JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

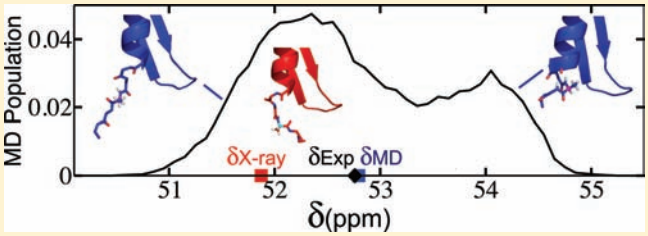
Interpreting Protein Structural Dynamics from NMR Chemical Shifts

Paul Robustelli, Kate A. Stafford, and Arthur G. Palmer, III*

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

Supporting Information

ABSTRACT: In this investigation, semiempirical NMR chemical shift prediction methods are used to evaluate the dynamically averaged values of backbone chemical shifts obtained from unbiased molecular dynamics (MD) simulations of proteins. MD-averaged chemical shift predictions generally improve agreement with experimental values when compared to predictions made from static X-ray structures. Improved chemical shift predictions result from population-weighted sampling of multiple conformational states and from sampling smaller fluctuations within conformational basins. Improved chemical shift predictions also result from discrete changes to conformations observed in X-ray structures, which may result from crystal contacts, and are not always reflective of conformational dynamics in solution. Chemical shifts are sensitive reporters of fluctuations in backbone and side chain torsional angles, and averaged ^{1}H chemical shifts are particularly sensitive reporters of fluctuations in aromatic ring positions and geometries of hydrogen bonds. In addition, poor predictions of MD-averaged chemical shifts can identify spurious conformations and motions observed in MD simulations that may result from force field deficiencies or insufficient sampling and can also suggest subsets of conformational space that are more consistent with experimental data. These results suggest that the analysis of dynamically averaged NMR chemical shifts from MD simulations can serve as a powerful approach for characterizing protein motions in atomistic detail.



INTRODUCTION

Efforts to understand and quantify the structural features that influence the NMR chemical shifts of nuclei in proteins have spanned several decades.[1−4] Progress in these efforts has precipitated the development of several methods that successfully map the dependency of protein chemical shifts on a multitude of structural factors with increasing speed and accuracy.[5−12] These chemical shift prediction tools have enabled computational methods for calculating protein structures using only chemical shifts as restraints. Initial implementations of these methods utilized a structural homology driven molecular fragment replacement approach.[13−15] More recently, chemical shifts have been successfully implemented in restrained conformational searches[16,17] in a manner that is identical to and compatible with conventional NOE and RDC driven NMR structure calculation protocols.

Isotropic chemical shifts, which reflect averages from ensembles of molecules on the millisecond time scale, have also been gaining attention as probes of conformational dynamics. Chemical shifts have been utilized to predict S^2 order parameters[18,19] reflecting their sensitivity to picosecond−nanosecond (ps−ns) time scale fluctuations, and to predict backbone[20,21] and side chain[22−24] dihedral angle population distributions, reflecting their sensitivity to conformational averaging on the millisecond time scale. Prediction of chemical shifts from structural ensembles of proteins derived from molecular dynamics (MD) simulations can result in improved agreements with experimental measurements when compared to chemical shifts predicted from individual static structures,[25,26] suggesting that available methods for predicting chemical shifts from protein structures may be sufficiently accurate and sensitive to conformational fluctuations to describe structural dynamics in proteins.

To elucidate the specific motional processes observed in MD simulations of proteins that are well described by currently available chemical shift prediction methods, we have conducted a detailed analysis of the dynamically averaged values of chemical shifts predicted from molecular dynamics simulations of proteins. In this investigation, we have focused on simulations of a pair of structurally homologous ribonuclease H enzymes from *Escherichia coli* (ecRNH) and *Thermus thermophilus* (ttRNH), which have been shown to possess substantial conformational dynamics on the ps−ns and μs−ms time scales.[27,28] We find that, on average, chemical shift values predicted from MD simulations are in better agreement with experimental values than those predicted from static structures. To identify the specific conformational changes and motional processes associated with improved agreement, we have examined the conformational fluctuations in the regions of the proteins where MD averaged chemical shift predictions provide the largest improvements over predictions from static structures. We observe that improved chemical shift predictions can result from a population weighted sampling of multiple conformational states and from sampling fluctuations within

Received: January 10, 2012
Published: March 1, 2012

   dx.doi.org/10.1021/ja300265w | *J. Am. Chem. Soc.* 2012, 134, 6365−6374

individual conformational basins. We find that chemical shifts are sensitive reporters of fluctuations in backbone and side chain torsional angles, and averaged ^{1}H chemical shifts are particularly sensitive reporters of fluctuations in aromatic ring positions and the presence and geometries of hydrogen bonds.

Importantly, we also find that improved chemical shift predictions can result from discrete changes to conformations observed in X-ray structures, which may result from crystal contacts, and may not be reflective of conformational dynamics in solution. These results emphasize the importance of identifying the specific structural and dynamic features of an MD simulation that influence the chemical shift predictions. In addition, we demonstrate that poor predictions of averaged chemical shifts can identify potentially spurious conformations and motions observed in MD simulations that may result from force field deficiencies or insufficient sampling and can also suggest subsets of conformational space that are more consistent with experimental data.

RESULTS AND DISCUSSION

In this investigation, we used several of the most recently developed and best performing chemical shift prediction tools, Sparta+, Shiftx2, ShiftX+, Camshift, and ShiftS,[5−12] to calculate the average values of backbone Cα, Cβ, C′, Hα, HN, and N chemical shifts predicted from unbiased explicit solvent MD simulations of a pair of structurally homologous Ribonuclease H enzymes from *E. coli* (ecRNH) and *T. thermophilus* (ttRNH). MD simulations of varying lengths, ranging from 100 to 1000 ns, were run using the Amber99SB[29] and Amber99SB-ILDN[30] force fields in TIP3P water.[31] Details of the MD simulation protocols and the preparation of the coordinates of each MD snapshot and X-ray crystal structure for input into the chemical shift predictors are presented in the Materials and Methods section. In all of the analyses presented here, the average chemical shift value reported for an MD trajectory (δMD) is determined by linear averaging of the predicted chemical shift for each snapshot saved in the trajectory.

Dynamic Averaging of Chemical Shifts from Molecular Dynamics Simulations of Proteins. A detailed analysis of the average chemical shifts predicted from MD simulations of the globular proteins ttRNH and ecRNH demonstrates that averaged chemical shifts, as predicted by the most recently developed semiempirical shift prediction methods, provide atomistic structural descriptions of a variety of conformational changes and dynamic processes in proteins. A comparison of the chemical shift prediction root mean square deviation (rmsd) values from experimental values obtained from X-ray structures and from unbiased 100 ns MD simulations using the amber99SB force field are shown for ecRNH and ttRNH in Table 1.

We observe a remarkable improvement in Sparta+ prediction rmsd values across all backbone atom types when comparing MD averaged predictions to predictions made from the 2.8 Å resolution X-ray structure (PDB code 1RIL)[32] of ttRNH. Smaller but consistent improvements were obtained in the case of ecRNH, where a higher resolution 1.5 Å X-ray structure was available (PDB code 2RN2).[33]

Recent analyses of chemical shifts predicted from MD simulations have focused on improvements of shift predictions as quantified by correlation coefficients or rmsd's from experimental values for given atom types across an entire protein.[25,26] We find that considerable additional insight is gained by examining structural dynamics in the regions of the proteins where MD averaged chemical shift predictions (δMD) produce the largest differences compared to predictions obtained from static X-ray crystal structures (δX-ray). Figure 1 compares the deviations between the experimentally measured values (δExp) of Cα and HN ttRNH chemical shifts, the δX-ray predictions, and the δMD predictions from a 100 ns simulation using the amber99SB force field. A substantial number of residues show differences between δX-ray and δMD that are very large relative to the reported prediction accuracy obtained by Sparta+ applied to databases of static structures (0.49 and 0.94 ppm for HN and Cα nuclei, respectively). This suggests the presence of substantial conformational dynamics in the MD trajectory or a large change in the average conformation of the protein observed in the MD trajectory compared to the X-ray structure. Similar plots are presented for the additional backbone nuclei of ttRNH in Supporting Information Figure S1 and for all available backbone nuclei of ecRNH in Supporting Information Figure S2.

Table 1. Comparison of rmsd's between Experimental Chemical Shifts and Sparta+ Chemical Shift Predictions Obtained from X-ray Structures and MD Simulations[a]

	Sparta+ prediction rmsd					
	Cα	Cβ	C′	HN	Hα	N
			ttRNH			
X-ray	1.00	1.14	1.21	0.44	0.32	2.70
MD	0.68	1.01	1.06	0.35	0.23	2.17
			ecRNH			
X-ray	0.74	-	-	0.39	0.25	2.51
MD	0.70	-	-	0.36	0.25	2.25

[a]PDB entries 1RIL and 2RN2 were used for X-ray predictions of ttRNH and ecRNH, respectively. MD values were calculated from unbiased 100 ns simulations using the amber99SB force field in TIP3P water.

Protein Structural Dynamics That Improve Chemical Shift Predictions. In the case of carbon nuclei, the largest differences between δX-ray and δMD are located in loop regions and at the ends of secondary structure elements, reflecting the primary dependence of carbon shifts on backbone ϕ and ψ angles. In Figure 2, the time course and distribution of the chemical shift predictions of Ala 145 Cα are shown for a 100 ns Amber99SB MD trajectory of ttRNH along with representative conformations from the MD trajectory and the X-ray structure. The time course and distribution of the Ala 145 Cα shift suggests that the presence of an equilibrium distribution of helical and nonhelical conformations, or helix "fraying", at the end of the C-terminal helix of ttRNH produces a more accurate prediction of the chemical shifts than the coil conformation observed in the crystal structure. Similarly close agreement is observed between the MD averaged predictions and the experimental measurements for all of the backbone nuclei of Ala 145 (Supporting Information Figure S3). Remarkably, the relative populations of the helical and coil states obtained within 100 ns of simulation produce chemical shift predictions that agree very well with the experimentally measured chemical shift, suggesting that an accurate estimate of the relative populations observed in solution has been obtained, within the limitations of the accuracies of the shift predictor. A second example where averaged chemical shifts appear to be a useful reporter of fraying terminal elements of secondary structure is shown in Figure 3. Here, the time course and distribution of the chemical shift predictions of Glu 4 HN

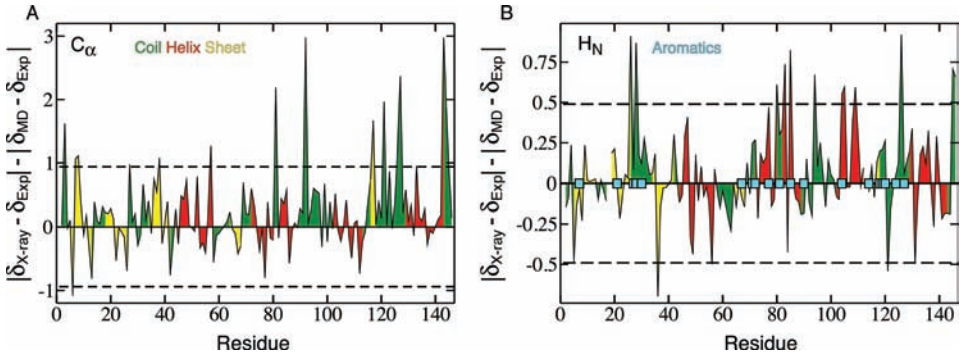


Figure 1. Comparison of Sparta+ chemical shift predictions for ttRNH obtained from an X-ray structure and a 100 ns MD simulation in the amber99SB force field. For each residue for which an experimentally measured chemical shift (δExp) was available, the magnitude of the deviation between the X-ray predicted value (δX-ray) and the experimental value, $|\delta$X-ray $- \delta$Exp$|$, and the magnitude of the deviation between the MD averaged prediction (δMD) and the experimental value, $|\delta$MD $- \delta$Exp$|$, are compared. $|\delta$X-ray $- \delta$Exp$| - |\delta$MD $- \delta$Exp$|$ is shown for Cα atoms in panel A and HN atoms in panel B. A positive value of $|\delta$X-ray $- \delta$Exp$| - |\delta$MD $- \delta$Exp$|$ indicates that δMD is in better agreement with experiment, while a negative value indicates that δX-ray is in better agreement. Residues are colored according to their secondary structure in the X-ray structure with green, red, and yellow corresponding to coil, helix, and sheet, respectively. Residues with aromatic side chains are displayed as cyan squares in panel B. The reported standard deviation of Sparta+ predictions obtained from a benchmark database of X-ray structures for Cα and HN atoms are displayed as dotted lines on for comparison.

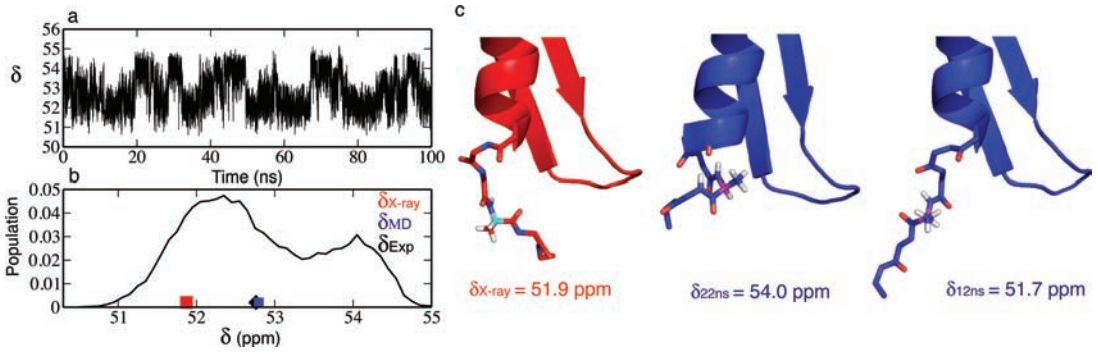


Figure 2. Chemical shift predictions of Ala 145 Cα in a 100 ns MD simulation of ttRNH in the amber99SB force field. (a) The value of the Sparta+ predicted chemical shift of Ala 145 Cα for snapshots saved every 4.5 ps of the MD trajectory. (b) The normalized distribution of the Sparta+ predicted shifts of Ala 145 Cα from the MD trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 1RIL) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (c) The conformation of Ala 145 observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding Sparta+ shift prediction of Ala 145 Cα for those conformations. Ala 145 Cα is colored cyan in the X-ray structure and magenta in the MD snapshots.

obtained from a 1 μs simulation of ecRNH demonstrates that the populations of a β-sheet conformation and more a disordered coil conformation observed in the simulation produce an excellent agreement between the predicted and experimental chemical shift.

The largest differences between δX-ray and δMD predictions for ^{1}H spins are clustered around aromatic side chains and reflect the large effects of ring currents. A recent study has demonstrated that the functional forms implemented in semiempirical chemical shift predictors to predict the magnitudes of ring current effects have been parametrized with a high level of accuracy, resulting in rmsd's of less than 0.1 ppm when compared to benchmark quantum calculations.[34] The effect that the dynamics of aromatic side chains can have on ^{1}H shifts is illustrated in Figure 4. Here, the time course and distribution of the chemical shift predictions of His 28 HN obtained from a 100 ns Amber99SB simulation of ttRNH are shown with the time course of the isolated ring current contribution to the chemical shift and representative conformations from the MD trajectory and the X-ray structure. The magnitude of the ring current effects experienced by the HN proton fluctuates substantially with the conformations of the proximal aromatic side chains throughout the trajectory. The agreement between the MD averaged chemical shift prediction and the experimental value, compared to the chemical shift predicted from the X-ray structure, suggests that the conformation observed in the X-ray structure is not significantly populated in solution. These results provide support for the relative populations of the side chain conformations observed in the simulation. The magnitude of the ring current effects observed in ttRNH demonstrates that a less detailed interpretation of dynamically averaged proton chemical shifts, in terms of their deviation from random coil values, used to infer secondary structure propensities may overlook and misinterpret significant contributions due to ring current effects.

The largest improvements between δX-ray and δMD for N shifts, which are sensitive to χ_1 side chain dihedral angles, are primarily due to side chain dynamics and sampling of multiple rotameric states in simulations. An example is presented in Figure 5 for the N atom of the ecRNH residue Val 5 from a 450 ns Amber99SB-ILDN MD simulation. The improvements in

dx.doi.org/10.1021/ja300265w | J. Am. Chem. Soc. 2012, 134, 6365−6374

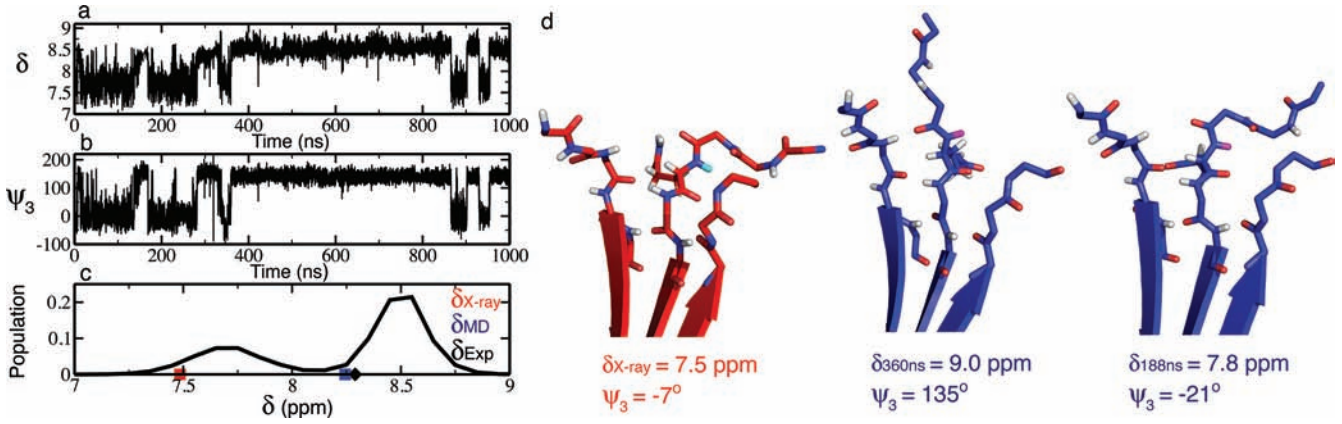


Figure 3. Chemical shift predictions of Glu 4 HN in a 1 μs MD simulation of ecRNH in the amber99SB force field. (a) The value of the Sparta+ predicted chemical shift of Glu 4 HN for snapshots saved every 4.5 ps of the MD trajectory. (b) The value of the ψ dihedral angle of Val 3 (ψ_3) for snapshots saved every 4.5 ps of the MD trajectory. (c) The normalized distribution of the Sparta+ predicted shifts of Glu 4 HN from the trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 2RN2) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (d) The conformation of Glu 4 observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding Sparta+ shift prediction of Glu 4 HN and the ψ value of Val 3 for those conformations. Glu 4 HN is colored cyan in the X-ray structure and magenta in the MD snapshots.

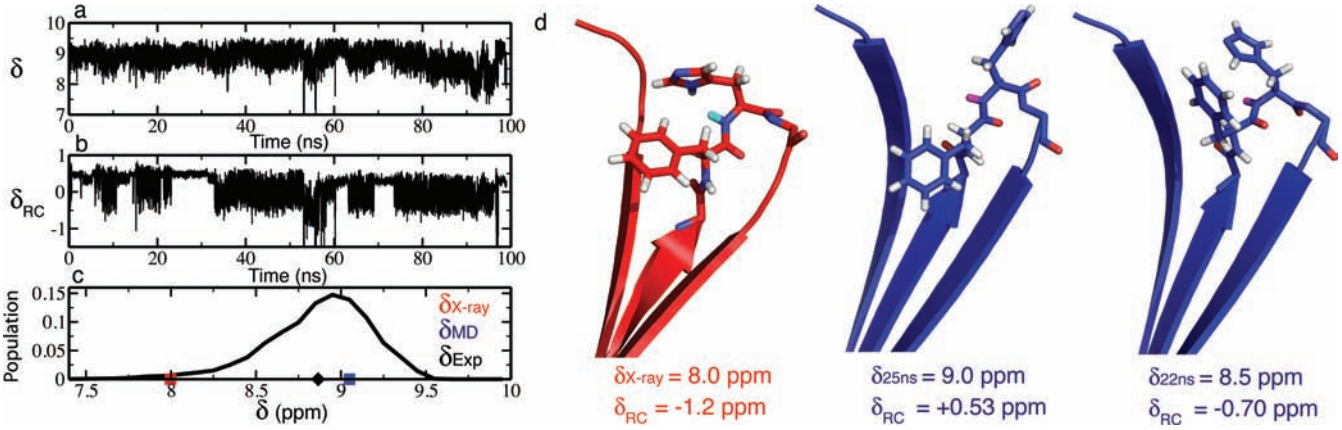


Figure 4. Chemical shift predictions of His 28 HN in a 100 ns MD simulation of ttRNH in the amber99SB force field. (a) The value of the Sparta+ predicted chemical shift of His 28 HN for snapshots saved every 4.5 ps of the MD trajectory. (b) The value of the contribution of the ring currents to the chemical shift of His 28, δ_{RC}, as calculated by Sparta+ for snapshots saved every 4.5 ps of the MD trajectory. (c) The normalized distribution of the Sparta+ predicted shifts of His 28 HN from the MD trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 1RIL) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (d) The conformation of His 28 observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding Sparta+ shift prediction of His 28 HN and the δ_{RC} value for those conformations. His 28 HN is colored cyan in the X-ray structure and magenta in the MD snapshots.

the N shift predictions suggest that MD simulations of 100−1000 ns can obtain reasonably accurate population estimates of the distribution of χ_1 torsion angles for side chain rotamers.

Changes in X-ray Structure Conformations That Improve Chemical Shift Predictions. The analysis of the time-course and distributions of individual chemical shift values is particularly useful for determining if an improvement in a chemical shift prediction is the result of a discrete change in a conformation observed in a crystal structure, such as those that arise from crystal packing, rather than significant conformational dynamics in solution. As shown in Figure 6, the time course and distribution of the chemical shift prediction of ttRNH Phe 27 C′ suggest that the simulation originates from a conformation that is incompatible with the experimental solution chemical shift, but ∼30 ns into the simulation, a conformational change occurs that produces predicted shifts in much better agreement with the experimental value. This conformational change involves the formation of a hydrogen bond not present in the crystal structure, and examination of the time course and of distribution of δMD of the trajectory suggests that this conformation is likely stable in solution. That is, the chemical shift predictions suggest that populating only the second conformation would produce a predicted average chemical shift that is closer to the experimental value than the average chemical shift prediction resulting from the sampled population distribution of the initial conformation and the second conformation observed in the simulation. It is interesting to note that in this instance, the chemical shift predictions of His 30 HN, which forms the hydrogen bond with Phe 27, do not exhibit the two state behavior observed in Phe 27 C′ which enable the discrimination between the conformations. An examination of the structural factors which

dx.doi.org/10.1021/ja300265w | J. Am. Chem. Soc. 2012, 134, 6365−6374

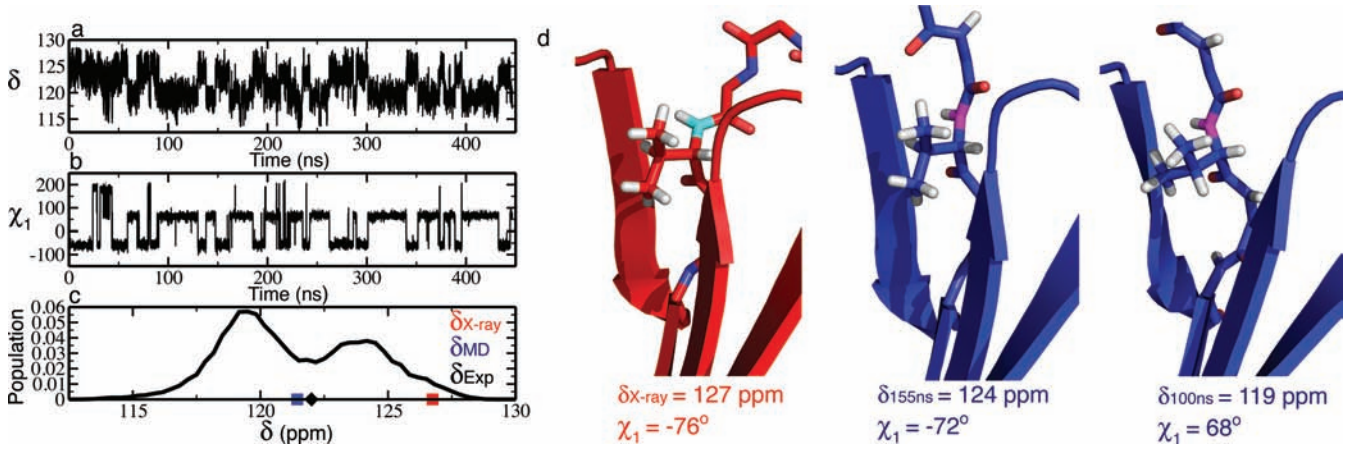


Figure 5. Chemical shift predictions of Val 5 N in a 450 ns MD simulation of ecRNH in the amber99SB-ILDN force field. (a) The value of the Sparta+ predicted chemical shift of Val 5 N for snapshots saved every 4.5 ps of the MD trajectory. (b) The value of the χ_1 dihedral angle of Val 5 for snapshots saved every 4.5 ps of the MD trajectory. (c) The normalized distribution of the Sparta+ predicted shifts of Val 5 N from the MD trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 2RN2) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (d) The conformation of Val 5 observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding SPARTA+ shift prediction of Val 5 N and Val 5 χ_1 value for those conformations. Val 5 N is colored cyan in the X-ray structure and magenta in the MD snapshots.

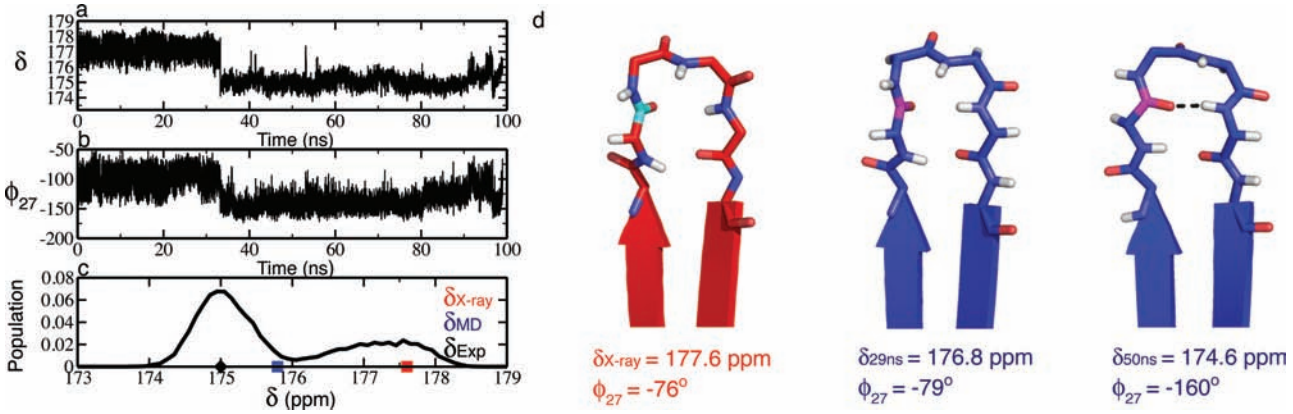


Figure 6. Chemical shift predictions of Phe 27 C′ in a 100 ns MD simulation of ttRNH in the amber99SB force field. (a) The value of the Sparta+ predicted chemical shift of Phe 27 C′ for snapshots saved every 4.5 ps of the MD trajectory. (b) The value of the ϕ dihedral angle of Phe 27 (ϕ_{27}) for snapshots saved every 4.5 ps of the MD trajectory. (c) The normalized distribution of the Sparta+ predicted shifts of Phe 27 C′ from the MD trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 1RIL) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (d) The conformation of Phe 27 observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding Sparta+ shift prediction of Phe 27 C′ and the ϕ value of Phe 27 (ϕ_{27}) for those conformations. Phe 27 C′ is colored cyan in the X-ray structure and magenta in the MD snapshots.

contribute to the shift predictions of His 30 HN reveals that, when the hydrogen bond is formed, the resulting downfield shift is offset by an upfield ring current shift of a similar magnitude, and the average chemical shift is predicted to be the same in the presence or absence of the hydrogen bond (data not shown). This illustrates how proximal spins can be sensitive to different structural features of the protein, and underscores the importance of deconvoluting the effects of the various structural features that contribute to each chemical shift prediction before interpreting their agreement with experimental values.

In general, we find that the more dramatic improvements in the accuracies of δMD values compared to δX-ray values are observed for ttRNH compared to ecRNH (Table 1) because a larger number of conformational features observed in the 2.8 Å resolution X-ray structure of ttRNH (PDB code 1RIL) appear to be inconsistent with solution chemical shifts when compared to the 1.5 Å resolution X-ray structure of ecRNH (PDB code 2RN2). Many of these inconsistent conformations are corrected to conformations in better agreement with solution chemical shifts within 100 ns of an MD trajectory, resulting in improved chemical shift predictions that are not solely due to the presence of conformational dynamics. We note that one could also obtain shift prediction improvements of this kind without considering shift predictions averaged over entire MD trajectories, but from considering individual structures obtained from X-ray structure refinement protocols that utilize explicit solvent and MD force fields to relax unfavorable X-ray conformations.[35]

A gradual decrease of δMD prediction rmsd's from experimental values over the course of a simulation might be interpreted as dynamical sampling converging on a given time

dx.doi.org/10.1021/ja300265w | J. Am. Chem. Soc. 2012, 134, 6365−6374

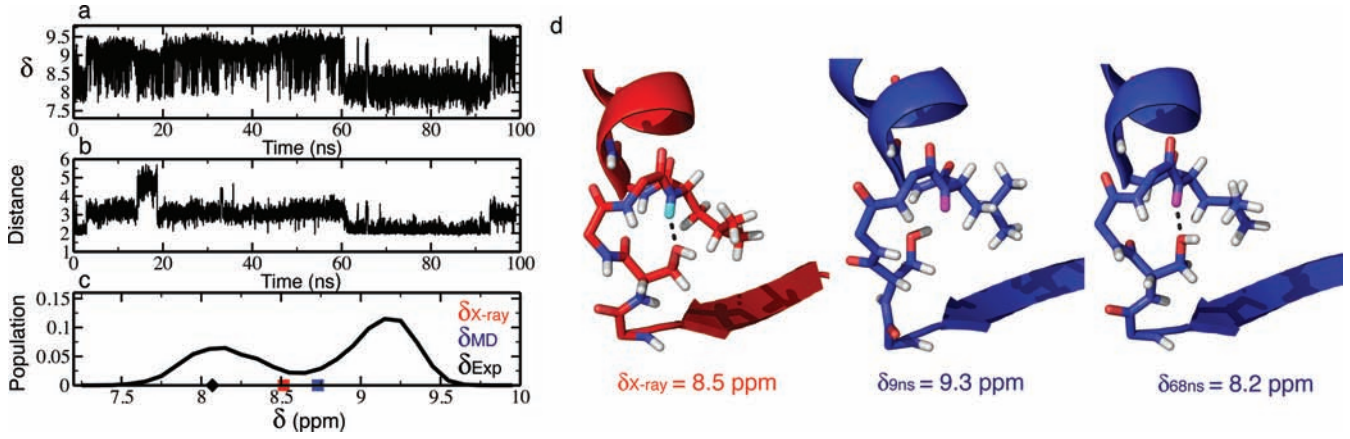


Figure 7. Chemical shift predictions of Leu 73 HN in a 100 ns MD simulation of ttRNH in the amber99SB force field. (a) The value of the Sparta+ predicted chemical shift of Leu 73 HN for snapshots saved every 4.5 ps of the MD trajectory. (b) The distance, in Å, between Leu 73 HN and the side chain hydroxyl O of Ser 70 for snapshots saved every 4.5 ps of the MD trajectory. (c) The normalized distribution of the Sparta+ predicted shifts of Leu 73 HN from the MD trajectory. The Sparta+ prediction obtained from the X-ray structure (pdb code 1RIL) is shown as a red square, the average value of the Sparta+ predictions over the entire MD trajectory is shown as a blue square, and the experimentally measured value is shown as a black diamond. (d) The conformation of Leu 73 HN observed in the X-ray structure (red) and from two representative MD snapshots (blue), along with the corresponding Sparta+ shift prediction of Leu 73 HN for those conformations. Leu 73 HN is colored cyan in the X-ray structure and magenta in the MD snapshots.

scale; however, such improvements could also be the result of several successive corrections of X-ray structure artifacts. We emphasize, however, that the rmsd of predictions averaged across all residues in a protein may not be the most informative metric, and note that even if the rmsd of δMD predictions is the same or higher than the rmsd of the X-ray predicted values for a given shift type, analysis of the per-residue differences in δX-ray and δMD may still reveal sites where conformational dynamics or conformational corrections to the X-ray structure provide a significantly improved agreement with experiment. This can be seen in the difference plot of HA predictions in ecRNH from the 100 ns Amber99SB MD simulation (Figure S2). The rmsd from experimental values for all HA atoms is 0.25 ppm for both δXray and δMD predictions, which is the reported accuracy of HA predictions of Sparta+ on static X-ray structures. However, multiple residues have values of $|\delta\text{MD} - \delta\text{Exp}|$ 0.4 ppm smaller or larger than $|\delta\text{Xray} - \delta\text{Exp}|$, suggesting the presence of conformational dynamics or significant conformational changes from the X-ray structure. This illustrates that the rmsd of δMD predictions, as considered at the macro atom type level across all residues, is not necessarily an indication that the accuracy of the dynamic modes and conformational changes observed in a trajectory are supported by experimental measurements. More detailed structural analysis of the features of a trajectory, such as those presented here, are required to determine if the improved predictions can be attributed to specific dynamic modes rather than conformational transitions away from X-ray structure conformations. In general, for a given structure of a protein, the extent to which chemical shift predictions will improve or worsen when calculated from a MD trajectory, and the extent to which those changes will result from differences in the average conformation compared to the presence of conformational dynamics, will vary on a case-by-case basis. These changes will be dependent on how well the structure matches the average solution conformation, the extent of the conformational dynamics present in solution, how accurately the conformational dynamics of the protein are captured by the MD simulation, and how sensitive the chosen chemical shift predictor is to those conformational dynamics.

Exposing Errors in Molecular Dynamics Simulations. In addition to identifying conformational dynamics that improve agreement with experimental values, an analysis of MD averaged chemical shifts can also identify erroneous conformations or motions in MD simulations that may result from force field deficiencies or insufficient sampling and can also suggest subsets of conformational space which are more consistent with experimental data. An example is presented in Figure 7. Here, the time course and distribution of the chemical shift predictions of ttRNH Leu 73 HN obtained from a 100 ns Amber99SB simulation is presented along with the time course of the distance between the side chain hydroxyl of Ser 70, which forms a helical N-Cap hydrogen bond with Leu 73 HN for portions of the trajectory. Examination of the time course and bimodal distribution of the Leu 73 HN shift makes clear that the shift is centered around a value of 8.2 ppm when the hydrogen bond is present, and a value of 9.2 ppm when the hydrogen bond is absent. The experimental value of 8.0 ppm suggests that the portion of the trajectory where the hydrogen bond is absent is not an accurate description of the dynamics in this region. This assertion is further supported by a comparison of the simulated ^{15}N generalized order parameter (S^2) of Leu 73 for both portions of the trajectory with the experimentally measured value.[27,36] In the portion of the trajectory where the N-cap hydrogen bond is formed (60−90 ns), a calculated S^2 value of 0.82 is obtained, in good agreement with the experimentally measured value of 0.84. In the portion of the trajectory where the hydrogen bond is not formed (20−60 ns), the calculated S^2 value is 0.60.

This example also illustrates the potential synergistic combination of NMR relaxation data, which provide information on amplitudes and correlation times of bond vector motions on the ps−ns time scale, and the analysis of dynamically averaged NMR chemical shifts. MD simulations and force field modifications are now routinely benchmarked based on agreements between experimental and simulated relaxation data.[29,36] Conversely, MD simulations are also

dx.doi.org/10.1021/ja300265w | J. Am. Chem. Soc. 2012, 134, 6365−6374

routinely employed to provide atomistic interpretations of relaxation data, for which the absolute accuracy of experimental measurements can be difficult to establish.[37] The additional consideration of dynamically averaged values of chemical shifts can provide an orthogonal source of validation of the correspondence between relaxation measurements and MD simulations. The analysis of averaged chemical shifts may enable one to discern between multiple atomistic models of protein motions that produce simulated relaxation data within the errors experimental measurements. Additionally, the different time scale sensitivities of relaxation measurements and chemical shifts may also provide temporal resolution of dynamic processes in proteins.

Comparing Molecular Dynamics Trajectories with Chemical Shift Predictions. Dynamically averaged chemical shifts also facilitate comparison of multiple MD trajectories of the same protein. Chemical shifts predicted for 100 ns simulations in the Amber99SB force field, 450 ns simulations using the Amber99SB-ILDN correction, and 1 μs simulations in the Amber99SB force field of ttRNH and ecRNH are summarized in SI Table 1 and SI Table 2, respectively. On average, the accuracy of the predictions across the simulations are similar, suggesting that, for the two force fields tested, significant differences are not observed on the time scales studied here. An exception is the improved N shift predictions in the 450 ns Amber99SB-ILDN and 1 μs Amber99SB trajectory compared to the 100 ns Amber99SB trajectory (SI Table 1), which appears to result from obtaining more accurate side chain rotamer population distributions. A comparison of the per-residue deviations of the values of $|\delta\text{Xray} - \delta\text{Exp}| - |\delta\text{MD} - \delta\text{Exp}|$ between trajectories (Figure S4), however, can illuminate specific differences among the trajectories. For example, the N-cap hydrogen bond between the side chain hydroxyl of Ser 70 and Leu 73 HN that breaks and reforms in the 100 ns Amber99SB simulation of ttRNH (Figure 7) remains stable in the 450 ns Amber99SB-ILDN and 1 μs Amber99SB simulation, resulting in more accurate δMD predictions. Comparing trajectories by differences in averaged chemical shift predictions may be a particularly useful option compared to other types of geometric or cluster based conformational analysis, as it quantifies differences in trajectories using only a set of conformational features that are known to influence the most readily accessible NMR observable, and may provide a good indication of the features of the trajectories that are most likely to be experimentally verifiable.

For simulations of ttRNH and ecRNH on time scales from 100 ns to 1 μs, the accuracy of δMD predictions generally do not improve with simulation length, and in some regions worsen. Our analyses of the trajectories suggest that this counterintuitive result may be due to the use of inflexible protonation states of side chain residues. The active site loop regions of ttRNH and ecRNH, ranging from residues 120−127, exhibit substantial conformational dynamics on the ps−ns and μs−ms time scales by NMR,[27,28] and contain a coupled network of His residues with predicted pK_a values ∼5.5, the experimental pH of the NMR measurements, suggesting constant pH simulations[38] may be necessary to accurately capture the dynamics of these regions.

Comparison of Chemical Shift Predictors. We have examined how the choice of chemical shift prediction software influences the interpretation of dynamically averaged chemical shifts calculated from MD trajectories. We used the chemical shift predictors Sparta+, Shiftx2, Shiftx+, Camshift, and ShiftS to calculate the average chemical shifts for 100 ns MD simulations of ecRNH and ttRNH in the amber99SB force field, and compared these values to chemical shifts calculated from the crystal structures using the same programs. The results of these calculations are displayed in SI Tables 3 and 4.

We found that the rmsd of MD averaged predictions from experimental values varied between the predictions in accordance with their previously reported prediction accuracies, with the most recently developed Sparta+ and Shiftx+/Shiftx2 yielding the lowest rmsd's. We observed that major trends and the most significant changes between δX-ray and δMD are largely conserved regardless of the choice of shift prediction software, suggesting that the major conformational dependencies of the chemical shift predictions are well captured by all the chemical shift predictors tested here. The differences between the shift predictors are generally manifested in small changes in the locations and widths of the predicted shift distributions.

A notable exception is the program ShiftS, which in addition to having substantially larger prediction rmsd's for values of δX-ray and δMD, also appears to have a less continuous chemical shift prediction surface than the other programs examined, resulting in noisier many-modal predicted shift distributions where other prediction tools generate smoother single and bimodal distributions (Figure S5). ShiftS was parametrized using a database of quantum mechanical chemical shift calculations performed on protein fragments, rather than using databases of high resolution X-ray structures and experimental chemical shifts. The ShiftS parametrization strategy could potentially enable a predictor to be more sensitive to smaller conformational fluctuations, as predictors that are trained to use the structural features of static X-ray structures to predict dynamically averaged experimental values may inherently contain some degree of dynamic averaging in their predictions. When a predictor is trained to predict an experimental chemical shift, which is in actuality an average value obtained from an ensemble of dynamic molecules, using the features of an individual static structure, the predictor may become less sensitive to the conformational fluctuations that are contributing to the experimentally measured average. That is, the predictor may become overly biased toward crystalline conformations and may erroneously attribute chemical shift changes that result from conformational dynamics in solution to the features of static crystal structures. However, inaccuracies in the quantum calculations used in the ShiftS parametrization, and the need to extrapolate relationships between the discrete conformations for which quantum calculations were performed, could also deleteriously affect the fits, and add additional noise to the predictions.

The results presented here suggest that the chemical shift predictors that were trained using high-resolution X-ray structures (Sparta+, Shiftx2, Shiftx+, and Camshift) are still sensitive to a range of conformational processes in proteins. The recently developed 4DSPOT chemical shift prediction model,[39,40] which trains its prediction models using averaged conformational properties observed in short MD simulations, rather than individual static structures, has shown some success in accounting for the effects of conformational dynamics during shift predictor parametrizations, and presents a promising approach for improving chemical shift predictors in the future.

The program Shiftx2 combines structure-based chemical shift predictions, provided by the module Shiftx+, with sequence

based predictions that are obtained from databases of experimental chemical shifts. This feature may reduce rmsd's of predictions if sequence homologues of a query protein are present in the Shiftx2 database, and in favorable cases may have the effect of shifting the center of δMD shift prediction distributions to be in better agreement with experiment. In general, however, comparisons of δX-ray and δMD values obtained solely from Shiftx+, using only structural inputs, will be more meaningful than those obtained from Shiftx2. A more detailed benchmark of the suitability of different chemical shift prediction tools for analyzing dynamics in MD simulations is needed to further explore these issues.

CONCLUSIONS

In this investigation, we have employed the most recently developed semiempirical NMR chemical shift prediction tools to evaluate the dynamically averaged values of chemical shifts obtained from unbiased molecular dynamics simulations of proteins. We have demonstrated that MD averaged chemical shift predictions generally improve agreement with experimental values when compared to predictions made from static X-ray structures and have conducted a detailed analysis of the structural dynamics and conformational changes associated with the improvements. We find that analyzing the time course and distributions of chemical shift predictions from molecular dynamics trajectories and examining the underlying dynamic modes that affect the chemical shift predictions can provide experimental support for atomistic descriptions of a number of motional processes in proteins.

Most previous structural interpretations of chemical shifts, both in terms of average structures and dynamics, have focused on the sensitivity of shifts to backbone ϕ and ψ angles.[20,21,25,26] We find that currently available tools for rapid backbone chemical shift predictions appear to be well suited to describe population weighted averages of multiple backbone conformational states and the sampling of smaller fluctuations within individual backbone conformational basins. We have also observed that predictions of backbone ^{1}H chemical shifts are sensitive reporters of fluctuations in aromatic ring currents and the presence and geometry of hydrogen bonds, and that backbone nitrogen chemical shift predictions are well suited to describe side chain χ_1 conformational dynamics.

We have also observed that MD averaged chemical shift predictions that provide an improved agreement with experiment do not necessarily reflect conformational dynamics and may result from one-time "corrections" in conformations observed in crystal structures that are unlikely to be populated in solution, such as those that arise from crystal packing artifacts. This suggests that an examination of prediction rmsd's from experimental values does not necessarily provide a validation of the dynamic modes observed in an MD trajectory, and underlies the importance of examining how a specific motional process of interest affects the agreement of chemical shift predictions with experimental values. This is likely of particular importance when using lower-resolution X-ray or NMR structures, structures obtained from homology models, and structures obtained from subsets of protein complexes as starting points of simulations.

In addition to examining structural dynamics in MD simulations that improved averaged chemical shift predictions, we have also demonstrated that poor predictions of averaged chemical shifts can identify spurious conformations and motions in MD simulations that may result from force field deficiencies or insufficient sampling and can also suggest subsets of conformational space that are more consistent with experimental data.

These results suggest that detailed analyses of dynamically averaged NMR chemical shifts from MD simulations will serve as a powerful tool for characterizing protein motions at an atomistic level. As NMR chemical shifts are the most readily accessible and ubiquitously measured NMR observable, and report on a large range of conformational properties in proteins, analyses such as those presented here should prove to provide one of the most accessible and informative links between simulations and experiments. Indeed, work has already begun to exploit some of the structural information contained in chemical shifts to aid in the refinement of molecular mechanics force fields.[41,42] As methods for predicting protein chemical shifts improve, expand to include more accurate predictions of side chain chemical shifts,[12,43,44] and are optimized to account for conformational dynamics during parametrizations,[39,40] atomistic descriptions of protein structural dynamics from NMR chemical shifts should become increasingly quantitative.

MATERIALS AND METHODS

Molecular Dynamics Simulations. Simulations were performed using Desmond Academic release 3 or source release 2.4.2.1.[45] The Amber99SB[29] or Amber99SB-ILDN[30] force fields were used with explicit TIP3P[31] water. All simulations used a 2.5 fs inner time step on a 1-1-3 RESPA cycle and were carried out in the NVT ensemble using a Nosé-Hoover thermostat after equilibration to constant box volume in the NPT ensemble. All simulations were run at 300 K. For each protein, *E. coli* ribonuclease H (ecRNH) and *T. thermophilus* ribonuclease H (ttRNH), three simulations were performed: a 100 ns simulation in the Amber99SB force field, a 1 μs simulation in the Amber99SB force field, and a 450 ns simulation in the Amber99SB-ILDN force field. Snapshots were saved every 4.5 ps. The 1RIL[32] and 2RN2[33] starting structures were used for ttRNH and ecRNH, respectively, and all structures were protonated in accordance with H++[46] pK_a predictions to replicate the pH of 5.5 used in previous NMR experiments on the *E. coli* and *T. thermophilus* proteins.[27,28,47,48] Crystallographic waters were removed and all structures were solvated with a minimum box buffer of 10 Å using Maestro version 8.5 or 9.1. Images were prepared in PyMol.

Chemical Shift Calculations. Prior to chemical shift calculations, waters were removed from each trajectory and each snapshot was saved as an individual PDB coordinate file. To enable the most robust parsing of each PDB file by the programs SPARTA+,[11] SHIFTX2,[12] SHIFTX+,[12] CamShift,[10] and ShiftS,[5−7] a brief 200 step steepest-descent minimization was applied to the crystal structures and to all MD snapshots. In the case of the crystal structures, some predictors require the addition of explicit hydrogens before a calculation can be performed, while others perform internal routines to add hydrogens. To enable the most direct comparisons between the prediction methods, hydrogens were added using the same external minimization routine instead of the internal routines, so each calculation was performed on identically prepared coordinates. Minimization was carried out in the Amber03[49] force field using the molecular simulation toolkit almost-1.0.4.[50] This program and force field were chosen as this combination has previously been shown to improve chemical shift predictions equally across multiple chemical shift prediction tools when calculating shifts from structures downloaded from the PDB.[10] The length of the minimization was empirically selected such that further minimization did not improve prediction rmsd's for crystal structures. The minimization caused very small fluctuations in the original positions of the atoms, on the order ~0.1 Å rmsd from the unminimized coordinates. The minimizations were also carried out for all MD snapshots in order to regularize the geometries of the hydrogen atoms between the MD snapshots and the crystal structures

and ensure that the differences in shift predictions between the crystal structures and the MD snapshots were not due to differences in the hydrogen atom geometries that resulted from the minimization routine. In the case of the MD snapshots, the average rmsd of the shift predictions fluctuated insignificantly after minimization (less than 1% on average). One important consideration was that, in addition to the minimization routine, a routine was also run to change atom and residue names to be recognizable to all predictors, which increased the number of residues which could be parsed by all of the shift predictors. All experimental shifts were rereferenced using the program Shiftcor.[51]

Abbreviations. NMR, nuclear magnetic resonance; MD, molecular dynamics; NOE, nuclear Overhauser effect; RDC, residual dipolar coupling; ecRNH, *Escherichia coli* ribonuclease H; ttRNH, *Thermus thermophilus* ribonuclease H; δMD, average chemical shift prediction obtained from a MD trajectory; δX-ray, chemical shift prediction obtained from an X-ray crystal structure; δExp, experimentally measured chemical shift; S^2, generalized order parameter.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S5; Tables S1−S4 as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

agp6@columbia.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was funded by an NSF postdoctoral research fellowship in Biology (grant 1002684) (P.R.), an NSF graduate research fellowship (K.A.S.), and NIH grant GM50291 (A.G.P.). We would like to thank Nikola Trbovic for providing the 100 ns molecular dynamics trajectory of ttRNH for analysis.

■ REFERENCES

(1) Pardi, A.; Wagner, G.; Wüthrich, K. *Eur. J. Biochem.* **1983**, *137*, 445−454.
(2) Osapay, K.; Case, D. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444.
(3) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311−433.
(4) Williamson, M. P.; Asakura, T. *J. Magn. Reson., Ser. B* **1993**, *101*, 63−71.
(5) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321−333.
(6) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408−423.
(7) Moon, S.; Case, D. A. *J. Biomol. NMR* **2007**, *38*, 139−150.
(8) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. J. *Biomol. NMR* **2003**, *26*, 215−240.
(9) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289−302.
(10) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894−13895.
(11) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13−22.
(12) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43−57.
(13) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615−9620.
(14) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685−4690.
(15) Wishart, D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. *Nucleic Acids Res.* **2008**, *36*, W496−502.
(16) Robustelli, P.; Cavalli, A.; Dobson, C. M.; Vendruscolo, M.; Salvatella, X. *J. Phys. Chem. B* **2009**, *113*, 7890−7896.
(17) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923−933.
(18) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970−14971.
(19) Berjanskii, M. V.; Wishart, D. S. *Nucleic Acids Res.* **2007**, *35*, W531−537.
(20) Cheung, M.-S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W. *J. Magn. Reson.* **2010**, *202*, 223−233.
(21) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795−2804.
(22) Hansen, D. F.; Kay, L. E. *J. Am. Chem. Soc.* **2011**, *133*, 8272−8281.
(23) Hansen, D. F.; Neudecker, P.; Kay, L. E. *J. Am. Chem. Soc.* **2010**, *132*, 7589−7391.
(24) Hansen, D. F.; Neudecker, P.; Vallurupalli, P.; Mulder, F. A. A.; Kay, L. E *J. Am. Chem. Soc.* **2010**, *132*, 42−43.
(25) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 1220−1221.
(26) Li, D.-W.; Brüschweiler, R. J. *Phys. Chem. Lett.* **2009**, *1*, 246−248.
(27) Butterwick, J. A.; Loria, J. P.; Astrof, N. S.; Kroenke, C. D.; Cole, R.; Rance, M.; Palmer, A. G. *J. Mol. Biol.* **2004**, *339*, 855−871.
(28) Butterwick, J. A.; Palmer, A. G. *Protein Sci.* **2006**, *15*, 2697−2707.
(29) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(30) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950−1958.
(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(32) Ishikawa, K.; Okumura, M.; Katayanagi, K.; Kimura, S.; Kanaya, S.; Nakamura, H.; Morikawa, K. *J. Mol. Biol.* **1993**, *230*, 529−542.
(33) Katayanagi, K.; Miyagawa, M.; Matsushima, M.; Ishikawa, M.; Kanaya, S.; Nakamura, H.; Ikehara, M.; Matsuzaki, T.; Morikawa, K. *J. Mol. Biol.* **1992**, *223*, 1029−1052.
(34) Christensen, A. S.; Sauer, S. P. A; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2078−2084.
(35) Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496−506.
(36) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G. *Proteins* **2008**, *71*, 684−694.
(37) Trbovic, N.; Cho, J.-H.; Abel, R.; Friesner, R. A.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2009**, *131*, 615−622.
(38) Khandogin, J.; Brooks, C. L. III. *Biophys. J.* **2005**, *89*, 141−157.
(39) Lehtivarjo, J.; Hassinen, T.; Korhonen, S.-P.; Peräkylä, M.; Laatikainen, R. *J. Biomol. NMR* **2009**, *45*, 413−426.
(40) Lehtivarjo, J.; Tuppurainen, K.; Hassinen, T.; Laatikainen, R.; Peräkylä, M. ; *J. Biomol. NMR* **2012**, *52*, 257−267.
(41) Li, D.-W.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(42) Li, D.-W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.
(43) Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 9620−9623.
(44) Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M. *J. Biomol. NMR* **2011**, *50*, 331−346.
(45) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossváry, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*; IEEE: New York, NY, 2006.
(46) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, W368−371.
(47) Yamazaki, T.; Yoshida, M.; Kanaya, S.; Nakamura, H.; Nagayama, K. *Biochemistry* **1991**, *30*, 6036−6047.
(48) Yamazaki, T.; Yoshida, M.; Nagayama, K. *Biochemistry* **1993**, *32*, 5656−5669.

(49) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(50) Almost-1.0.4 Home Page. www.open-almost.org.
(51) Zhang, H.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173−195.