifts in constructing RNA dynamic ensembles using simulated chemical shift and a known target *reference* ensemble. We used replica-exchange molecular dynamics (REMD) simulations³⁸ to generate a broad conformation pool for the TAR apical loop. The TAR apical loop has previously been shown to undergo complex motions at multiple timescales and therefore provides a good model system for testing this approach⁴⁴. We then generated 21 *reference* ensembles that feature different levels of dynamics by randomly selecting a reference conformer from the 45,000 membered pool and then randomly selecting 100 conformers that are within 2, 3 and 4 Å of the reference conformer. In so doing we generated a total of 21 *reference* ensembles (see Methods). For each *reference* ensemble, ‘experimental’ ensemble-averaged H1’, H2, H5, H6 and H8 chemical shifts were computed using SHIFTS. Selections on the other hand were carried out using NUCHEMICS predicted shifts. This corresponds to ~0.24 ppm prediction error, as judged based on comparison of the average CS^{RMSD} between SHIFTS and NUCHEMICS for the 18 benchmark RNAs studied herein (see Methods). One bond C-H RDCs were also computed assuming a fixed alignment tensor determined experimentally in Pf-1 phage. The RDCs were noise corrupted by adding Gaussian white noise with standard deviation of 2.0 Hz corresponding to the uncertainty in RDC measurements in elongated RNA⁴⁵. In order to minimize overlap between *reference* ensembles, we skewed the conformational pool by replicating underrepresented conformers prior to the selection of *dynamical* ensembles.

In all cases, convergence was achieved for the chemical shift based selections at $N=2$ ($\text{CS}^{\text{RMSD}} = 0.13, 0.11$ and 0.10 ppm for the 2, 3 and 4 Å *reference* ensembles; Table 1). We found that increasing the value of N for chemical shift selections did not lead to significant improvements in the chemical shift predictions (Table 1). By comparison, $N \sim 8$ was required to achieve convergence for RDC and chemical shift+RDC selections; the RDC^{RMSD} for the 2, 3 and 4 Å ensemble was 1.71, 1.72 and 1.72, and 1.73, 1.73 and 1.66 Hz, respectively. We next investigated whether the chemical shift ensembles were able to recapitulate the *reference* ensembles RDCs. For the $N=2$ ensemble the $\text{RDC}^{\text{RMSD}} = 15.8, 14.1$, and 14.1 Hz for the 2, 3 and 4 Å ensembles, respectively, and increasing N did not result in any significant improvement in RDC agreement (Table 1). The chemical shift ensembles therefore were unable to satisfy the RDCs to within the 2.0 Hz error thresholds; a similar trend was observed when back-predicting RDCs from ensemble constructed using experimental chemical shift (data not show). In contrast, the RDC ensembles predicted the chemical shifts to within the 0.24 ppm threshold ($\text{CS}^{\text{RMSD}} = 0.25, 0.24$ and 0.23 for the 2, 3 and 4 Å ensembles, respectively), suggesting that though chemical shifts based *dynamical*

ensembles may not themselves attain the required accuracy, chemical shifts maybe used to validate or invalidate *dynamical* ensembles generated using other input structural data (e.g. RDCs).

To further interrogate the chemical shift *dynamical* ensembles, we used the **S**-matrix method (Methods) to determine their structural similarity with the *reference* ensembles. We observed that for $N=2$ the $\langle s_{ij} \rangle_A$ was 0.88, 0.78 and 0.84 for the 2, 3 and 4 Å *reference* ensembles (Table 2). Increasing N did not result in any significant enhancement in the overlap between the chemical shift and *reference* ensembles (Table 2). By comparison, the $\langle s_{ij} \rangle_A$ for randomly selected ensembles were 1.06, 0.93 and 0.78, indicating that there was better correspondence between chemical shift based *dynamical* ensemble and *reference* ensembles than randomly selected *dynamical* ensembles and *reference* ensembles. However, the RDC, as well as the, chemical shift+RDC ensembles, exhibited much better overlap with the *reference* ensembles; for the $N=8$, $\langle s_{ij} \rangle_A$ for 2, 3 and 4 Å ensembles was 0.44, 0.44 and 0.43, and 0.44, 0.41 and 0.39, respectively.

Taken together, although the chemical shift-based *dynamical* ensembles exhibited greater resemblance to the *reference* ensembles than the randomly constructed ensembles, they were unable to achieve the same degree of overlap as the RDC and chemical shift+RDC, and consequently were unable to adequately predict the *reference* ensemble RDCs. These effects can be attributed to the comparatively larger error threshold used to define convergence for chemical shifts (threshold is ~22% of chemical shift total range) as compared to RDCs (threshold is ~2.5% of the RDC range). Indeed, repeating the simulations with zero error resulted in chemical shift *dynamical* ensembles that exhibited enhanced overlap with the *reference* ensembles, and thus, better predicted *reference* ensembles RDCs (data not shown). The ability of a chemical shift *dynamical* ensemble to recover the *reference* ensemble is therefore limited by the accuracy of chemical shifts predictions. Currently, SHIFTS and NUCHEMICS predict ^1H chemical shifts to within ~ 0.30 ppm, slightly higher than the 0.24 ppm error threshold used to determine convergence in the theoretical simulations. Significant improvement in the accuracy of the ^1H chemical shift predictions will be needed in order to construct accurate *dynamical* ensembles of RNAs based solely on chemical shift data.

Conclusions

NMR structure determination of nucleic acids has traditionally been challenging due to the paucity of inter-proton NOE-derived distance restraints, extended nature of the structures, and high degree of flexibility. There has been a long-standing quest to measure different sources of structural information, and indeed, the measurement of RDCs has revolutionized the ability to determine the structure and dynamics of nucleic acids. There is now renewed interest in utilizing NMR chemical shifts to solve RNA structure, as they are the most accessible and accurately measured NMR observable. In this report we demonstrated that ^1H chemical shifts can be used to resolve RNA structure, allowing discrimination of native structure from non-native states. We show using SHIFTS and NUCHEMICS, which on average predict ^1H chemical shifts to within 0.30 ppm, that ^1H chemical shifts can be used to resolve with to within 2–4 Å, and <3 Å when neglecting highly flexible residues. ^1H chemical shifts can immediately be used for RNA structure validation. In time, as more accurate ^1H chemical shift prediction methods emerge the resolution limit should decrease well below 4 Å. When combined with improvements in RNA structure prediction, we can anticipate that methodologies such as CS-ROSETTA will evolve that allow high-resolution RNA structure determination based on chemical shift data alone.

We also investigated whether ^1H chemical shifts could be used to generate accurate *dynamical* ensembles of RNAs. Using theoretical simulation on the hexa-nucleotide HIV-1

TAR apical loop our results indicate that though *dynamical* ensembles constructed using ^1H chemical shifts exhibited greater structural overlap with known *reference* ensembles than randomly constructed *dynamical* ensembles, they failed to achieve the same degree of overlap as the corresponding RDC *dynamical* ensembles. This result hinted to an inherent degeneracy in the chemical shifts *dynamical* ensembles and in fact, we observe that the chemical shift based *dynamical* ensembles were unable to reproduce the RDCs back-calculated from the *reference* ensembles. Here again, more accurate ^1H chemical shifts prediction methods should enable more accurate *dynamical* ensembles to be generated, as should the incorporation of chemical shifts from other nuclei e.g. ^{13}C , ^{15}N and ^{31}P .

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

H.M.A. acknowledges funding from NIH NIGMS R21GM096156. I.A acknowledges funding the US National Science Foundation (NSF Career Award CHE-0918817). A.T.F was supported through NSF Graduate Research Fellowship.

References

1. de Dios AC, Pearson JG, Oldfield E. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science*. 1993; 260:1491–1496. [PubMed: 8502992]
2. Wishart DS, Case DA. Use of chemical shifts in macromolecular structure determination. *Meth Enzymol*. 2001; 338:3–34. [PubMed: 11460554]
3. Case DA. Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR*. 1995; 6:341–346. [PubMed: 8563464]
4. Cornilescu GG, Delaglio FF, Bax AA. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*. 1999; 13:289–302. [PubMed: 10212987]
5. 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.
6. Le H, Pearson JG, de Dios AC, Oldfield E. Protein structure refinement and prediction via NMR chemical shifts and quantum chemistry. *J. Am. Chem. Soc.* 1995; 117:3800–3807.
7. Shen Y, Bryan PN, He Y, Orban J, Baker D, Bax A. De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci*. 2010; 19:349–356. [PubMed: 19998407]
8. Shen Y, Lange O, Delaglio F, Rossi P, Aramini J, Liu G, Eletsky A, Wu Y, Singarapu K, Lemak A, Ignatchenko A, Arrowsmith C, Szyperski T, Montelione G, Baker D, Bax A. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences*. 2008; 105:4685–4690.
9. Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G. CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res*. 2008; 36:W496–W502. [PubMed: 18515350]
10. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA*. 2007; 104:9615–9620. [PubMed: 17535901]
11. Rule, GS. *Fundamentals of Protein NMR Spectroscopy (Focus on Structural Biology)* [Hardcover]. Springer; 2005.
12. Chary, KVR.; Govil, G. *NMR in Biological Systems: From Molecules to Human (Focus on Structural Biology)*. 1st ed.. Springer; 2008.
13. Berjanskii MM, Wishart DSD. NMR: prediction of protein flexibility. *CORD Conference Proceedings*. 2006; 1:683–688.

14. Berjanskii MV, Wishart DS. The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.* 2007; 35:W531–W537. [PubMed: 17485469]
15. Camilloni C, Robustelli P, Simone AD, Cavalli A, Vendruscolo M. Characterization of the Conformational Equilibrium between the Two Major Substates of RNase A Using NMR Chemical Shifts. *J. Am. Chem. Soc.* 2012; 134 120222164411006.
16. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure.* 2010; 18:494–506. [PubMed: 20399186]
17. Jensen MR, Salmon L, Nodet G, Blackledge M. Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J. Am. Chem. Soc.* 2010; 132:1270–1272. [PubMed: 20063887]
18. Ball KA, Phillips AH, Nerenberg PS, Fawzi NL, Wemmer DE, Head-Gordon T. Homogeneous and heterogeneous tertiary structure ensembles of amyloid- β peptides. *Biochemistry.* 2011; 50:7612–7628. [PubMed: 21797254]
19. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol.* 2011; 21:426–431. [PubMed: 21530234]
20. Li D-W, Brüschweiler R. Certification of Molecular Dynamics Trajectories with NMR Chemical Shifts. *The Journal of Physical Chemistry Letters.* 2010; 1:246–248.
21. Li D-W, Brüschweiler R. NMR-Based Protein Potentials. *Angew. Chem. Int. Ed.* 2010; 49:6778–6780.
22. Robustelli P, Stafford KA, Palmer AG III. Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J. Am. Chem. Soc.* 2012; 134:6365–6374. [PubMed: 22381384]
23. Lam SL, Chi LM. Use of chemical shifts for structural studies of nucleic acids. *Prog Nucl Magn Reson Spectrosc.* 2010; 56:289–310. [PubMed: 20633356]
24. Dejaegere A, Bryce RA, Case DA. An empirical analysis of proton chemical shifts in nucleic acids. *ACS Symposium Series.* 1999; 732:194–206.
25. Cromsig JA, Hilbers CW, Wijmenga SS. Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J Biomol NMR.* 2001; 21:11–29. [PubMed: 11693565]
26. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR.* 2007; 38:289–302. [PubMed: 17610132]
27. Kohlhoff KJ, 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. [PubMed: 19739624]
28. 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. [PubMed: 21086975]
29. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL. BioMagResBank. *Nucleic Acids Res.* 2008; 36:402–408.
30. Tolbert BS, Miyazaki Y, Barton S, Kinde B, Starck P, Singh R, Bax A, Case DA, Summers MF. Major groove width variations in RNA structures determined by NMR and impact of ^{13}C residual chemical shift anisotropy and ^1H - ^{13}C residual dipolar coupling on refinement. *J Biomol NMR.* 2010; 47:205–219. [PubMed: 20549304]
31. Nozinovic S, Fürtig B, Jonker H, Richter C, Schwalbe H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* 2010; 38:683. [PubMed: 19906714]
32. Zhang Q, Kim NK, Peterson RD, Wang Z, Feigon J. Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc Natl Acad Sci USA.* 2010; 107:18761–18768. [PubMed: 20966348]
33. Kang M, Peterson R, Feigon J. Structural Insights into Riboswitch Control of the Biosynthesis of Queuosine, a Modified Nucleotide Found in the Anticodon of tRNA. *Molecular Cell.* 2009; 33:784–790. [PubMed: 19285444]

34. Zhang Q, Kang M, Peterson RD, Feigon J. Comparison of Solution and Crystal Structures of PreQ1 Riboswitch Reveals Calcium-Induced Changes in Conformation and Dynamics. *J. Am. Chem. Soc.* 2011; 133:5190–5193. [PubMed: 21410253]
35. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation.* 2008; 4:435–447.
36. Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 1995; 117:5179–5197.
37. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 79:926–935.
38. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters.* 1999; 314:141–151.
39. Das. Baker Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci USA.* 2007
40. Chen Y, Campbell S, Dokholyan N. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J.* 2007
41. Frank AT, Stelzer AC, Al-Hashimi HM, 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. [PubMed: 19369218]
42. Metropolis N, Rosenbluth AW, Rosenbluth NN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 1953; 21:1087–1092.
43. De Simone A, Richter B, Salvatella X, Vendruscolo M. Toward an accurate determination of free energy landscapes in solution states of proteins. *J. Am. Chem. Soc.* 2009; 131:3810–3811. [PubMed: 19292482]
44. Dethoff E, Hansen A, Musselman C, Watt E, Andricioaei I, Al-Hashimi H. Characterizing Complex Dynamics in the TAR Apical Loop and Motional Correlations with the Bulge by NMR, MD and Mutagenesis. *Biophys J.* 2008
45. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature.* 2007; 450:1263–1267. [PubMed: 18097416]

Figure 1.

Measured vs. predicted chemical shift in RNA database. Root mean square difference (CS^{RMSD}) between experimental and (A) NUCHEMICS and (B) SHIFTS predicted chemical shifts prior to (red) and post (blue) minimization of input structure. In cases where multiple conformers exist we report the results for the structure that gives the lowest CS^{RMSD} after minimization.

Figure 2.

Resolving structure using 1H chemical shifts. (A) Correlation between SHIFTS (red) and NUCHEMICS (black) CS^{RMSD} and $structure^{RMSD}$ for the A-form duplex, UUCG tetraloop, hTR internal loop, and pseudoknot. CS^{RMSD} and $structure^{RMSD}$ are calculated over conformational pool consisting of native-like and unfolded conformers (see text). Plot where made by binning data along $structure^{RMSD}$. Bin widths were 0.50 Å. Averages and error were calculated over data in each bin. (B) Correlation plots between measured and predicted SHIFTS and NUCHEMICS chemical shifts for the A-form duplex, UUCG tetraloop, internal loop and pseudoknot. In each case, the chemical shifts shown are those calculated from conformer with lowest CS^{RMSD} , which is indicated on each plot. (C) Structural overlay of the average NMR structure with conformers with the lowest SHIFTS and NUCHEMICS unweighted (left) and weighted (right) CS^{RMSD} . The NMR, SHIFTS, and NUCHEMICS structures are shown in blue, red and black, respectively. The $structure^{RMSD}$ between the structures is indicated below each.

Table 1

Back-predicting chemical shifts and RDCs from chemical shift based *dynamical* ensembles. Shown are the RMSD and Pearson correlation coefficient (*R*) between chemical shifts calculated from the *reference* ensembles and those calculated from the 2, 3 and 4 Å chemical shift based *dynamical* ensembles. For comparison, the RMSD and *R* between RDCs calculated from the *reference* ensembles and those calculated from the chemical shift based *dynamical* ensembles are also shown.

N	CS: RMSD (ppm)/R			RDC: RMSD (Hz)/R		
	2 Å	3 Å	4 Å	2 Å	3 Å	4 Å
1	0.24/0.97	0.25/0.97	0.26/0.97	25.5/0.77	28.1/0.73	25.9/0.78
2	0.13/0.99	0.11/0.99	0.10/0.99	15.8/0.89	14.1/0.91	14.1/0.91
4	0.10/0.99	0.09/0.99	0.08/0.99	15.4/0.89	12.6/0.92	13.2/0.92
6	0.09/0.99	0.08/0.99	0.08/0.99	14.7/0.90	12.2/0.92	12.5/0.92
8	0.09/0.99	0.08/0.99	0.07/0.99	14.4/0.90	11.3/0.92	12.3/0.92

Table 2

Overlap between chemical shifts based *dynamical* ensembles and *reference* ensembles. Shown are the $\langle s_{ij} \rangle_A$ between chemical shift based *dynamical* ensembles and the 2, 3 and 4 Å *reference* ensembles for $N=1, 2, 4, 6$ and 8. For $N=8$, $\langle s_{ij} \rangle_A$ between the *reference* ensemble and RDC and chemical shift+RDC based *dynamical* ensembles are included. For comparison, the $\langle s_{ij} \rangle_A$ between randomly selected ensembles and the *reference* ensembles are included.

N	2 Å	3 Å	4 Å
1	1.47	1.59	1.56
2	0.88	0.78	0.84
4	0.76	0.64	0.65
6	0.75	0.61	0.56
8 ^a	0.75/0.44/0.44	0.59/0.44/0.41	0.53/0.43/0.39
random	1.06	0.93	0.78

^afor comparison $\langle S_{ij} \rangle_A$ for the RDC and chemical shift+RDC ensemble are also listed