Journal of Biomolecular NMR (2005) 32: 151–162
DOI 10.1007/s10858-005-5366-0

Article

Relating side-chain mobility in proteins to rotameric transitions: Insights from molecular dynamics simulations and NMR

Hao Hu[a], Jan Hermans[a] & Andrew L. Lee[a,b,*]
[a]*Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA;* [b]*Division of Medicinal Chemistry and Natural Products, School of Pharmacy, A.L.L., University of North Carolina, 310 Beard Hall, CB# 7360 Chapel Hill, NC, 27599-7360*

Received 15 February 2005; Accepted 28 March 2005

Key words: S^2_{axis}, molecular dynamics simulation, NMR, order parameter, rotamer populations, side-chain dynamics

Abstract

The dynamic aspect of proteins is fundamental to understanding protein stability and function. One of the goals of NMR studies of side-chain dynamics in proteins is to relate spin relaxation rates to discrete conformational states and the timescales of interconversion between those states. Reported here is a physical analysis of side-chain dynamics that occur on a timescale commensurate with monitoring by ^{2}H spin relaxation within methyl groups. Motivated by observations made from tens-of-nanoseconds long MD simulations on the small protein eglin c in explicit solvent, we propose a simple molecular mechanics-based model for the motions of side-chain methyl groups. By using a Boltzmann distribution within rotamers, and by considering the transitions between different rotamer states, the model semi-quantitatively correlates the population of rotamer states with 'model-free' order parameters typically fitted from NMR relaxation experiments. Two easy-to-use, analytical expressions are given for converting S^2_{axis} values (order parameter for C–CH_3 bond) into side-chain rotamer populations. These predict that S^2_{axis} values below 0.8 result from population of more than one rotameric state. The relations are shown to predict rotameric sampling with reasonable accuracy on the ps–ns timescale for eglin c and are validated for longer timescales on ubiquitin, for which side-chain residual dipolar coupling (RDC) data have been collected.

Abbreviations: p_{major} – fractional population of the major rotameric state of a side chain; S^2 – Lipari-Szabo 'model-free' order parameter; S^2_{axis} – 'model-free' order parameter for the methyl symmetry axis; MD – molecular dynamics; MM – molecular mechanics; NMR – nuclear magnetic resonance; RDC – residual dipolar coupling.

Introduction

While structural coordinates of proteins are being deposited into the Protein Data Bank (PDB) in unprecedented numbers and at ever higher resolution (Westbrook et al., 2003), the dynamic aspect of protein structure, while appreciated by many, has proven to be more elusive in its characterization. Studies of protein dynamics are motivated by both the knowledge that considerable residual motion exists in the native state and that protein flexibility represents an important component of function. A wide range of theoretical and experimental methods have been applied to characterize and understand these motions, from small-amplitude, fast-timescale bond librations to large domain motions that occur on much

*To whom correspondence should be addressed. E-mail: drewlee@unc.edu

longer timescales. While it may be argued that a gap exists between theoretical and experimental work, this gap appears to be shrinking (Case, 2002; Lee et al., 2002; Brüschweiler, 2003; Best et al., 2004). One of the limitations of experimental characterizations of dynamics is that while accurate characterizations are made, the resultant picture is often 'blurred.' An example of this is the widely used Lipari-Szabo order parameter, S^2, which describes an individual bond vector's 'rigidity' in the frame of the macromolecule, on a scale of 0–1 (Lipari and Szabo, 1982a). Such order parameters are most frequently used in NMR studies of internal dynamics on the ps–ns timescale (Palmer, 2001). These studies typically employ auto-correlated spin relaxation for determining order parameters. (Analysis of cross-correlated relaxation can add to the level of detail of the motions, particularly along the backbone (Fischer et al., 1997).) Although S^2 values have the flexibility to be projected onto different motional models, this is not usually done, and the order parameters are typically evaluated on their own terms. While this is sufficient for many applications, the 'picture' of side-chain methyl dynamics provided by NMR-derived S^2_{axis} values (Muhandiram et al., 1995; Wand, 2001) becomes noticeably blurred, as the order parameter may be lowered by sampling of multiple rotamer states in addition to diffusive motions within individual rotamer states. Thus, it often remains unclear how the S^2_{axis} values attain their low values. It is therefore desirable to extract a more detailed view of side-chain dynamics from experimentally determined S^2_{axis} values.

Here, we have combined molecular dynamics simulations and NMR relaxation studies on a small protein to gain further insights into side-chain dynamics. The protein used for these studies is eglin c, a small (8 kDa) serine protease inhibitor from the leech *Hirudo medicinalis*. A 40 ns MD simulation performed previously to assist interpretation of the dynamic response to pH (Hu et al., 2003) has been extended to 80 ns and is used here to provide insight into ps–ns motions of methyl-bearing side chains. Specifically, analysis of side-chain dynamics from this simulation of eglin c has provided a striking realization of the degree to which multiple rotamers are sampled and contribute to S^2_{axis} . This picture is supported by the qualitative agreement between experimental and simulated dynamics of hydrophobic side chains.

In an attempt to build rotamer sampling directly into the interpretation of experimental S^2_{axis} values in the absence of MD simulations, we use a simple molecular-mechanics (MM) based model for side-chain rotations. Numerous physical models have been developed previously for the interpretation of model-free order parameters, such as diffusion in a cone (Lipari and Szabo, 1982b), jump models (Wittebort and Szabo, 1978), the Gaussian axial fluctuation (GAF) model (Brüschweiler and Wright, 1994), and the 'GAF-and-jump' model (Bremi et al., 1997). Although these have been successful towards their application to backbone directed dynamics, only the GAF-and-jump model is applicable to side-chain rotameric sampling, with the model requiring up to five internal parameters. We show that the MM-based model allows S^2_{axis} values to be projected into rotamer populations by a simple relation, independent of the timescale of interconversion. In addition, we show that this approximation works well in reverse by calculating S^2_{axis} values from rotamer populations based on ubiquitin RDC data and showing that these are in good agreement with S^2_{axis} values determined from explicit fitting of vector distributions to the RDC data. These results, along with the previously shown correlation of RDC-derived S^2_{axis} values with those derived from ^{2}H relaxation (Chou et al., 2003), imply that a large portion of rotamer interconversions occur on the ps–ns timescale, at least in ubiquitin. Recently, a method was reported to incorporate S^2_{axis} values into refinement of NMR ensembles, resulting in observation of side-chain rotameric sampling (Lindorff-Larson et al., 2005). The approach used here does not involve calculation of structures and is therefore a complementary, simpler method. The convenient ability to approximate rotamer interconversion populations will facilitate testing whether side chains in other proteins also interconvert primarily on the ps–ns timescale. Having a simple method for estimating the extent of these structural changes should help to sharpen our view of dynamic processes in proteins.

Methods

The backbone and side-chain dynamics of eglin c at pH 7 have been measured previously by ^{15}N and ^{2}H spin relaxation methods, respectively (Hu

et al., 2003). MD simulations were carried out for eglin c at pH 7 using the CHARMM22 force field (MacKerell et al., 1998) and the Sigma program (Mann et al., 2002). The details of the simulation have been described previously (Hu et al., 2003), although here the length of the simulation has been extended to 80 ns.

Angular distributions of backbone N–H bonds and side-chain terminal C–C bonds were generated by first extracting the coordinates of the whole protein molecule from simulated trajectories, then removing the global translation and rotation by superimposing the core of the structure (residues 8–38, 50–70) with respect to the initial structure, and normalizing the bond vectors. By considering each vector to start from a common origin and end on the surface of a sphere of unit radius, one obtains a straightforward visualization of the distribution of bond vectors and identification of different conformational states.

S^2_{axis} values were calculated for all terminal C–CH_3 bond vectors from multiple 2.88 ns windows from the MD simulation and averaged, as was done previously (Hu et al., 2003). For each MD window, S^2_{axis} values were calculated as S^2 for C–C(H_3) bond vectors using the following expression (Henry and Szabo, 1985):

$$S^2 = 3/2\left[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle xz\rangle^2 + 2\langle yz\rangle^2\right] - 1/2 \quad (1)$$

As the experimental correlation time for global rotation, τ_m, is 4.6 ns, it is inappropriate to use the entire 80 ns simulation for calculating autocorrelation functions to be compared to NMR relaxation data, which is insensitive to motions slower than τ_m. Splitting the long trajectory into multiple windows has the advantage of enabling estimation of the statistical errors. Varying the sampling window size from 2.88 to 5.4 ns essentially does not change the results (data not shown).

Results and discussion

MD simulation as a model for protein dynamics

In the present study, we focus on using MD simulations to aid in the interpretation of order parameters (S^2_{axis}) of the 3-fold symmetry axes of side-chain methyl groups (Lipari and Szabo, 1982b; Nicholson et al., 1992), which are known to exhibit both high and low values, often despite the degree of interior burial in the structure. MD simulations were carried out previously on the 70-residue protein eglin c (Hu et al., 2003). Important to this study is evidence that the simulation reproduces the main features of the ps–ns dynamics as determined from the NMR-based characterization (Peng and Wagner, 1992; Hu et al., 2003). The simulated backbone (S^2) and side-chain methyl (S^2_{axis}) order parameters of eglin c were calculated and compared to the experimentally determined values (Hu et al., 2003) (Figure 1). On a qualitative level, the simulation and experimental results agree well with each other, though the agreement is better for the

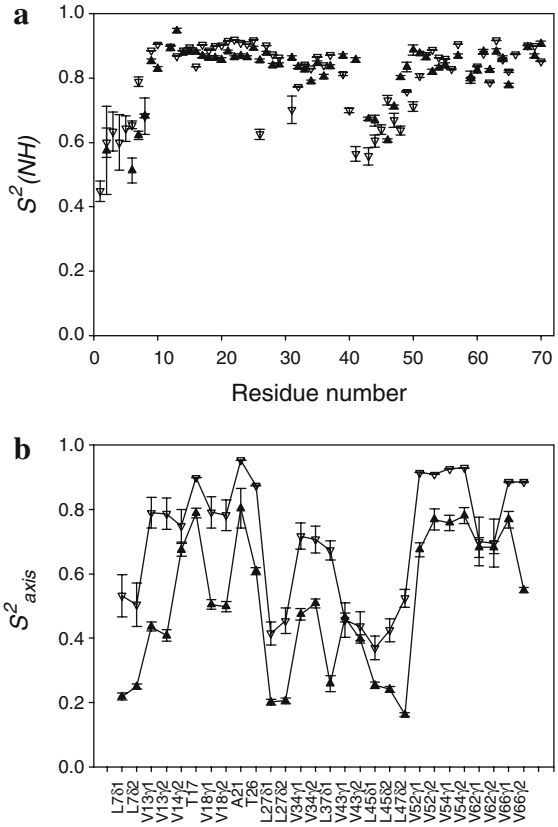

Figure 1. Comparison of eglin c order parameters determined by NMR and by MD simulation. (a) S^2 values for backbone NH bonds; (b) S^2_{axis} values for side-chain methyl group symmetry axes. Open symbols, data obtained from MD simulation; filled symbols, data obtained from NMR relaxation measurements.

backbone. Of particular note is the clear correlation of side-chain order parameters from MD and NMR (the correlation coefficient, r, between experimental and simulated side-chain order parameters is 0.84), as these have typically proven to be more difficult to match than with backbone S^2 values. Eglin c therefore appears to be a favorable case for such comparisons. In both the simulation and experiment, the order parameters of backbone sites are generally higher than those of side chains. This phenomenon has been observed in many other proteins, and is usually attributed to the greater torsional freedom of side-chain groups which leads to more complicated motions. In summary, the MD simulations have reproduced the amplitudes of dynamics of methyl-bearing side chains to a degree that merits further investigation of the MD simulations.

Side-chain rotamer sampling

The angular distributions of specific bond vectors from the simulation were computed and selected examples are shown in Figure 2 (top level, N–H vector with S^2 of 0.56; lower level, C–CH_3 vectors with S^2_{axis} of 0.91, 0.62, and 0.16, from left to right). By comparing the distributions of bond vectors with the computed order parameters (Figure 2), one sees that the value of S^2_{axis} can reflect the presence of multiple rotameric states, as pointed out previously (Nicholson et al., 1992; Yang et al., 1998; Lindorff-Larson et al., 2005). Here, because the simulated dynamics are in agreement with NMR-derived order parameters and span the same motional timescale, a direct connection can be made between S^2_{axis} and population of multiple rotameric states. From inspection of multiple side-chain methyl groups, we observe that (1) a single rotamer state is populated when the S^2_{axis} value is high; (2) multiple rotamer states show up when S^2_{axis} is below a certain value; and (3) there is a correlation between the number of rotameric states sampled and decreasing value of S^2_{axis}. Such a correlation is not observed for the backbone NH vectors, which implies a qualitative difference between the backbone and side-chain motions. Backbone groups are typically more rigid and are usually undergoing quasi-harmonic motions, unless a large conformational

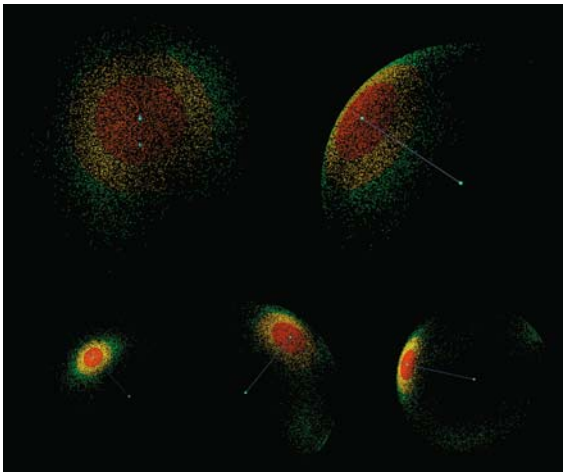

Figure 2. Representative spatial distributions of backbone NH and side-chain C–CH_3 bonds observed in MD simulations. Upper portion of panel: the distribution of an NH bond with S^2 of 0.56 shown in two different views. Lower portion of panel: left, a side-chain C–CH_3 bond with S^2_{axis} of 0.91; middle, a side-chain C–CH_3 bond with S^2_{axis} of 0.62; right, a side-chain C–CH_3 bond with S^2_{axis} of 0.16. The blue line is the mean vector of all bond vectors. Orange points, 40% of all points that are closest to the mean axis; yellow points, next 40%; green points, last 20%. Although the three side-chain vectors appear variable in size, this effect is due to different orientations, and all three distributions are scaled equally.

transition occurs, which is rare, especially for the stable, compact core of the protein. Local motions of terminal side-chain methyl groups due to bond vibration and bond angle bending are significant; however, motions due to changes of dihedral angles are much less restricted in the side chains. These rotational motions are modulated by non-bonded interactions with neighboring residues, which provides the principal source of differences between order parameters of different residues. As shown in Figure 3, the internal rotational motions about the preceding C–C bond (C–N bond for Ala) will have the largest influence on the distribution of side-chain C–CH_3 bond vectors.

Molecular mechanics model for methyl dynamics

To gain insight into the determinants of side-chain methyl dynamics, we focus on the factors that most immediately affect the motion of the C–CH_3 bond. We start by exploring the motions of a side-chain C–CH_3 bond in terms of energetics. The most impor-

tant and direct forces influencing its motion are the bending of the bond angle C′–C–CH_3, and the rotation of the dihedral C″–C′–C–CH_3. These two angles will be referred to as θ and χ, respectively. Without considering the interactions with the rest of the protein and the correlated motion requiring torsion of more than one dihedral, as between the side chain and the backbone, the potential surface for this bond vector is simply determined by

$$E_{\mathrm{C-CH_3}}(\theta,\chi) = \frac{K_\theta}{2}(\theta-\theta_0)^2 + \frac{V_\chi}{2}[1+\cos(n\chi-\chi_0)] \qquad (2)$$

in which the first term is a harmonic angular bending term, and the second is the energy of an internal rotation with n minima. The functional form of this equation is typical of MM force fields. The spatial distribution of the bond vector can be easily computed according to a Boltzmann distribution, i.e.,

$$\rho(\theta,\chi) = \frac{1}{Z}\exp[-E_{\mathrm{C-CH_3}}(\theta,\chi)/k_BT] \qquad (3)$$

in which k_B is the Boltzmann constant, T the temperature, and Z the normalization factor or partition function. Therefore, given the parameters of the potential function, K_θ, θ_0, V_χ, χ_0, and n, one can compute the order parameter from

$$\begin{aligned} S^2 &= \frac{3}{2}\left[\langle x^2\rangle^2+\langle y^2\rangle^2+\langle z^2\rangle^2+2\langle xy\rangle^2 +2\langle xz\rangle^2+2\langle yz\rangle^2\right]-\frac{1}{2} \\ &= \frac{3}{2}\left[\begin{array}{l} \left(\int \rho(\theta,\chi)x^2_{\theta,\chi}d\theta d\chi\right)^2 \\ +\left(\int \rho(\theta,\chi)y^2_{\theta,\chi}d\theta d\chi\right)^2 \\ +\left(\int \rho(\theta,\chi)z^2_{\theta,\chi}d\theta d\chi\right)^2 \\ +2\left(\int \rho(\theta,\chi)x_{\theta,\chi}y_{\theta,\chi}d\theta d\chi\right)^2 \\ +2\left(\int \rho(\theta,\chi)x_{\theta,\chi}z_{\theta,\chi}d\theta d\chi\right)^2 \\ +2\left(\int \rho(\theta,\chi)y_{\theta,\chi}z_{\theta,\chi}d\theta d\chi\right)^2 \end{array}\right]-\frac{1}{2} \end{aligned} \qquad (4)$$

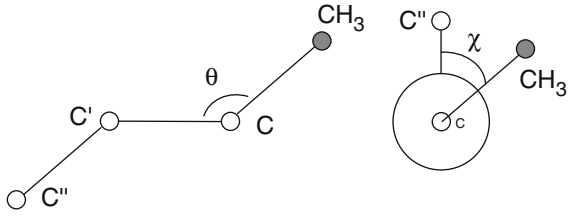

Figure 3. The two local interactions that dominate the motion of side-chain methyl groups. The filled sphere is the carbon atom of the CH_3 group; open spheres are other heavy atoms such as carbon or nitrogen. See text for other notations.

This approach is well-suited for analysis of S^2_{axis} parameters as derived from ^{2}H relaxation (Muhandiram et al., 1995), as bond vibrations are factored out by being subsumed into the quadrupolar coupling constant used in the model-free analysis. The model parameters, i.e., the angular bending force and the rotation barrier, used in Equation (2) are very similar for methyl groups (with the exception of methionine methyls) and vary little between vacuum and aqueous states. Indeed, different molecular mechanics force fields [CHARMM (MacKerell et al., 1998), AMBER (Cornell et al., 1995), GROMOS (van Gunsteren et al., 1996), and OPLS-AA (Jorgensen, 1998)] tend to have similar values of K_θ and V_χ for the side-chain methyl groups of Val, Ile, Leu, and Thr. Here we use (typical) values for K_θ of 110 kcal/(mol rad^2), and for V_χ of 3.0 kcal/mol, with a value for n of 3. Results of a high level *ab initio* quantum calculation (6-311++g**/B3LYP) for leucine dipeptide in vacuo (results not shown) agree well with these. The force constant for deforming the bond angle C^β–C^γ–C^δ is so large that even in the interior of a packed protein, the distribution of the bond angle is not expected to differ significantly between residues in the core and on the surface of the protein. However, the rotational energy profile, and in particular, the energies of different rotamer states will in most cases be modified by interactions with other groups, especially if a side chain is part of a tightly packed protein core, and even a small energy difference will strongly affect the relative distribution over the different rotamer states.

As an approximation, we assume that the side-chain C–CH_3 bond vector follows a distribution that consists of one or several Boltzmann

distributions of rotamer states, each for an energy given by Equation (2). However, the rotamer states are unequally populated. Transitions between different rotamer states will be rare and transient, and the population of conformations at the high-energy barriers will make a negligible contribution to the order parameter. Analysis of the dynamics of the six leucine side chains of eglin c shows that during the course of the 80 ns simulation, all are found in at least six of nine possible rotameric states. In each state, the distributions of χ_1 and χ_2 oscillate around a specific set of mean values (Butterfoss et al., 2005), with root-mean-square deviations of χ_2 (the angle that corresponds to χ of Equation (2)) that vary in a range of 10–14 degrees.

The relative population of each rotamer state then becomes the dominant factor determining the value of the order parameter and one can rewrite Equation (3) as

$$\begin{aligned}
&\rho(\theta,\chi|_{(n\chi-\chi_0)\in[-\pi,\pi]}) \\
&\quad = \frac{\lambda_1}{Z}\exp[-E_{\mathrm{C-CH_3}}(\theta,\chi)/k_BT] \\
&\rho(\theta,\chi|_{(n\chi-\chi_0)\in[\pi,3\pi]}) \\
&\quad = \frac{\lambda_2}{Z}\exp[-E_{\mathrm{C-CH_3}}(\theta,\chi)/k_BT] \\
&\rho(\theta,\chi|_{(n\chi-\chi_0)\in[-3\pi,-\pi]}) \\
&\quad = \frac{\lambda_3}{Z}\exp[-E_{\mathrm{C-CH_3}}(\theta,\chi)/k_BT] \\
&\lambda_1+\lambda_2+\lambda_3 = 1
\end{aligned} \tag{5}$$

in which λ_1, λ_2, and λ_3 are the population fractions of each rotamer state, the distributions of θ and χ are identical for rotamer states, but scaled according to the overall populations between rotamer states.

Equations (4) and (5) relate order parameters to the distribution over rotamers. To demonstrate the results, three different extreme distribution schemes of the rotamer populations are considered. In Case 1, the C–CH_3 bond is trapped in a single rotamer. In Case 2, the bond has an equilibrium distribution between two rotamer states, i.e., λ_3 is zero. In Case 3, the bond has an equilibrium distribution between three rotamer states.

In Case 1, computation with the assumed values of the potential function results in a lower bound to the order parameter, as non-bonded interactions between this bond and other groups would serve to increase the value of V_χ and hence narrow the width of the distribution of χ. When using V_χ of 3.0 kcal/mol, this lower bound for S^2_{axis} is 0.84. S^2_{axis} values below this value are predicted to sample additional rotamers. In Case 2, by allowing equilibration between two rotamer states, the variation of order parameter vs. the population of the major rotamer state can be computed (Figure 4, solid line). In this case, the upper bound for S^2_{axis} is 0.84 and the lower bound corresponds to an equal distribution (0.5:0.5) between two states, with S^2_{axis} of 0.29. Case 3 can be broken down into two limiting extremes, as the order parameter is now affected by the populations of two additional rotamer states which can change independently. If we focus on the population of one major rotamer state, then the order parameter has an upper bound corresponding to only one minor rotamer state being populated (which turns to be Case 2), and a lower bound corresponding to the two minor states being equally populated. Given the population of the major rotamer, the order parameter can vary between the upper bound (Case 2) and the lower bound. The relationship between order parameters and the population of the major rotamer states corresponding to the limit of lower bound is also shown in Figure 4 (dashed line). Using this simple molecular mechanics-based model, the minimum value of the order parameter is 0.11, corresponding to an equal population of three rotamer states. It is important to point out that interpreting S^2_{axis} values in this manner provides an underestimation of the actual equilibrium population distribution of rotamers; the extent of this underestimation is determined by the interconversion timescale to which the measurement (for obtaining S^2_{axis}) is sensitive. In addition, the S^2_{axis} rotamer population correlation is independent of the identity of specific rotameric configurations.

Estimation of rotamer populations sampled on sub-τ_m timescales from relaxation-derived S^2_{axis} values

Figure 4 shows the predicted order parameters (S^2, or S^2_{axis}) vs. the population of (major) rotamer states. If only two rotamer states are sampled on a

sufficiently fast timescale, the relation between the population of the major rotamer, p_{major} and the order parameter, S^2_{axis} is simply

$$p_{\text{major}} = \sqrt{\frac{S^2 - 0.29}{2.19}} + 0.5 \qquad (6)$$

the values of the constants having been estimated by numerical integration of Equation (4). For order parameters higher than 0.4, this equation can be used more generally, as the differences between Cases 2 and 3 are small (Figure 4, see below). For example, an S^2_{axis} value of 0.55 is expected to reflect a major rotamer populated at the 84% level with the remaining 16% distributed over one or both of the minor rotamers. It is also possible to compute the 'lower bound' relationship for Case 3, in which the two minor states are equally populated as

$$p_{\text{major}} = \sqrt{\frac{S^2 - 0.11}{1.65}} + 0.33 \qquad (7)$$

Therefore, based on the molecular mechanics model proposed above, one can use these simple expressions to convert side-chain order parameters into populations of interconverting rotamer states that are sampled on the relevant timescale.

To assess the ability of these expressions to correctly predict the degree of rotamer averaging, we have compared p_{major} for side-chain terminal

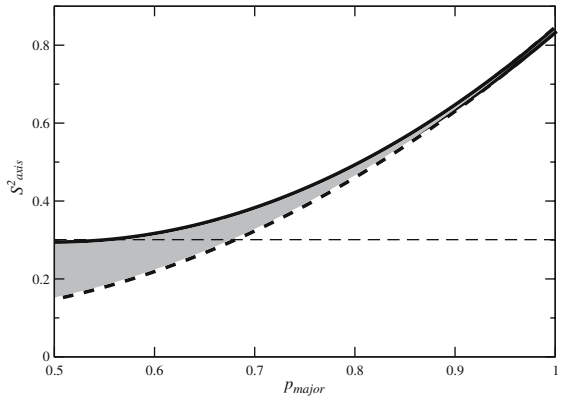

Figure 4. The relationship between populations of rotamer states and predicted order parameters. The *x*-axis is the fraction of the most highly populated rotamer state. Solid line: model for two rotamer states (Equation (5)); dashed line: model for three rotamer states with the two minor states equally populated (Equation (6)). The dashed line can be considered a general lower limit for the major rotamer when there are transitions between all three rotamers. The hatched area corresponds to the general case of transitions between all three rotamers.

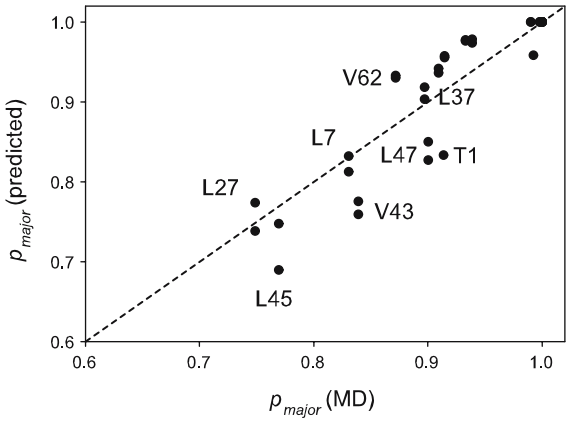

Figure 5. Correlation between the population of the major rotamer state (p_{major}) predicted from Equation (6) (after calculation of S^2_{axis} values, see text) and the actual values observed directly from the MD simulations. To have consistency with predicted p_{major} values, the actual p_{major} values taken directly from the simulations are mean values calculated from p_{major} values determined for 2.88 ns time windows. Val, Ile, Leu, and Thr p_{major} values are shown.

rotamers in eglin c, as calculated from Equation (6), to p_{major} values observed directly from the MD simulation (Figure 5, $r = 0.93$). The calculated p_{major} values (y-axis) are based on the S^2_{axis} values determined directly from the simulation (as shown in Figure 1), and are mean values based on averaging of multiple 2.88 ns windows in order to have self-consistency with S^2_{axis} values. The correlation shown in this plot shows that Equations (6) and (7) are reasonably accurate at predicting p_{major} and should, therefore, be useful for assessing experimental S^2_{axis} parameters. In general, the correlation reinforces the MM-based potential function as a reliable, semi-quantitative method that can be used to extract approximate S^2_{axis} and rotamer population information of side chains in proteins.

To further test the MM-based model, S^2_{axis} parameters were calculated directly from the MD simulations for C–CH_3 vector distributions within single rotamers, thereby simulating side chains with no rotamer averaging. Consistent with the predictions of the model, methyl order parameters from threonine and valine residues were clustered about a value of 0.9 (Figure 6a). Although the majority of these 'single-rotamer' S^2_{axis} values were greater than 0.84, a small group were found to have values that lie between 0.74 and 0.84. Further analysis of the MD simulation showed that these low S^2_{axis} values result from additional motion of

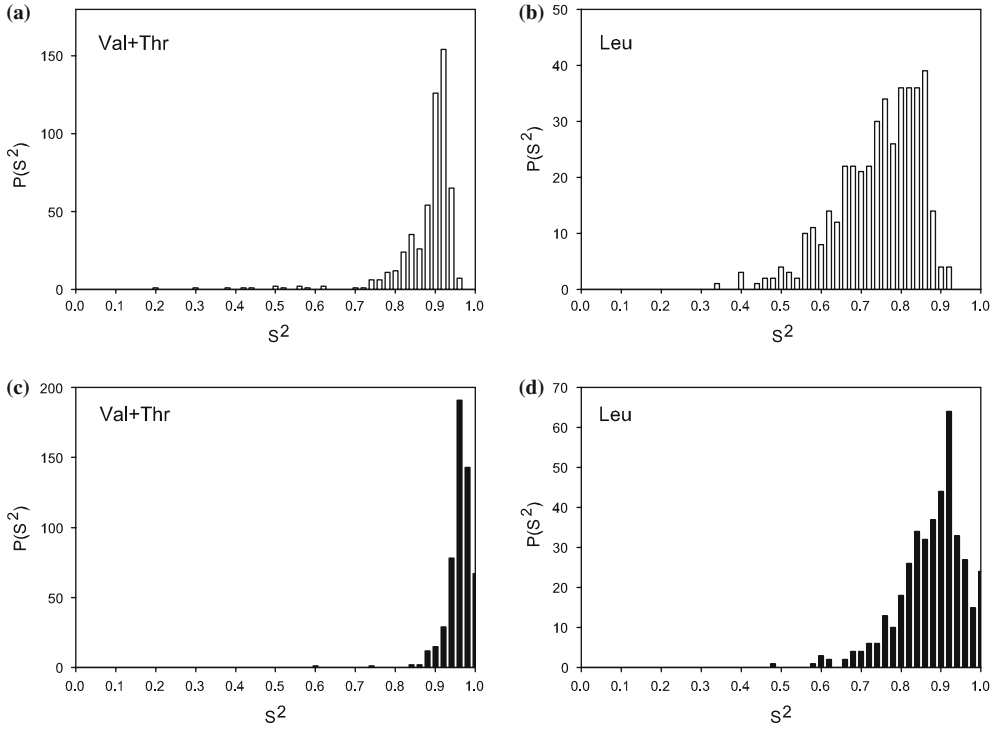

Figure 6. Distributions of 'single-rotamer' order parameters (panels a,b). Single-rotamer order parameters were corrected (panels c,d) by dividing S^2_{axis} by the order parameter corresponding to the preceding C–C bond (e.g., C'–C^{α} in Val, C^{α}–C^{β} in Leu). Panel c shows that in the absence of rotation about C'–C^{α}, rotations about χ_1 within a single rotamer of Val and Thr residues yield C–CH_3 (S^2_{axis}) order parameters >0.84.

the preceding C′–C^{α} bond (Figure 6c), which the simple model does not account for. Given this, rather than using a strict value of 0.84 as a threshold for the onset of rotamer averaging, in practice we choose to interpret experimental S^2_{axis} values of <0.8 as the onset of rotameric averaging in alanine, threonine, and valine side chains.

Leucines warrant additional consideration. It is well-known that leucines have a tendency to undergo simultaneous transitions of χ_1 and χ_2 dihedrals (Nicholson et al., 1992; Wong and Daggett, 1998). One might therefore expect that agreement between MD-observed and calculated p_{major} values (as in Figure 5) would deteriorate for leucines, compared to other methyl-bearing residues. However, as shown in Figure 5, this appears not to be the case, at least for eglin c studied here. No specific amino acid types in Figure 5 fall consistently above or below the line of slope 1. Interestingly, analysis of the effect of χ_1 rotations on leucine S^2_{axis} parameters shows that single-rotamer S^2_{axis} values can be significantly below 0.8 (Figure 6b). Nevertheless, the good correlation in Figure 5 shows that this is at least partially compensated and that use of Equations (6) and (7) still yields reasonable results. Due to the limited number of methyl sites in eglin c, further study on additional proteins will help to establish if leucines are more problematic than other residues. As eglin c has no isoleucines or methionines, information on these methyl types will also need further characterization.

For order parameters in the range of ~0.3–0.4, the possible populations over different rotamer states vary substantially. For example, an S^2_{axis} value of 0.3 could correspond to either a population over two rotamer states with a ratio of 0.58:0.42, or a population over three rotamer states with a ratio of 0.68:0.16:0.16. In general, it is difficult to distinguish these two cases experimentally, and therefore it is only possible to conclude that that particular side chain is

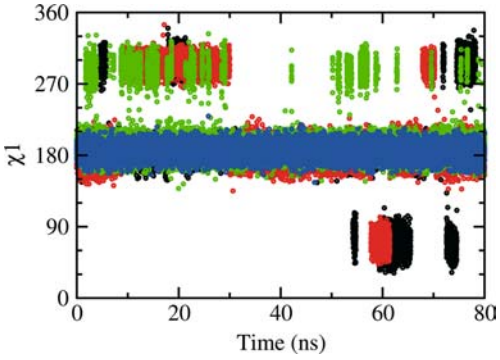

Figure 7. The distribution of χ_1 dihedrals of V14 (black), V18 (red), V34 (green), V54 (blue) observed in 80 ns of simulation recorded at 7.2 ps intervals.

undergoing extensive side-chain conformational averaging.

Timescales of rotamer sampling in eglin c

Although discussion so far has been in reference to motional processes as would be detected from NMR spin-relaxation measurements corresponding to sub-τ_m timescales (e.g., faster than a several nanoseconds), rotamer sampling can clearly occur on slower timescales and would therefore not be detected by ^{2}H relaxation measurements. This can be seen in the distribution of side-chain χ_1 dihedrals for four valines (V14, V18, V34, V54) observed in the 80 ns simulations (Figure 7). The mean S^2_{axis} values observed for these valines in the simulation are 0.75, 0.79, 0.71, and 0.93, respectively. It can be seen that while many transitions occur on sub- or low-ns timescales, some transitions require up to 60 ns of simulation time (e.g., V14 and V18 sampling $\chi_1 = 60°$) to occur. In such cases, relaxation measurements would not detect these large excursions and from this point of view, NMR relaxation-derived order parameters will always underestimate the extent of dynamics occurring, and the term 'protein dynamics' only has meaning in the context of the timescales considered.

Application to side-chain rotamers in ubiquitin

In a recent study, the rotamer populations of side-chain groups in ubiquitin were determined from residual dipolar couplings (RDC's) measured in

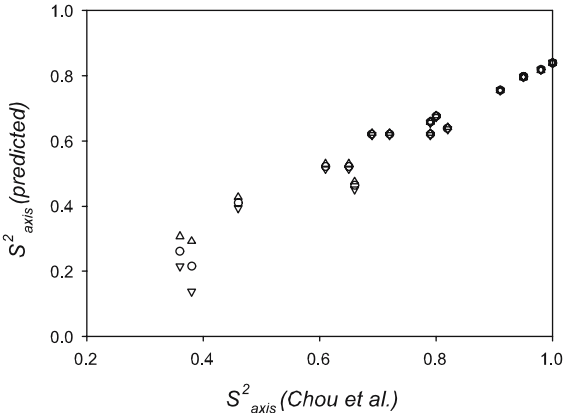

Figure 8. Comparison of order parameters (S^2_{axis}) for ubiquitin predicted by the MM-model in this paper with those reported previously based on two independent sets of RDC data (Chou et al., 2003). S^2_{axis} (predicted) were calculated using as input the major rotamer state populations from Table 1 of Chou et al.; these calculations used Equation (6) (open triangle up), Equation (7) (open triangle down), and the mean of the two (open circle). S^2_{axis} (Chou et al., 2003) are the order parameters obtained from explicit modeling of side-chains, as reported by Chou et al. (Chou et al., 2003); they were taken from Table 3 of that paper. (column 'S^2_D mean'). Val, Ile, Leu, and Thr S^2_{axis} values are shown.

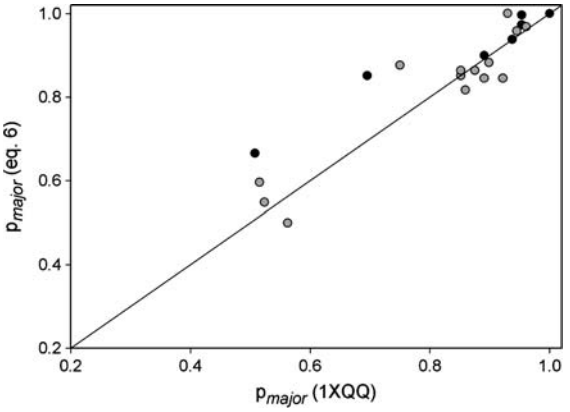

Figure 9. Comparison of p_{major} determined from DER (Lindorff-Larson et al., 2005) and from Equation (6). Values of p_{major} based on the DER family of ubiquitin structures were calculated directly from the ensemble (pdb code 1XQQ). For values of p_{major} calculated using Equation (6) for leucine and valine residues, experimental S^2_{axis} values for pro-R and pro-S methyl groups were averaged prior to use. Rotamer averaging in χ_1 (e.g. Val, Thr) and χ_2 (e.g. Leu, Ile-δ) are shown as black and gray circles, respectively. The upper right-most black circle corresponds to a superposition of eight data points, all from γ-methyl groups. The correlation coefficient (r) for all points is 0.94.

two different alignment media (Chou et al., 2003). Through explicit modeling of rotamer distributions, Chou et al. calculated methyl symmetry axis order parameters (S^2_{axis}) from these dipolar coupling data; the S^2_{axis} values therefore included side-chain motions occurring over much slower timescales (up to millisecond). A nice feature of the MM-based model presented here (Equations (5)–(7)) is that it applies to rotamer interconversions in a timescale-independent manner and is therefore applicable to RDC data, which can report on slower motions. To demonstrate this general applicability, we compare the ubiquitin S^2_{axis} values modeled from RDC data, as provided in Table 3 of Chou et al. to S^2_{axis} values calculated from our MM-based relations, using as input the major rotamer state populations from Table 1 (Chou et al., 2003) (Figure 8). Good agreement is found between the order parameters from dipolar coupling and those predicted from our model (Figure 8). Although a slope of 1 might be expected in the correlation plot, a slope of less than 1 is observed. A significant portion of this deviation is most likely due to the nature of the structural ensemble of Chou et al. which appears to have compressed distributions of χ_1 angles within major rotamers, as evidenced by standard deviations of only ~3 degrees (Table 1 of Chou et al.). Nevertheless, the strong agreement indicates that the assumptions in the model are reasonable, and that in many cases explicit modeling may not be necessary. That good agreement was also found between RDC (and J-coupling) derived and ^{2}H-relaxation derived S^2_{axis} parameters (Chou et al., 2003) indicates that side-chain rotamer interconversions are determined to a large extent by transitions on sub-τ_m timescales, although transitions between side-chain rotamer states can clearly happen on a timescale slower than that sensed by usual spin-relaxation experiments.

Recently, NMR-derived S^2_{axis} values were used as input for the refinement of the solution structure of ubiquitin (Lindorff-Larson et al., 2005). The resultant ensemble of structures therefore captures ps–ns motions of side-chains, including rotameric transitions contributing to S^2_{axis}. We have compared the side-chain rotamer populations from this DER (dynamic ensemble refinement) ensemble to those calculated using the same experimental S^2_{axis} values inserted into Equation (6). There is a relatively strong correlation (r = 0.94, Figure 9), showing that the simple MM-based model compares favorably with the method of Vendruscolo and co-workers and reinforces the utility of Equations (6) and (7) for directly estimating rotamer averaging. It should be pointed out that ubiquitin, unlike eglin c, has numerous leucine and isoleucine residues. Even with the caveat of χ_1 motion contributing to S^2_{axis} of the δ-methyls of leucine (Figure 6) and isoleucine, the two methods yield very similar values of p_{major} for χ_2 (Figure 9).

Motions in the interiors of proteins

This study has shown that even interior side-chain groups may actively sample multiple rotameric states, and that this is reflected in S^2_{axis} values obtained from spin relaxation corresponding to the ps–ns timescale. Considering that proteins are tightly packed molecules, this may be somewhat surprising: A jump to another rotameric state for a packed side chain potentially introduces a large strain on the system and would therefore be highly unlikely. The strain would result from movement of side-chain atoms into volumes occupied by other residues, unless concerted transitions occur simultaneously at several contacting residues. Yet the present analysis, in addition to previous studies on side-chain rotamers (Wong and Daggett, 1998; Mittermaier and Kay, 2001; Chou et al., 2003; Lindorff-Larson et al., 2005) clearly points toward the existence of rotameric transitions. It may be that the motion reflects the changing packing environment more than the inherent tendencies of the side chain itself. It will be important in future studies to determine the degree of coordination of these rotational transitions. The situation of rotameric jumps of methyl-bearing side chains is reminiscent of aromatic ring flips (Wagner, 1980), in that concerted motions must occur to allow for large rearrangements to be accommodated. On average, the interior methyl-bearing residues of both ubiquitin and eglin c appear to spend ~10% of the time in alternate χ_1 rotameric configurations, whether considering timescales up to a millisecond (ubiquitin) or up to a few nanoseconds (eglin c). How this 10% of 'stretched' conformational space is utilized may ultimately be related to function.

Correlated motions of side-chain with backbone atoms will render our model less accurate.

It is possible that such correlations could cause the rotamer population to be overestimated, and there are fewer transitions to minor rotamers. However, the agreement of rotamer predictions with actual rotameric states in the eglin c MD simulation (Figure 5); the agreement with the explicitly calculated ubiquitin RDC data (Figure 7); and the agreement between RDC- and relaxation-derived S^2_{axis} parameters (Chou et al., 2003) suggest that intra-residue correlated motions do not have a large effect, or that opposing effects cancel out.

Conclusions

We have presented a motional model for side-chains with methyl groups (excluding methionine), rationalized on the basis of molecular mechanics and observations in MD simulations. Using this model, we have semi-quantitatively correlated side-chain order parameters with the interconversion populations of rotamer states. In addition, we have presented an easy-to-use formula for estimating rotamer interconversion populations based on S^2_{axis} values obtained by any means (e.g., NMR ^{2}H and ^{13}C spin relaxation). This has been made possible by our confidence in the simulations, based on agreement of side-chain dynamical parameters with NMR data. The results presented in this paper underscore the utility of modeling tools in the analysis of NMR experiments. Combination of these two methods are highly complementary and serve to validate one another, resulting in a deeper understanding of structure, dynamics and thermodynamics of protein molecules.

Acknowledgements

We thank Dr Ernesto J. Fuentes and Michael W. Clarkson for helpful discussions. We also thank the scientific computing support at the University of North Carolina at Chapel Hill. This work was supported by research grants GM066009 (to A.L.L.) and RR08012 (to J.H.) from the National Institutes of Health, and MCB-0344354 (to A.L.L.) from the National Science Foundation.

References

Best, R.B., Clarke, J. and Karplus, M. (2004) *J. Am. Chem. Soc.*, **126**, 7734–7735.

Bremi, T., Bruschweiler, R. and Ernst, R.R. (1997) *J. Am. Chem. Soc.*, **119**, 4272–4284.

Brüschweiler, R. (2003) *Curr. Opin. Struct. Biol.*, **13**, 175–183.

Brüschweiler, R. and Wright, P.E. (1994) *J. Am. Chem. Soc.*, **116**, 8426–8427.

Butterfoss, G., Richardson, J. and Hermans, J. (2005) *Acta Cryst. D*, **61**, 88–98.

Case, D.A. (2002) *Accounts Chem. Res.*, **35**, 325–331.

Chou, J.J., Case, D.A. and Bax, A. (2003) *J. Am. Chem. Soc.*, **125**, 8959–8966.

Cornell, W.D., Cieplak, P., Bayly, C., Gould, I.R., Merz, K.M.J., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.

Fischer, M.W.F., Zeng, L., Pang, Y., Hu, W., Majumdar, A. and Zuiderweg, E.R.P. (1997) *J. Am. Chem. Soc.*, **119**, 12629–12642.

Henry, E.R. and Szabo, A. (1985) *J. Chem. Phys.*, **82**, 4753–4760.

Hu, H., Clarkson, M.W., Hermans, J. and Lee, A.L. (2003) *Biochemistry*, **42**, 13856–13868.

Jorgensen, W.L. (1998) In *Encyc. Comp. Chem.*, Vol. 3, Schleyer, P.v.R. (Ed.), Wiley, New York.

Lee, A.L., Sharp, K.A., Kranz, J.K., Song, X.J. and Wand, A.J. (2002) *Biochemistry*, **41**, 13814–13825.

Lindorff-Larson, K., Best, R.B., DePristo, M.A., Dobson, C.M. and Vendruscolo, M. (2005) *Nature*, **433**, 128–132.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559–4570.

MacKerell A.D. Jr., Brooks B., Brooks III C.L., Nilsson L., Roux, B., Won, Y. and Karplus M. (1998). In *Encyclopedia Computational Chemistry*, Vol. 1, P.v.R. Schleyer et al. (Eds.), John Wiley & Sons, Chichester, pp. 271–277.

MacKerell, A.D. Jr., Bashford, D., Bellott, M., Dunbrack, R.L. Jr., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., ReiherIII, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998) *J. Phys. Chem. B*, **102**, 3586–3616.

Mann, G., Yun, R.H., Nyland, L., Prins, J., Board, J. and Hermans, J. (2002). In *Computational methods for macromolecules: Challenges and applications – Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling, New York, October 12–14, 2000*, Schlick, T. and Gan H.H. (Eds.), Springer-Verlag, Berlin and New York, pp. 129–145..

Mittermaier, A. and Kay, L.E. (2001) *J. Am. Chem. Soc.*, **123**, 6892–6903.

Muhandiram, D.R., Yamazaki, T., Sykes, B.D. and Kay, L.E. (1995) *J. Am. Chem. Soc.*, **117**, 11536–11544.

Nicholson, L.K., Kay, L.E., Baldisseri, D.M., Arango, J., Young, P.E., Bax, A. and Torchia, D.A. (1992) *Biochemistry*, **31**, 5253–5263.

Palmer, A.G. (2001) *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 129–155.

Peng, J.W. and Wagner, G. (1992) *Biochemistry*, **31**, 8571–8586.
van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996). *Biomolecular simulation: the GROMOS96 manual and user guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich.
Wagner, G. (1980) *FEBS Lett*., **112**, 280–284.
Wand, A.J. (2001) *Nat. Struct. Biol*., **8**, 926–931.
Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H.M. (2003) *Nucleic Acids Res*., **31**, 489–491.
Wittebort, R.J. and Szabo, A. (1978) *J. Chem. Phys*., **69**, 1722–1736.
Wong, K.B. and Daggett, V. (1998) *Biochemistry*, **37**, 11182–11192.
Yang, D., Mittermaier, A., Mok, Y.K. and Kay, L.E. (1998) *J. Mol. Biol*., **276**, 939–954.