A Simple Method to Measure Protein Side-Chain Mobility Using NMR Chemical Shifts

Mark V. Berjanskii and David S. Wishart*

Departments of Computing Science and Biological Sciences, University of Alberta, Edmonton, AB, Canada T6G 2E8

S *Supporting Information*

ABSTRACT: Protein side-chain motions are involved in many important biological processes including enzymatic catalysis, allosteric regulation, and the mediation of protein–protein, protein–DNA, protein–RNA, and protein–cofactor interactions. NMR spectroscopy has long been used to provide insights into the motions of side-chain groups. Currently, the method of choice for studying side-chain dynamics by NMR is the measurement of methyl ^{2}H autorelaxation. Methyl ^{2}H autorelaxation exhibits simple relaxation mechanisms and can be straightforwardly converted to meaningful dynamic parameters. However, methyl groups can only be found in 6 of 19 side-chain bearing amino acids. Consequently, only a sparse assessment of protein side-chain dynamics is possible. Therefore, there is a significant interest in developing novel methods of studying side-chain motions that can be applied to all types of side-chains. Here, we show how side-chain chemical shifts can be used to determine the magnitude of fast side-chain motions in proteins. The chemical shift method is applicable to all side-chain bearing residues and does not require any additional measurements beyond standard NMR experiments for backbone and side-chain assignments.

Side-chain dynamics play a crucial role in many aspects of protein function including enzymatic catalysis, allosteric regulation, and the mediation of protein–ligand interactions.[1] NMR spectroscopy provides structural biologists with a unique set of tools to assess the amplitudes and time-scales of side-chain motions at the atomic level. NMR studies of protein side-chain dynamics commonly include measurements of auto- or cross-correlated relaxation rates for various atoms found along the length of amino acid side-chains, such as methyl, methylene, methine, and aromatic groups (reviewed by Igumenova et al.).[2] These relaxation rates can be subsequently analyzed using model-independent methods (e.g., Lipari–Szabo model-free approach)[3] and/or model-dependent methods (e.g., Woessner's random jumps and stochastic diffusion models,[4] the wobbling-in-a-cone model,[5,6] and Gaussian axial fluctuation and jump model[7]) to extract the amplitude and time-scale of side-chain motions. While these experimental strategies have solid theoretical foundations and have provided many important insights into the role of side-chain motions in protein function, they also have their share of limitations and disadvantages. More specifically, as the protein size increases, the evaluation of methylene and methine groups quickly becomes very challenging due to increasing spectral overlap and/or diminishing signal intensity. The dynamic interpretation of side-chain atom ^{13}C relaxation is complicated by the presence of multiple relaxation mechanisms. Complex NMR experiments and expensive isotope labeling schemes are often required to suppress multiple ^{13}C relaxation pathways. Hence, the vast majority of NMR studies on side-chain dynamics are currently conducted using deuterated methyl groups via ^{2}H autorelaxation experiments. Methyl group ^{2}H autorelaxation is characterized by simple relaxation mechanisms, offers good spectral sensitivity, and can be straightforwardly converted to meaningful dynamic parameters. However, methyl groups can be found only in 6 of the 19 side-chain bearing amino acids, meaning only a sparse assessment of protein side-chain dynamics is possible. In addition, the NMR-measured dynamics of methyl groups is mostly affected by rotameric jumps[8] and thus may not always represent the collective motions of the whole side-chain. Finally, the time-scale of fast side-chain motions (ps–ns time-scale) that are amenable to study by NMR relaxation is limited by the protein's overall tumbling rate.[3] Given the aforementioned shortcomings, there is a significant need in the NMR community to develop novel ways of studying fast side-chain dynamics without the limitations of the existing relaxation methods.

Here, we present a simple method for extracting amplitudes of side-chain motions from NMR chemical shifts. The method is applicable to all 19 side-chain bearing amino acids, reports the mobility of the whole side-chain, and does not need any additional measurements beyond the standard NMR experiments for backbone and side-chain assignments. Further, we show that this new method provides a good experimental estimate of residue-specific accessible surface area and can be used to quantitatively predict B-factors of X-ray protein models and per residue root-mean-square deviations (RMSD) of side-chain atoms that are observed in MD simulations and NMR structural ensembles.

Previously we demonstrated that the amplitude of protein backbone motions could be determined using backbone chemical shifts via the random coil index (RCI).[9–11] Because of the RCI's widespread use in the biomolecular NMR community, we were encouraged to develop a similar approach to characterize side-chain dynamics. However, a simple extension of the existing RCI protocol to side-chain atoms does not produce a satisfactory assessment of side-chain motional amplitudes, as shown in Figure S1. In order to achieve

more accurate predictions, we had to fundamentally rethink how to define side-chain motions and how to use chemical shifts to assess their amplitudes.

In this work, we hypothesized that side-chain and backbone chemical shifts are related to a combination of all motional events that contribute to changes in the chemical environment of side-chain atoms. Such motions include interconversions among side-chain rotamers, fluctuations of backbone torsion angles (local and nearest neighbors), and large-amplitude segmental motions in protein loops and termini. We define the side-chain reorientations due to these combined motional events as total side-chain motions. Comparison of backbone RMSD ($RMSD_{BB}$) and side-chain RMSD ($RMSD_{SC}$) for total side-chain motions from MD simulations (Figure 1A) indicated that total side-chain motions can have differing dominant mechanisms in different protein regions.
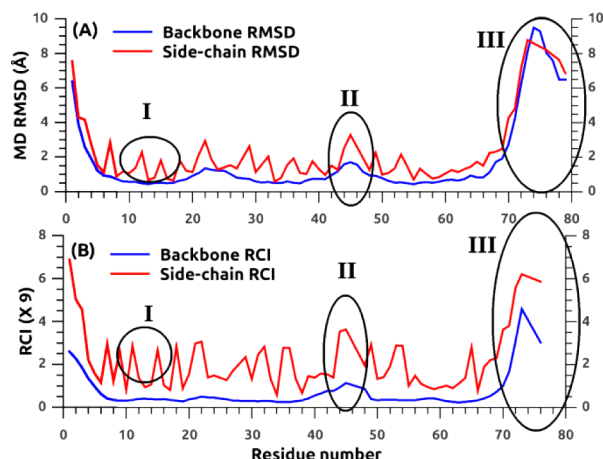

Figure 1. Mechanisms of total side-chain motions identified by MD (A) and side-chain RCI (B) for PyJ (PDB ID: 1FAF). Mechanisms are labeled with Roman numerals. Mechanism I corresponds to the dominant contribution of side-chain rotameric jumps to side-chain displacements. Mechanism II indicates comparable contributions of side-chain and backbone flexibilities to side-chain total motions. Mechanism III corresponds to the dominant contribution of backbone motions to the amplitude of side-chain movements. Ellipses indicate protein regions corresponding to particular mechanisms of total side-chain motions. Additional details about the simulations are in the SI.

For example, amplitudes of total side-chain motions in rigid parts (α-helix, β-sheet) of a protein will likely be dominated by transitions among side-chain rotamers (Mechanism I, $RMSD_{SC} \gg RMSD_{BB}$, Figure 1A), whereas side-chain displacements in very flexible loops and the termini will be primarily driven by backbone motions (Mechanism III, $RMSD_{SC} \approx RMSD_{BB}$, Figure 1A). Between these two extremes, multiple dynamic processes may have comparable contributions to total side-chain fluctuations (Mechanism II, Figure 1A). The dominant mechanism(s) of total side-chain motions will also depend on the residue type. While total mobility of residues with long side-chains (e.g., Arg, Lys, Leu, Ile, Met) will be greatly affected by transitions among side-chain group rotamers, total mobility of short (e.g., Ala, Ser) and rigid (Pro) side-chains will be significantly influenced by backbone motions. Regardless of the different dominant dynamic mechanisms, the principle outcomes in all scenarios are the same: side-chain reorientations change the chemical environment of side-chain atoms.

To calculate the amplitudes of total side-chain motions from experimental chemical shifts, we developed a new RCI specifically for side-chains (RCI_{SC}) that combines side-chain and backbone shifts as follows:

$$RCI_{SC} = A^*\langle \sum_a (k|\Delta\delta_a|)\rangle^{-1} + B^*RCI_{BB} \qquad (1)$$

where a indicates side-chain atoms of a given residue, $|\Delta\delta_a|$ is absolute secondary chemical shift of side-chain atom a (i.e., experimental shift minus the corresponding random coil value), k, A, and B are weighting coefficients, and RCI_{BB} is the RCI of the backbone chemical shifts.[9] Random coil values measured by Wishart et al.[12] were used in this work. ^{13}C, ^{15}N, and ^{1}H secondary chemical shifts were scaled by 2.5, 1, and 10, respectively, to account for the characteristic resonance frequencies of these nuclei. The minimal value for the scaled $|\Delta\delta_a|$ was set to 0.5 to avoid infinitely large RCI_{SC} values when secondary chemical shifts approach zero. For predicting accessible surface area (*vide infra*), the minimal value $|\Delta\delta_a|$ had to be increased to 1.5. The most recent weighting coefficients[10] and the standard procedure[11] for calculating the backbone RCI_{BB} were used.

We optimized the weighting coefficients in the RCI_{SC} expression to predict the amplitudes of side-chain motions (i.e., $RMSD_{SC}$) observed in MD simulations. To perform the optimization, we built training and testing sets, each containing 15 proteins with complete or near complete NMR assignments (Tables 1 and S1). The data sets consisted of 2858 residues,

Table 1. Spearman Coefficient of RCI_{SC} Correlation with Per Residue $RMSD_{SC}$ of MD and NMR Ensembles and Fractional ASA for the Testing Set

protein	BMRB ID	PDB ID	MD RMSD	NMR RMSD	ASA
At1g70830	7339	2I9Y	0.75	0.67	0.72
sperm flagellar protein 1	10147	2EE7	0.77	0.74	0.71
engrailed homeo-domain	7401	2P6J	0.89	0.87	0.77
F20O9.120	10090	1WJJ	0.81	0.79	0.79
HopPmaL	17739	2LF6	0.82	0.78	0.78
LIM domain	11350	1X4L	0.73	0.82	0.75
MM1357	6505	1YEZ	0.83	0.81	0.81
replication protein A	15849	2K5V	0.74	0.76	0.71
atrophin-1 interacting protein	10064	1UEW	0.78	0.84	0.73
hepatoma growth factor	10123	1N27	0.86	0.86	0.78
Alr3790	16382	2KL3	0.82	0.75	0.75
SpaI	17534	2LVL	0.85	0.75	0.72
ubiquitin	17769	1D3Z	0.77	0.64	0.71
STAM2	10264	1X5B	0.74	0.74	0.74
yxeF	15211	2JOZ	0.88	0.75	0.8

spanning a range of protein sizes from 32 to 149 residues and included various protein fold classes (all α, all β, mixed α/β) with both ordered and disordered regions.

To obtain the amplitudes of the total side-chain motions, we calculated MD trajectories for each protein using Gromacs 4.5.5.[13] The MD simulations were done in explicit solvent with the GROMOS96 43a1 force field[14] and had an average length of 3 ns. The $RMSD_{SC}$ was calculated for each MD trajectory using the Gromacs program. Details of the MD simulations and

RMSD calculations are available in the Supporting Information (SI).

The weighting coefficients (k, A, and B) in the RCI_{SC} expression were optimized by a grid search to maximize the Spearman rank-order correlation coefficient[15] between RCI_{SC} and the side-chain RMSD. Coefficients k were optimized separately for each residue type, whereas coefficients A and B were tuned globally for the whole training set. The average Spearman correlation coefficient for the training set was 0.81. The performance of RCI_{SC} for individual proteins in the training set is shown in Table S1. The optimized weighting coefficients can be found in Table S6. A leave-one-out cross-validation protocol was applied to the calculation of the RCI_{SC} coefficients to ensure that they were not overfitted. More specifically, weighting coefficients for each protein were optimized without its own data. The average rank-order correlation coefficient between the RCI_{SC} and $RMSD_{SC}$ from MD simulations in the leave-one-out test was 0.81 (identical to that obtained with the training set).

In addition, the RCI_{SC} method was validated on a testing set of 15 proteins, which were not included in the grid search (Tables 1 and S2). The average rank-order coefficient of correlation between RCI_{SC} and $RMSD_{SC}$ of these proteins was 0.80. The mean Pearson correlation coefficients[16] were 0.86 and 0.82 for the training and testing sets, respectively. Examples of per-residue correlations between RCI_{SC} and $RMSD_{SC}$ are shown in Figures 2A and S3.

It is important to note that the MD force-field and MD simulation period used to train RCI_{SC} were chosen to match the MD parameters that had been used in optimizing the original RCI_{BB}.[9] However, testing RCI_{SC} on subsets of its training/testing data indicated that RCI_{SC} still correlates very well with MD $RMSD_{SC}$ obtained with other force fields or with longer MD simulations (Tables S3 and S4).

We tested the generality of the RCI_{SC} estimates by calculating the correlation between RCI_{SC} and the magnitude of the side-chain total deviations that are observed in NMR ensembles (i.e., side-chain RMSD). Side-chain NMR RMSDs are influenced by protein dynamics and by the presence and strength of experimental restraints (e.g., NOEs). $RMSD_{SC}$ values in NMR ensembles were measured by MolMol[17] after aligning the ensembles by their secondary structure elements. Average Spearman coefficients of correlation between the RCI_{SC} and NMR $RMSD_{SC}$ are 0.77 for both the training and testing sets (Tables 1 and S1). Average Pearson correlation coefficients are 0.81 and 0.80 for the training and testing sets, respectively (Tables S1 and S2). Examples of the per-residue agreement between RCI_{SC} and NMR $RMSD_{SC}$ are shown in Figures 1B and S4.

In addition, we validated the RCI_{SC} method by testing its ability to predict thermal B-factors (B) of side-chains in X-ray protein models. B-factors are related to mean-square displacements $\langle x^2 \rangle$ of side-chain atoms via a well-known relationship:[18] $B = 8\pi^2 \langle x^2 \rangle$. For a set of 18 X-ray structures, RCI_{SC} showed a good correlation to mean side-chain B-factors, with a Spearman correlation coefficient of 0.72 and a Pearson correlation coefficient of 0.74 (Table S5, Figures 2C and S5).

We also investigated the relationship between RCI_{SC} and the fractional accessible surface area (ASAf) of protein residues. When inspecting side-chains of PyJ (PDB ID: 1FAF), we noticed that side-chains with low RCI_{SC} values (<0.11) were mostly buried, whereas residues with high RCI_{SC} (>0.15) values were solvent exposed (Figure S6). Calculations of coefficients
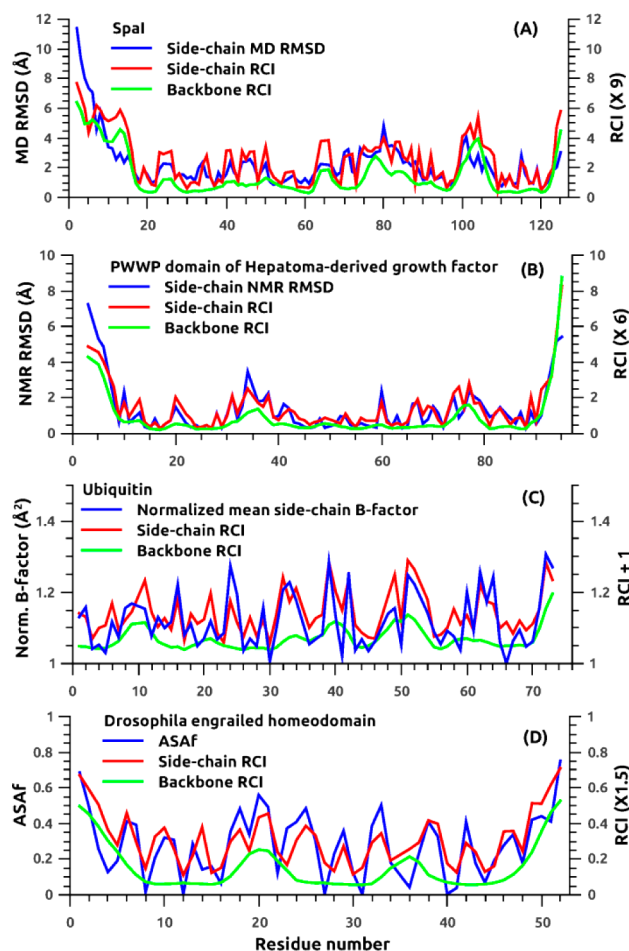

Figure 2. Correlation of the RCI_{SC} with side-chain MD RMSD (A), side-chain NMR RMSD (B), normalized mean side-chain B-factor (C), and fractional ASA (D). The backbone RCI (green lines) is also included to demonstrate that it alone cannot properly predict side-chain mobility and ASAf.

of RCI_{SC} correlation with ASAf (as reported by MolMol) confirmed that RCI_{SC} agrees with ASAf reasonably well, with mean Spearman coefficients of 0.76 and 0.75, and Pearson coefficients of 0.72 and 0.73 for the training and testing tests, respectively (Tables 1, S1, and S2). Examples of ASAf predictions can be seen in Figures 1D and S7.

Not surprisingly, our preliminary tests did not find a good correlation between RCI_{SC} and experimental CH_3 order parameters (data not shown). CH_3 order parameters are known to have a poor correlation with ASA[2,19−23] and primarily depend on transitions among side-chain rotamers.[8] Since the side-chain rotamer transitions and total side-chain motions (which RCI_{SC} was trained to predict) can correspond to different motional events, a disagreement of RCI_{SC} with CH_3 order parameters should not be unexpected. RCI optimization to predict CH_3 order parameters is beyond the scope of this paper and will be attempted in follow-up studies.

Using our training set, we obtained the following scaling relationships to quantitatively estimate side-chain RMSDs, ASAf, and B-factors from RCI_{SC} values:

$$\text{MD RMSD}_{SC} = RCI_{SC} \ast 9.0\text{Å} \qquad (2)$$

$$\text{NMR RMSD}_{SC} = RCI_{SC} \ast 6.0\text{Å} \qquad (3)$$

$$\text{ASAf} = \text{RCI}_{SC}*1.5 \tag{4}$$

$$B_{SC} = \text{RCI}_{SC}*1.0\text{Å}^2 + 1.0 \tag{5}$$

where B_{SC} is the average side-chain B-factor normalized by the smallest observed B-factor value (B_{MIN}), using equation $B_{SC} = (B - B_{MIN})/100 + 1$. Average median scores of absolute prediction errors for MD RMSD$_{SC}$, NMR RMSD$_{SC}$, and ASAf were, 0.40 Å, 0.37 Å, and 0.09, respectively, for the training set and 0.46 Å, 0.40 Å, and 0.09, respectively, for the testing set. The average median score for predicting the normalized B-factor was 0.07 Å^2. Prediction errors for individual types of residues are shown in Tables S7−S9 and can be displayed by the RCI$_{SC}$ program (*vide infra*) to indicate prediction uncertainty.

It is important to note that the side-chain RCI method allows one to identify the dominant mechanism of side-chain displacements. This is demonstrated for the PyJ protein (Figure 1B). A comparison of the RCI$_{SC}$ with its backbone RCI$_{BB}$ component can indicate what motional events contribute most into side-chain movements. When the RCI$_{SC}$ is significantly bigger than the RCI$_{BB}$ (Mechanism I, Figure 1B), it is reasonable to conclude that side-chain motions primarily depend on jumps among the side-chain rotamers. When the RCI$_{BB}$ is comparable to the RCI$_{SC}$ (Mechanism III, Figure 1B), backbone motions play the dominant role in side-chain displacements. For the remaining cases, one should expect comparable contributions of side-chain rotameric jumps and backbone flexibility to the amplitude of the total side-chain motions (Mechanism II, Figure 1B).

In summary, we have developed a new chemical shift-based method, called side-chain RCI, for predicting the amplitudes of total side-chain motions in proteins. Furthermore, we have demonstrated that this method can be used to obtain quantitative estimates of residue-specific side-chain RMSDs of MD and NMR ensembles as well as residue-specific ASAf and X-ray B-factors for a wide variety of proteins. Comparison of the side-chain and backbone RCI can also provide insights about the mechanism(s) of side-chain displacements, while the ASAf predicted from RCI$_{SC}$ should be useful for evaluating quality of protein models. The side-chain RCI approach has a number of advantages over many commonly used NMR relaxation methods. In particular, it reports the collective mobility of the complete side-chain (not just a specific side-chain group) for all 19 side-chain bearing amino acids. Side-chain RCI does not require any additional NMR measurements or the preparation of additional selectively labeled NMR samples beyond the NMR experiments and samples typically used for standard backbone and side-chain assignments.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Details of the MD simulations and RCI weighting coefficients and correlation of RCI$_{SC}$ with MD and NMR RMSD$_{SC}$, B-factors, and ASAf for several proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
david.wishart@ualberta.ca

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Teilum, K.; Olsen, J. G.; Kragelund, B. B. *Cell. Mol. Life. Sci.* **2009**, *66*, 2231.
(2) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672.
(3) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(4) Woessner, D. E. *J. Chem. Phys.* **1962**, *36*, 1.
(5) Kinosita, K., Jr.; Kawato, S.; Ikegami, A. *Biophys. J.* **1977**, *20*, 289.
(6) Lipari, G.; Szabo, A. *Biophys. J.* **1980**, *30*, 489.
(7) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 4272.
(8) Best, R. B.; Clarke, J.; Karplus, M. *J. Mol. Biol.* **2005**, *349*, 185.
(9) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970.
(10) Berjanskii, M. V.; Wishart, D. S. *J. Biomol. NMR* **2008**, *40*, 31.
(11) Berjanskii, M.; Wishart, D. S. *Nat. Protoc.* **2006**, *1*, 683.
(12) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67.
(13) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.
(14) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.
(15) Spearman, C. *Am. J. Psychol.* **1904**, *15*, 72.
(16) Pearson, K. *Philos. Trans. R. Soc. London* **1896**, *187*, 253.
(17) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51.
(18) Petsko, G. A.; Ringe, D. *Annu. Rev. Biophys. Bioeng.* **1984**, *13*, 331.
(19) Mittermaier, A.; Kay, L. E.; Forman-Kay, J. D. *J. Biomol. NMR* **1999**, *13*, 181.
(20) Mittermaier, A.; Davidson, A. R.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 9004.
(21) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C. H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T., II *Biochemistry* **1998**, *37*, 7965.
(22) Flynn, P. F.; Bieber Urbauer, R. J.; Zhang, H.; Lee, A. L.; Wand, A. J. *Biochemistry* **2001**, *40*, 6559.
(23) Liu, W.; Flynn, P. F.; Fuentes, E. J.; Kranz, J. K.; McCormick, M.; Wand, A. J. *Biochemistry* **2001**, *40*, 14744.