JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Determining Valine Side-Chain Rotamer Conformations in Proteins from Methyl ^{13}C Chemical Shifts: Application to the 360 kDa Half-Proteasome

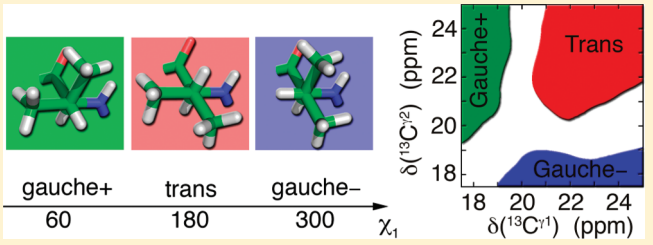
D. Flemming Hansen*,† and Lewis E. Kay‡

†Institute of Structural and Molecular Biology, Division of Biosciences, University College London, Gower Street, London WC1E 6BT, United Kingdom

‡Departments of Molecular Genetics, Biochemistry and Chemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada

Supporting Information

ABSTRACT: A method is presented for determining Val side-chain χ_1 rotamer distributions in proteins based exclusively on measured $^{13}C^{\gamma 1}$ and $^{13}C^{\gamma 2}$ chemical shifts. The approach selects an ensemble of 20 χ_1 values, calculates average methyl $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts via theoretical quantum chemical calculations and maximizes the agreement with the experimentally measured shifts using a genetic algorithm. The methodology is validated with an application involving six proteins for which $^{13}C^{\gamma}$ chemical shifts and three-bond methyl-backbone scalar couplings are available. The utility of the methodology is demonstrated with an application to the 360 kDa 'half-proteasome' where the χ_1 rotameric distributions of Val residues are calculated on the basis of chemical shifts. For the most part the χ_1 profiles so obtained compare very well with those generated from the high-resolution (2.3 Å) X-ray structure of the proteasome. Both NMR and X-ray distributions are cross-validated by comparing calculated $^{1}H-^{13}C$ methyl residual dipolar couplings with measured values, and the level of agreement is at least as good for the NMR derived χ_1 values. Notably, as the resolution of the X-ray data improves (rotamer distributions from 3.4 and 2.3 Å X-ray structures are compared with the NMR data), the agreement with the NMR gets significantly better. This emphasizes the importance of NMR approaches for the study of high molecular weight complexes that can be recalcitrant to high resolution X-ray analysis.



INTRODUCTION

Proteins are not static,[1] and the excursions that they undergo are often critical for function.[2−4] It has long been recognized that NMR spectroscopy is a powerful tool for characterizing these excursions in detail.[5,6] Even the simplest and most easily measured of all NMR parameters, chemical shifts, are useful probes, as they report on a weighted average of rapidly interconverting substates that are populated along a trajectory,[7,8] thus providing information about how such states are distributed in the system. It is now possible to calculate the overall *average backbone* structure of small, rigid proteins solely from chemical shifts using a combination of database approaches[9,10] and advanced computational procedures.[11,12] Such chemical shifts, when available through relaxation dispersion NMR measurements, can also be used to obtain detailed structural models of 'invisible' excited states[13−15] that are transiently populated at low levels. This opens the possibility for exploring important states of proteins that cannot be analyzed using the more traditional tools of structural biology. It is also possible to provide estimates of the amplitudes of protein backbone motions from a comparison of measured chemical shifts of backbone ^{13}C, ^{15}N, and ^{1}H nuclei with the corresponding 'random coil' values.[16]

The success of characterizing protein structure and dynamics through backbone chemical shifts has led to an interest in developing similar approaches involving side-chains. Our motivation in this regard stems from our work in two areas including (i) studies of supramolecular systems with molecular masses in the hundreds of kDa and (ii) investigations of the structure and dynamics of excited protein states. In the first case it is clear that in many applications the measurement of backbone chemical shifts is not possible due to sensitivity issues, yet measurement of methyl group chemical shifts remains feasible using methyl-TROSY[17,18] based approaches. Second, sensitive and robust experiments have been developed for the accurate measurement of methyl group chemical shifts[19,20] in low populated, transiently formed protein states that can potentially be exploited to provide detailed information on the structure and dynamics of these elusive conformers.

London and co-workers have recently described the dependence of ^{13}C side-chain chemical shifts on dihedral angles as a tool for conformational analysis in proteins,[21] and simple relations have been described for 'translating' Ile ($^{13}C^{\delta 1}$)[22] and Leu

Received: February 16, 2011
Published: May 05, 2011

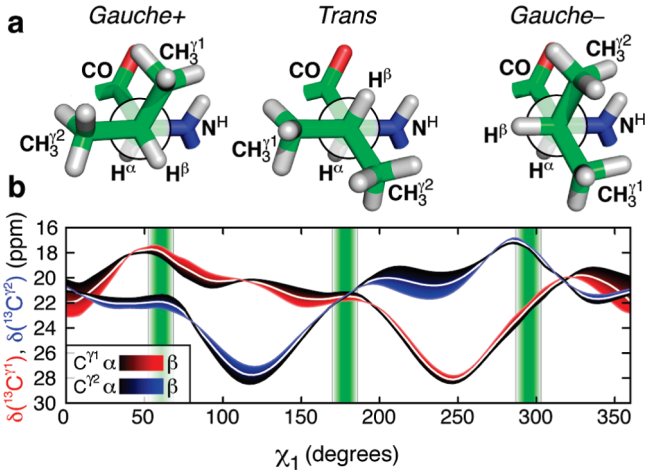


Figure 1. (a) Three canonical rotamer conformations of the Val side-chain based on a database of high-resolution crystal structures.[30,31] (b) $^{13}C^{\gamma 1}$ (red; *pro-R*) and $^{13}C^{\gamma 2}$ (blue; *pro-S*) chemical shifts derived from a DFT calculation for β-sheet (light) and α-helical (dark) backbone conformations (see Supporting Information for details); the average is shown with the white line. The green bars indicate the 'canonical values' from the set of high-resolution crystal structures, see text ($\chi_1 = 63° \pm 9°$ for gauche+; $\chi_1 = 174° \pm 9°$ for trans; $\chi_1 = 296° \pm 9°$ for gauche−).

$(^{13}C^{\delta 1,\delta 2})^{23,24}$ methyl ^{13}C chemical shifts into χ_2 rotamer populations for these residues. Our goal here is to extend this work with an approach for obtaining χ_1 torsion angle distributions for Val residues in proteins on the basis of measured $^{13}C^{\gamma 1,\gamma 2}$ shifts exclusively. Stereospecific assignments of the methyl pro-chiral carbons that are required in this analysis can be readily obtained using the method of Neri et al.[25] or alternatively by $^{13}CH_3$ labeling of either *pro-R* or *pro-S* methyls using the appropriate precursors.[26,27] We present a fully automated method that takes Val ^{13}C methyl shifts as input and generates a set of 20 rotamers that together best reproduces the experimental shift data using a genetic algorithm and a density functional theory (DFT) derived surface that relates $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts and χ_1. The utility of the approach is established by considering a database of 22 Val residues from six different proteins for which chemical shifts and three-bond scalar couplings have been measured with high precision. Subsequently, an application to the 360 kDa half-proteasome,[18] $\alpha_7\alpha_7$, is presented, where χ_1 rotamer distributions are obtained for the Val residues. For the majority of Val residues the distribution is localized predominantly to a single χ_1 value, and there is excellent agreement with angles predicted from the proteasome X-ray structure,[28,29] providing further validation of the methodology. The method is robust and simple to use and promises to extend the utility of methyl chemical shifts as probes of side-chain conformation and dynamics.

RESULTS AND DISCUSSION

An analysis of a database of high-resolution crystal structures[30,31] with *B*-factors less than 20 Å^2 establishes that Val side-chains can populate three different χ_1 rotameric states, including gauche+ (g+, $\chi_1 \approx 63°$), trans (t, $\chi_1 \approx 174°$), and gauche− (g−, $\chi_1 \approx 296°$) and that the average populations of each of the three states are $\overline{p}_{g+} = 6\%$, $\overline{p}_t = 74\%$, $\overline{p}_{g-} = 20\%$, Figure 1a. Importantly, the rmsd of χ_1 values within each of these states is less than 10°,[31] indicating that Val χ_1 dihedral angle motions are best described in terms of averaging about and jumps between each of the three (g+, t, g−) canonical conformations.

Previously it was shown that the chemical shifts of Ile $(^{13}C^{\delta 1})^{22}$ and Leu $(^{13}C^{\delta 1,\delta 2})^{23,24}$ methyl carbons can be estimated from a simple γ-gauche model[32,33] whereby the dominant effect is the number, orientation, and type of heavy atoms in the γ position relative to the methyl ^{13}C spin in question. This means that in the case of Leu, for example, the methyl $^{13}C^{\delta j}$ chemical shift depends in a sensitive manner on the dihedral angle between $C^{\delta j}$ and C^α. The Ile $^{13}C^{\delta 1}$ chemical shift, in turn, depends on γ-gauche effects caused by both C^α and $C^{\gamma 2}$ atoms that reflect the dihedral angles between $C^{\delta 1}$ and C^α and between $C^{\delta 1}$ and $C^{\gamma 2}$, respectively, both reporting on χ_2.

Prior to using Val $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts to analyze χ_1 rotamer distributions, we were interested in establishing through DFT calculations how such shifts depend on the side-chain torsion angle; such DFT calculations are integral to the methodology proposed below. Moreover, we wanted to explore whether the γ-gauche effect that had been exploited in our previous studies of Ile and Leu dynamics[22,23] was also a dominant contributor to methyl ^{13}C shifts for Val. The carbonyl carbon (CO) and the amide nitrogen (N) of Val both occupy γ positions relative to the methyl carbon atoms, and it is expected, therefore, that the chemical shifts of the $^{13}C^{\gamma 1,\gamma 2}$ nuclei ($\gamma 1$= *pro-R*, $\gamma 2$ = *pro-S*) will be sensitive to a sum of two γ-gauche effects that both depend on the χ_1 dihedral angle. Empirically, we can write the $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts as $\delta(^{13}C^{\gamma i}) = \delta^{\gamma g}(\chi_1)_{C\gamma i} + \langle\delta\rangle_{\chi 1}$, where $\delta^{\gamma g}(\chi_1)_{C\gamma i}$ is the contribution from the γ-gauche effect, while $\langle\delta\rangle_{\chi 1}$ includes contributions from other terms that are independent of χ_1. Inspection of Figure 1a allows one to write,

$$\begin{aligned}
&\delta^{\gamma g}(60°)_{C\gamma 1} = \gamma g(N) + \gamma g(CO) && \delta^{\gamma g}(60°)_{C\gamma 2} = \gamma g(CO)\\
&\delta^{\gamma g}(180°)_{C\gamma 1} = \gamma g(CO) && \delta^{\gamma g}(180°)_{C\gamma 2} = \gamma g(N)\\
&\delta^{\gamma g}(300°)_{C\gamma 1} = \gamma g(N) && \delta^{\gamma g}(300°)_{C\gamma 2} = \gamma g(N) + \gamma g(CO)
\end{aligned} \qquad (1)$$

where $\gamma g(N)$ and $\gamma g(CO)$ are γ-gauche effects from N and CO, respectively, and 60°, 180°, 300° correspond to the g+, t, and g− conformations (Figure 1a).

A series of DFT calculations was carried out on small model complexes as a function of the χ_1 dihedral angle, neglecting any tertiary interactions (details are provided in Supporting Information). Figure 1b plots $\delta(^{13}C^{\gamma i})$; in general, the chemical shifts derived from the DFT calculations follow the predictions based on the empirical γ-gauche model.[32,33] Interestingly, for the trans conformation ($\chi_1 = 180°$) $\delta(^{13}C^{\gamma 1}) \approx \delta(^{13}C^{\gamma 2})$, indicating that the γ-gauche effects from the amide nitrogen and the carbonyl carbon are similar.

We have carried out DFT calculations for both α-helical and β-sheet backbone conformations, illustrated in Figure 1b. Also shown is the average over both secondary structure elements. For all three rotamer conformations the $^{13}C^{\gamma}$ side-chain chemical shifts depend little on whether the Val residue is in an α-helix or a β-sheet backbone conformation (<1 ppm) so that, in studies where the protein secondary structure is not known *a priori*, the average shift value can be used without introducing significant error. Of course, in cases where the secondary structure is known, more accurate $\delta(^{13}C^{\gamma i})$ surfaces can be used in the analyses that follow below.

dx.doi.org/10.1021/ja2014532 |*J. Am. Chem. Soc.* 2011, 133, 8272–8281

Determining Side-Chain Rotamer Distributions from Methyl Chemical Shifts. Prior to presenting the methodology used here for obtaining $\{p_{g+}, p_t, p_{g-}\}$ values describing the χ_1 rotamer populations in Val residues based on methyl $^{13}C^{\gamma1,\gamma2}$ chemical shifts, it is worth summarizing how the corresponding χ_2 rotamer populations are obtained for Leu and Ile residues. As described previously, these distributions are derived from $^{13}C^{\delta1,\delta2}$ (Leu) and $^{13}C^{\delta1}$ (Ile) chemical shifts.[21,22,24] A significant advantage for both Leu and Ile is that to excellent approximation only a pair of χ_2 rotamers are sampled in each case: (g+,t)[23,31] for Leu and (t,g−)[22,31] for Ile. This significantly simplifies the problem because it implies, as least in principle, that from a single methyl chemical shift measurement it is possible to extract the populations of the two states in each case. For example, for Leu, the population of the trans χ_2 rotamer conformation is well approximated by the relation $p_{trans} = [\Delta\delta(^{13}C) + 5\ ppm]/10\ ppm$,[24] where $\Delta\delta(^{13}C^{\delta}) = \delta(^{13}C^{\delta1}) - \delta(^{13}C^{\delta2})$. Moreover, by using the difference between $^{13}C^{\delta1}$ and $^{13}C^{\delta2}$ chemical shifts the effects of electrostatics, ring currents, and other tertiary contributions cancel to first order, allowing for an accurate estimation of p_{trans}. In the case of Ile it has recently been shown that the χ_2 populations can be calculated from a linear relationship that relates $\delta(^{13}C^{\delta1})$ to $p_{gauche-}$ and $p_{trans} = 1 - p_{gauche-}$ and that the populations so obtained agree reasonably well with those predicted from measured three-bond scalar couplings.[22]

Unfortunately, the situation is more complex for Val. An approach that uses $\Delta\delta(^{13}C^{\gamma}) = \delta(^{13}C^{\gamma1}) - \delta(^{13}C^{\gamma2})$ to determine the χ_1 rotameric distributions for this residue is not possible because Val side-chains populate all three rotamer states, Figure 1a, and since $p_{g+} + p_t + p_{g-} = 1$, there are two unknown populations. Consequently, two experimental parameters are necessary, and the 'raw' chemical shifts, $\delta(^{13}C^{\gamma1})$ and $\delta(^{13}C^{\gamma2})$, are therefore used. We show below that effects from ring currents and electrostatics can be ignored to a good approximation, despite the fact that individual shifts are used. This is consistent with what has been noted previously in our studies of Ile χ_2 rotamer distributions calculated on the basis of only the $^{13}C^{\delta1}$ chemical shift.[22]

A Method for Obtaining Val χ_1 Rotameric Distributions from Methyl Chemical Shifts. The goal in the present study is to develop a simple approach for calculating Val χ_1 distributions based only on methyl $^{13}C^{\gamma1,\gamma2}$ chemical shifts. While it is clear that in the cases of small to moderately sized proteins these distributions can be obtained from other data, such as scalar or residual dipolar couplings, our choice of restricting the input to only chemical shifts is motivated by the desire to develop methodology that can be applied to studies of supramolecular protein systems, such as the proteasome described below, or to excited protein states. As discussed in the introduction, in these cases it is not possible to measure the couplings that are required to generate the desired rotameric distributions.

The basic crux of our approach is to 'describe' each Val χ_1 rotamer in terms of a set of 20 χ_1 dihedral angles that 'fit' the measured $^{13}C^{\gamma1,\gamma2}$ chemical shifts best. In cases where the χ_1 angle is 'frozen', then all 20 angles in the chosen set would be expected to be very narrowly distributed about the χ_1 value. By contrast, if the Val in question is dynamic, then a wider distribution will be obtained that reflects the motion. A value of 20 has been chosen rather arbitrarily; it is large enough to ensure that a random coil distribution, $\overline{p}_{g+} = 6\%$, $\overline{p}_t = 74\%$, $\overline{p}_{g-} = 20\%$ can be calculated if necessary. These 20 χ_1 values are subsequently 'translated' into $^{13}C^{\gamma1,\gamma2}$ chemical shifts from a lookup table that tabulates $\delta(^{13}C^{\gamma1,\gamma2})$ as a function of χ_1 as calculated by DFT (see below and Supporting Information). It is assumed that the 20 rotameric states are in fast exchange so that the chemical shifts of each of the states can be averaged to obtain the observed shift. Once the 'best' set of 20 states is found, it is straightforward to calculate the populations,$\{p_{g+}, p_t, p_{g-}\}$ which is the desired goal (see below).

To validate the approach, the first step we consider is a set of six proteins for which four three-bond scalar couplings $\{^3J(C^{\gamma1}, N), ^3J(C^{\gamma2},N), ^3J(C^{\gamma1},CO), ^3J(C^{\gamma2},CO)\}$, and $^{13}C^{\gamma1,\gamma2}$ chemical shift assignments are available for Val residues. This set includes ubiquitin,[34] protein GB3,[34] the C-terminal SH2 domain from phospholipase Cγ (PLCC SH2),[35] protein L,[36] HIV protease,[34] and maltose binding protein, for which data was measured in the present study (Supporting Information). Our immediate goal is to calculate a reference set of χ_1 rotameric distributions for each Val residue that satisfies the measured three-bond scalar couplings and show that the calculated chemical shifts from this distribution agree with those measured. This will establish that it is possible to 'convert' the distribution into accurate values of $\delta(^{13}C^{\gamma1})/\delta(^{13}C^{\gamma2})$ using the DFT hyper-surfaces we have calculated. A second goal is to establish that the distributions obtained from scalar couplings and chemical shifts are very similar. It is well recognized that accurate dihedral angle distributions can be obtained from scalar couplings;[34] establishing that the corresponding χ_1 rotamer profiles obtained from chemical shifts are in good agreement indicates that reasonable χ_1 distributions for Val can be generated exclusively from chemical shifts.

Each of the 22 Val residues from the six 'database' proteins for which a set of four three-bond scalar couplings was measured was considered independently. A best set of 20 χ_1 dihedral angles, corresponding to the '*J*-reference rotameric distribution' for that residue, was selected using a genetic algorithm described in Materials and Methods. Here a target function is minimized,

$$E = E_{MF} + E_J \tag{2}$$

where E_{MF} and E_J are energies that 'force' the dihedral angles in the best set to lie within the χ_1 ranges observed experimentally for Val (E_{MF}) and that ensure that the scalar couplings predicted from the distribution, using well-established Karplus relationships[34] (E_J), are in good agreement with the experimental values. Details are given in Materials and Methods.

Once the scalar coupling-based χ_1 distribution profile was obtained, it was used to calculate methyl chemical shifts using the DFT derived surface (Figure 1b) that relates shifts and dihedral angles. Values of $\delta(^{13}C^{\gamma1})/\delta(^{13}C^{\gamma2})$ were calculated as

$$\delta_{calc}(^{13}C^{\gamma i}) = \frac{1}{N}\sum_k \delta_{DFT,i}(\chi_{1,k}) \tag{3}$$

where $\chi_{1,k}$ is the χ_1 angle of the kth member of the *J*-reference ensemble ($1 \le k \le 20$), $N = 20$, and $\delta_{DFT,i}(\chi_{1,k})$ is the chemical shift calculated for $^{13}C^{\gamma i}$ from the DFT hyper-surface (Figure 1b) for a dihedral angle of $\chi_{1,k}$. In all calculations, a DFT surface was used that is an average over those calculated assuming α-helical and β-sheet geometries, although the differences in the two surfaces are small (Figure 1b). The correlation between calculated and experimentally measured chemical shifts is shown in Figure 2. The rmsd between δ_{calc} and δ_{exp}, 0.83 ppm, obtained here for Val $^{13}C^{\gamma1,\gamma2}$ chemical shifts is slightly better than that obtained for the prediction of $^{13}C^{\alpha}$ (rmsd = 1.0 ppm) and

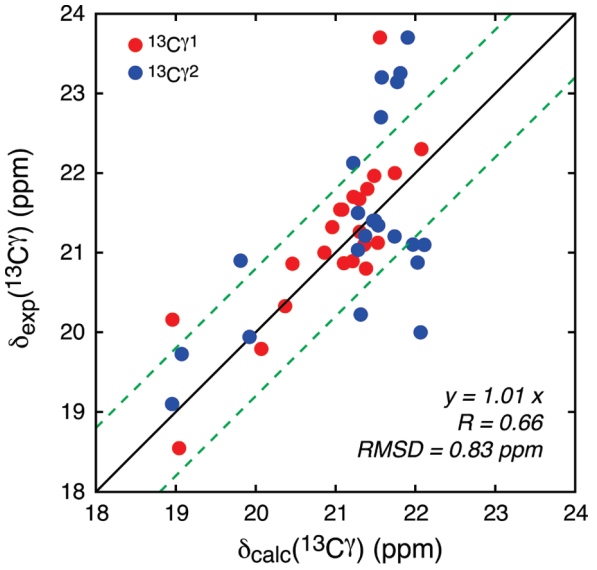


Figure 2. Comparison of experimentally measured $^{13}C^{\gamma}$ chemical shifts, δ_{exp}, with chemical shifts derived from the *J*-reference ensemble and calculated using a DFT-derived hyper-surface (Figure 1b) and eq 3, δ_{calc}. $^{13}C^{\gamma 1}$,*pro-R* ($^{13}C^{\gamma 2}$, *pro-S*) chemical shifts are shown in red (blue); green dotted lines are given by $y = x \pm \sigma$, where $\sigma = \text{rmsd}(\delta_{calc} - \delta_{exp})$.

$^{13}C^{\beta}$ (rmsd = 1.1 ppm) chemical shifts from protein structure.[37−39] Interestingly, the agreement is significantly better for $C^{\gamma 1}$ (rmsd = 0.59 ppm) than it is for $C^{\gamma 2}$ (rmsd = 1.0 ppm), likely due to the large dependence of chemical shift on χ_1, $\partial\delta_{calc}/\partial\chi_1$, for $C^{\gamma 2}$ around the most populated trans conformation. Thus, errors in the rotamer distribution obtained from scalar couplings, reflecting errors in the couplings themselves and/or errors in the Karplus relations that relate χ_1 to ^3J, will affect calculated $^{13}C^{\gamma 2}$ chemical shifts much more than the corresponding $^{13}C^{\gamma 1}$ values, contributing to the poorer correlation for $^{13}C^{\gamma 2}$ that is observed in Figure 2. The correlation in Figure 2 indicates that Val side-chain $^{13}C^{\gamma 1}$ and $^{13}C^{\gamma 2}$ chemical shifts can be determined, to an accuracy of better than 0.83 ppm, from a χ_1 distribution using the calculated DFT surface. This is an upper bound on potential errors in the mapping approach to produce chemical shifts from a distribution, since it assumes that both the measured scalar couplings and the Karplus equations are error free. As shown below, with an application to the 360 kDa $\alpha_7\alpha_7$ particle, this level of accuracy is sufficient to provide robust measures of χ_1 values.

A χ_1 ensemble can also be calculated directly from $^{13}C^{\gamma 1}$ and $^{13}C^{\gamma 2}$ chemical shifts, following an analogous approach to that described above for obtaining distributions from scalar couplings. Here we define the following energy function to bias the search

$$E = E_{MF} + E_{\delta} \quad (4)$$

where E_{MF} ensures that the distribution of χ_1 values lies within the allowed ranges (see above) and E_{δ} is an energy term that 'forces' the calculated chemical shifts (eq 3) to agree with the experimentally derived values.

The scalar coupling and chemical shift rotamer distributions obtained for Val 21 of protein GB3 are shown in a and b of Figure 3, respectively. This residue has been chosen in particular because the agreement between the two distributions is among the worst of those obtained (see Figure 3d). Yet nevertheless, a comparison of a and b of Figure 3 shows that the two distributions

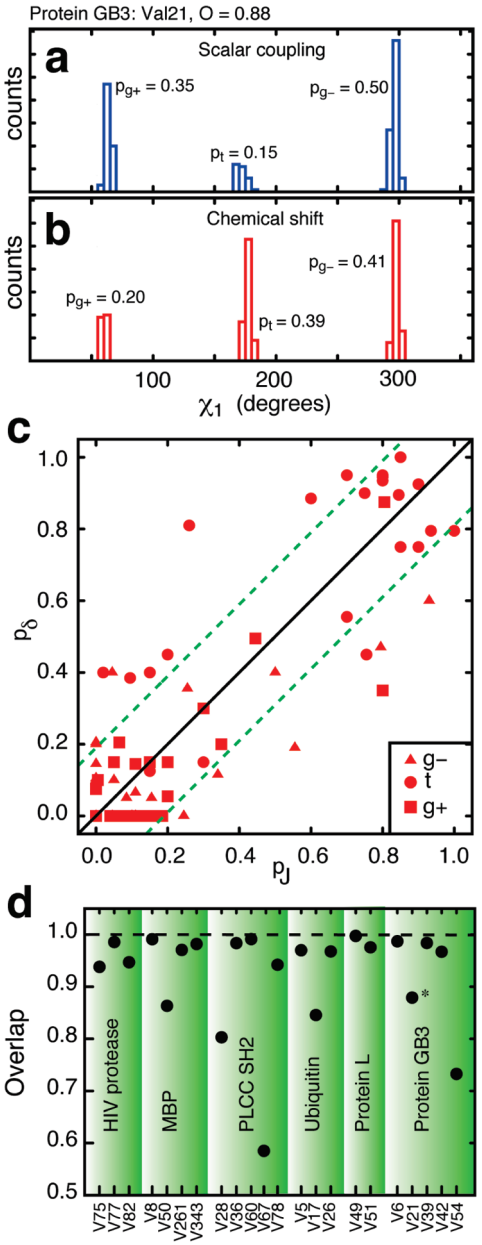


Figure 3. Comparison of rotamer (g+, t, g−) populations determined from Val methyl-backbone three-bond scalar couplings and from chemical shifts, $\delta(^{13}C^{\gamma 1})$ and $\delta(^{13}C^{\gamma 2})$. (a,b) Distributions calculated for Val 21 of protein GB3, based on input scalar couplings (a) or chemical shifts (b). The overlap, *O*, between the two distributions is calculated using eq 5. (c) Correlation plot of rotamer populations derived from chemical shifts (p_δ) vs the corresponding values generated using scalar couplings (p_J), as described in the text. (d) The overlap, eq 5, calculated for all the Val residues used in the simulations described above. Val 21 of protein GB3, that is 'featured' in a,b is marked with an asterisk.

are very similar. Thus, the dynamic behavior of this residue, as evidenced by the sizable populations of all three rotameric states, is captured in both distributions. Figure 3c shows a direct comparison of the fractional populations for all Val residues used in this study, estimated from the chemical shift (*Y*-axis) and scalar coupling (*X*-axis) distributions. The three populations, p_{g+}, p_t, and p_{g-} are calculated from each of the χ_1 distributions by

dx.doi.org/10.1021/ja2014532 |*J. Am. Chem. Soc.* 2011, 133, 8272–8281

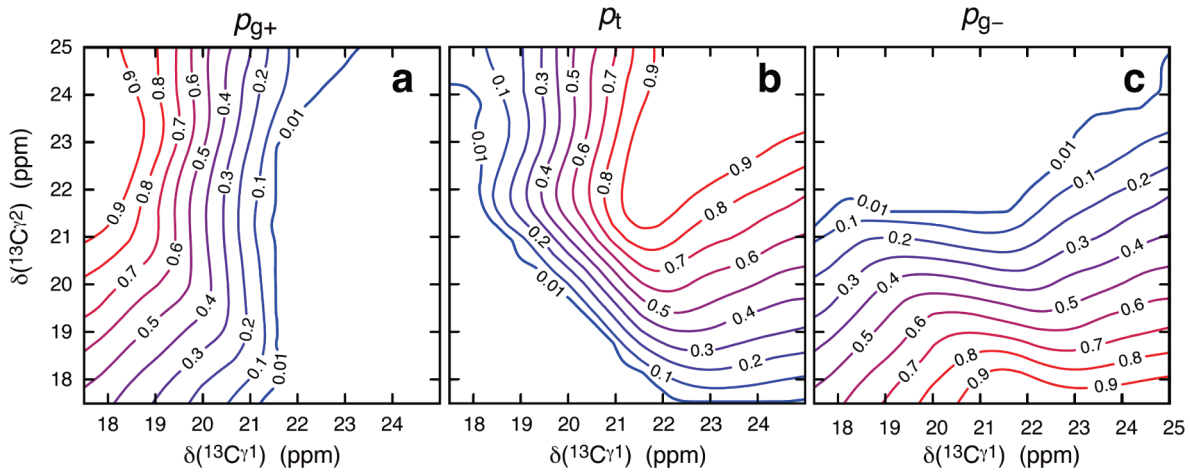


Figure 4. DFT-derived hyper-surfaces for the three Val χ_1 rotamer populations. (a) Populations of g+ ($\chi_1 \in [0°,120°)$), t ($\chi_1 \in [120°,240°)$) and g− ($\chi_1 \in [240°,360°)$) conformers are respectively shown in a, b, c, as functions of $^{13}C^{\gamma 1}$ and $^{13}C^{\gamma 2}$ chemical shifts.

projecting the χ_1 onto the three canonical states. That is, if a given χ_1 angle in an ensemble is between 0° and 120°, $\chi_1 \in [0°,120°)$, it is assigned to the g+ population, values of $\chi_1 \in [120°,240°)$ are assigned to state t, while values of $\chi_1 \in [240°,360°)$ are assigned to g−. Most of the χ_1 distributions, including those for Val 21 of protein GB3, are closely centered around the canonical dihedral angles (63° for g+, 174° for t, 296° for g−) so that the approximation of projecting all χ_1 angles onto three states is reasonable. The rmsd between the populations determined from the scalar coupling-driven (p_J) and the chemical shift-based (p_δ) distributions can be calculated as

$$\sqrt{\frac{1}{3N}\sum_{\text{Val}}\sum_{r\in\{\text{g+,t,g-}\}}\left(p_J(r)-p_\delta(r)\right)^2}$$

where N is the number of Val residues included in the analysis. A value of 0.19 is obtained, indicated by the dashed green lines of Figure 3b. In direct analogy with the rmsd obtained between δ_{calc} and δ_{exp}, the value of 0.19 calculated presently is an upper bound for the uncertainty of predicting rotamer populations from chemical shifts since it assumes that the scalar coupling J-reference ensemble is 'correct'. As discussed above, uncertainties in the scalar coupling distributions are expected to arise as a consequence of errors in both the measured couplings and in the Karplus relations that relate the χ_1 dihedral angle to the scalar couplings. The distributions of the three populations, p_{g+}, p_t, and p_{g-} for the six proteins considered in this study, obtained from either scalar couplings or chemical shifts, are tabulated in Supporting Information.

Another convenient way of comparing the distributions is to calculate the overlap between them in a manor analogous to taking a dot-product between two vectors. Thus, from the three populations, p_{g+}, p_t, and p_{g-} we calculate an overlap, O, as

$$O = \frac{\sum_r p_J(r)p_\delta(r)}{\sqrt{\sum_r p_\delta(r)p_\delta(r)}\sqrt{\sum_r p_J(r)p_J(r)}} \qquad (5)$$

In general, a very good overlap is noted between distributions derived from chemical shifts and from scalar couplings, as seen in Figure 3d where $O > 0.9$ for 16 of the 22 Val residues. It is noteworthy that $O = 0.88$ for Val 21 of protein GB3, yet similar distributions are still observed (Figure 3a,b). In summary, the strong correlation between the χ_1 population distributions generated on the basis of chemical shifts and scalar couplings (Figure 3c,d) validates using Val $^{13}C^{\gamma}$ chemical shifts to estimate rotamer preferences. On the basis of the results of Figure 3c (green dashed lines) an upper bound on the uncertainties in p_{g+}, p_t, and p_{g-} values is 0.19.

A Population Hyper-Surface. In the plots of Figure 3 a full simulation is performed for each Val residue; that is, the full genetic algorithm is run for each set of experimentally measured chemical shifts $\{\delta(^{13}C^{\gamma 1}), \delta(^{13}C^{\gamma 2})\}$ to obtain the set of 20 χ_1 angles that best fits the data. This calculation takes approximately two minutes per residue on a standard PC with a 2 GHz processor. Faster calculations of rotamer populations from chemical shifts will often be required, for example, if the chemical shifts are to be used in restrained MD simulations. Therefore, we have calculated the full hyper-surface for the range of Val $^{13}C^{\gamma}$ chemical shifts that are normally encountered in proteins; the hyper-surfaces for the three populations are shown in Figure 4. For each $\{\delta(^{13}C^{\gamma 1}), \delta(^{13}C^{\gamma 2})\}$ grid point of the hyper-surface (one grid point for every 0.3 ppm) the best set of 20 χ_1 angles is calculated as described above (using the full genetic algorithm), values of $\{p_{g+}, p_t, p_{g-}\}$ subsequently determined and smoothed two-dimensional (2D) contour plots obtained using a 2D spline (calculated with the *scipy* and *pylab* extension libraries to Python). With the hyper-surfaces of Figure 4 it is straightforward to obtain Val rotamer populations directly from $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts; yet if the full distributions are to be obtained (e.g., Figure 3b), a complete calculation will still be required.

An Application to the 1/2 Proteasome. The 20S core particle (20S CP) proteasome is a key cellular component that plays a major role in protein degradation, thereby regulating processes such as cell division, gene expression, signal transduction, and apoptosis.[40−42] Detailed X-ray studies have established that the 20S CP consists of a barrel-like structure, comprising four rings, arranged as $\alpha_7\beta_7\beta_7\alpha_7$, where each ring is a heptamer with either seven α (α_7) or seven β (β_7) subunits.[29] In the case of the archaeal version studied here each of the α or β subunits is identical. Detailed solution NMR studies have shown that the proteasome is highly dynamic and that its function critically

dx.doi.org/10.1021/ja2014532 |*J. Am. Chem. Soc.* 2011, 133, 8272–8281

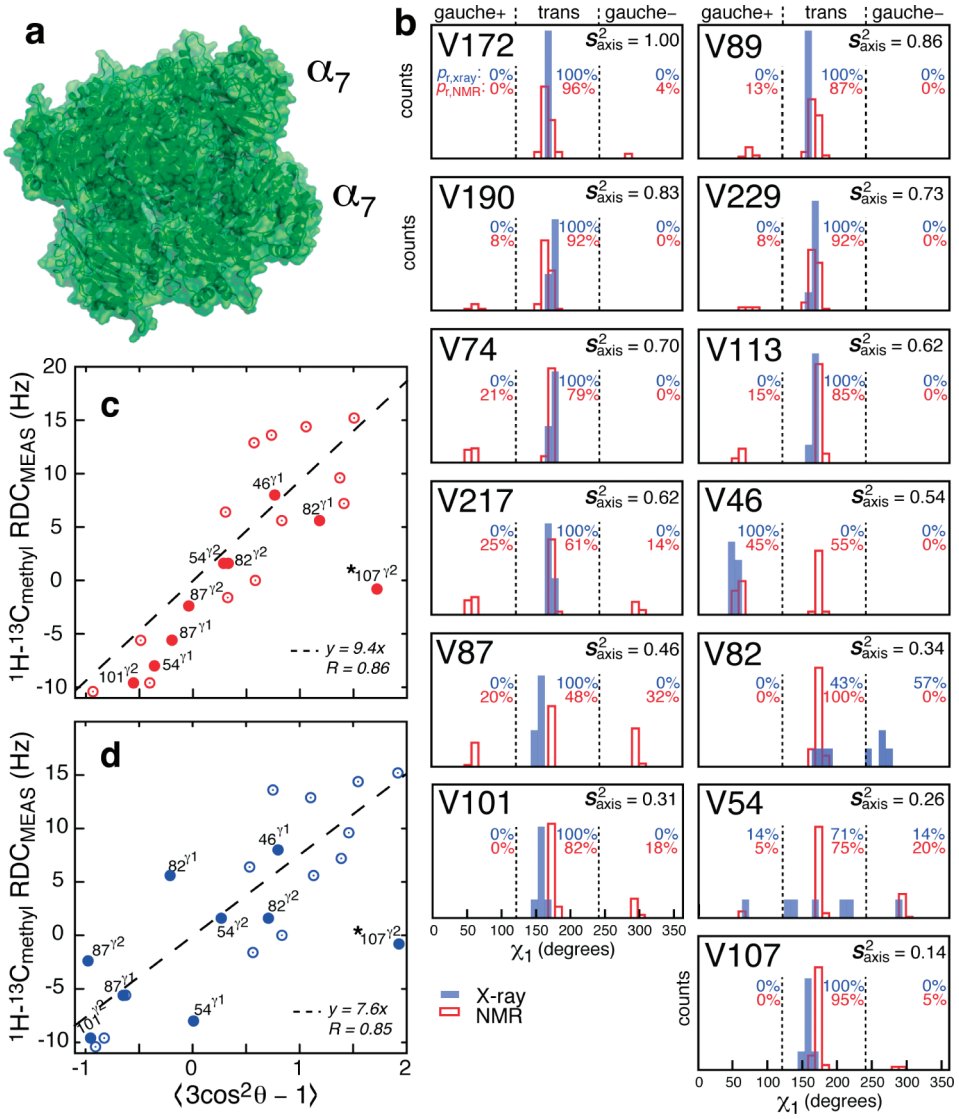


Figure 5. Application to the 1/2 proteasome. (a) Structure of the $\alpha_7\alpha_7$ (360 kDa) particle considered here. (b) Val χ_1 torsion angle distributions in $\alpha_7\alpha_7$ derived from NMR (red) and the X-ray structure of the proteasome (blue), as described in the text. Values of S^2_{axis}, averaged over the $\gamma1$ and $\gamma2$ positions, are reported. (c) Correlation plot of RDC_{MEAS} vs $RDC'_{NMR} = \langle 3\cos^2\theta - 1\rangle$ for both $\gamma1$ and $\gamma2$ methyl groups with solid (open) circles denoting residues with S^2_{axis} values less (greater) than 0.6. (d) As in (c) but RDC_{MEAS} vs $RDC'_{X\text{-}RAY} = \langle 3\cos^2\theta - 1\rangle$. Note that not all of the RDC values in $\alpha_7\alpha_7$ could be measured due to overlap. In (c) and (d) Val 107 $C^{\gamma2}$ (*) was omitted in the calculation of the best-fit line; R is the Pearson's correlation coefficient for the best-fit line.

depends on motions[18,43] that span a wide range of time-scales. Studies to date have focused on (i) measurement of rapid picosecond−nanosecond time-scale dynamics from ^{13}C, ^2H, and ^1H relaxation experiments,[18,44] (ii) quantification of ms processes by relaxation dispersion,[18,20] and (iii) characterization of yet slower events on the seconds time-scale through magnetization exchange spectroscopy.[43] A determination of torsion angle distributions would also provide valuable dynamical information[34,36] of a nature that would complement existing measures. We have calculated such distributions for Val, based on $^{13}C^{\gamma1,\gamma2}$ chemical shifts measured on a sample of the 1/2 proteasome ($\alpha_7\alpha_7$), Figure 5a, comprising pairs of stacked α_7 rings. Previous studies have shown that $\alpha_7\alpha_7$ is a good model system for studying the dynamics of α-subunits in the context of the 20S CP.[18] However, its smaller size relative to the full 20S particle (360 vs 670 kDa) translates into better spectra and more quantitative measurements.

Val rotamer population distributions in $\alpha_7\alpha_7$, obtained using the chemical shift-based approach described here (red), are plotted in Figure 5b. Included in this figure are the corresponding χ_1 profiles from the high-resolution (2.3 Å, pdb: 1YA7) X-ray structure of the 20S CP (blue),[28] where seven distinct values of χ_1 are obtained from the seven α-subunits in each α_7 ring. For each residue the value of the three-fold methyl axis order parameter squared, S^2_{axis}, averaged over $\gamma1$ and $\gamma2$ positions is also shown.[18] The majority of Val residues (9/13) are predominantly in the trans χ_1 conformation in both solution and the solid state (i.e., $p_t \geq 70\%$). The remaining four residues have lower than average S^2_{axis} values, and not surprisingly there are deviations between what is observed in solution and in the solid state. In

dx.doi.org/10.1021/ja2014532 |J. Am. Chem. Soc. 2011, 133, 8272–8281

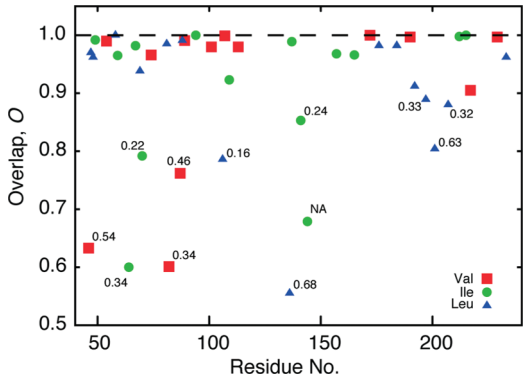


Figure 6. Plot of *O* values (eq 5) as a function of residue number for Ile, Leu, and Val residues of $\alpha_7\alpha_7$. Values of S^2_{axis}, averaged over the $\gamma 1$ and $\gamma 2$ positions, are listed for residues where *O* values are ≤ 0.9.

solution three of the four residues populate at least two rotameric states, while only a single conformer prevails in the solid. Interestingly for Val 82 $p_t = 100\%$ is calculated from the NMR data, while $\{p_t = 43\%, p_{g-} = 57\%\}$ is obtained from the X-ray structure. As we show below, measured methyl $^1H-^{13}C$ residual dipolar couplings for Val 82 $\gamma 1$ are in better agreement with predictions based on the NMR rotamer distribution.

In order to cross-validate the Val χ_1 torsion angle distributions for the 1/2 proteasome that have been obtained from the chemical shift methodology presented here we have calculated $RDC'_{NMR} = \langle 3\cos^2\theta - 1\rangle$ for each Val residue where θ is the angle between the $H_3C_{methyl}-C$ bond and the symmetry axis of the seven-fold symmetric $\alpha_7\alpha_7$ particle, and the angular brackets denote averaging over all χ_1 angles in the distribution. In computations of RDC'_{NMR} only Val χ_1 torsion angles were allowed to vary from the X-ray structure in the calculations, and identical χ_1 values were assumed for the same Val residues in each of the 14 α-subunits (i.e., for the 14 Val 172 residues, for example). Calculated RDC'_{NMR} values are compared with measured RDCs, $RDC_{MEAS} = A\langle 3\cos^2\theta - 1\rangle$, where *A* is a constant that depends, among other things, on the degree of alignment and, in this case, $\langle..\rangle$ indicates averaging of the $^1H-^{13}C$ methyl bond vectors from molecular dynamics. Assuming that the extent of averaging is the same for each methyl, it follows that a plot of RDC_{MEAS} vs RDC'_{NMR} would be linear with zero intercept. Figure 5c plots RDC_{MEAS} vs RDC'_{NMR} for both $\gamma 1$ and $\gamma 2$ methyl groups with solid (open) circles denoting residues with S^2_{axis} values less (greater) than 0.6. The best-fit line $y = mx$ is shown with dashes along with Pearson's correlation coefficient, $R = 0.86$ ($R^2 = 0.78$). We have also calculated $RDC''_{X\text{-}RAY}$ values by averaging over $(3\cos^2\theta - 1)$ for the seven copies of each residue in each α_7 ring that were not enforced to be the same in the high-resolution X-ray structure.[28] The corresponding correlation between RDC_{MEAS} vs $RDC'_{X\text{-}RAY}$ is shown in Figure 5d. Of note, with the exception of Val 107 $C^{\gamma 2}$, the experimental RDC data is reasonably well reproduced by both NMR and X-ray distributions. Certainly the quality of the 'NMR' correlation is no worse than the 'gold-standard' high-resolution X-ray model, perhaps even slightly better. Deviations arise to a large extent from residue-specific variations in dynamics that are not accounted for in the simple correlation presented (see below). A comparison of RDC_{MEAS} vs RDC′ values for Val 82 shows that, while similar levels of agreement between predicted and measured couplings are obtained for the $\gamma 2$ methyl, the deviation between RDC_{MEAS} and RDC′ for the $\gamma 1$ moiety is significantly larger in the case of the X-ray derived RDC′ value (compare c and d of Figure 5). Recall that Val 82 is one of four residues where significant differences in torsion angle distributions are predicted from NMR and X-ray data; the RDC data establish that the NMR rotamer distribution is preferred in this case. For the remaining three residues a similar comparison of RDC data does not identify a preference between the NMR and X-ray distributions.

Combined analysis of torsion angle distributions and order parameters leads to a number of conclusions. For example, high-order parameters and a significant population of only one of the allowed torsion angles implies limited dynamics over a wide range of time-scales extending from picoseconds to hundreds of microseconds. In cases where low order parameters are measured yet only a single side-chain conformation is predicted (for example, Val 107), motion is confined to the backbone or to dynamics within a well but does not involve jumps between different rotamers. The very poor agreement between predicted and measured RDC values for Val 107 $\gamma 2$ (Figure 5c,d; the corresponding data for $\gamma 1$ is not available) is likely the result of large range backbone dynamics in solution that are not captured in the X-ray structure. The combination of low S^2_{axis} values and multiple χ_1 states (for example, Val 87) implies more extensive motion involving rotamer hopping and likely dynamics within wells or involving the backbone. Finally, it is noteworthy that while high order parameters generally correlate with significant populations of only one of the 'allowed' three torsion angles, low-order parameters do not necessarily imply significant torsion angle hopping. This is illustrated for Val 82 and 107 where $p_t \geq 95\%$, yet S^2_{axis} values of 0.34 and 0.14, respectively, are measured. Thus, picosecond−nanosecond time-scale motions that are reported by order parameters are not necessarily correlated with slower regime dynamics that are typically associated with dihedral angle rotations.

We have also calculated the overlap, *O* (eq 5), for the chemical shift and X-ray-based Val χ_1 distributions in $\alpha_7\alpha_7$, Figure 6 (red), as a function of residue number. For completeness we have also determined *O* values for Ile (green) and Leu (blue) χ_2 distributions that have been obtained either from chemical shifts (using eq 3 of ref 22 and eq 2 of ref 24) or from the 2.3 Å X-ray structure of $\alpha_7\beta_7\beta_7\alpha_7$.[28] High *O* values are, in general, observed except in some cases where residues are dynamic, as noted for Val as well, further establishing the utility of the proposed methodology, even in studies of very high molecular weight particles. As a final note, it is worth mentioning that we have performed analyses very similar to those described above using a poorer quality X-ray structure of the proteasome solved to 3.4 Å resolution.[29] Not surprisingly, the level of agreement between the rotamer distributions so obtained and those based on either the solution data or on the high-resolution 2.3 Å crystal structure is very poor. For example, in the 3.4 Å structure all but one of the Val χ_1 torsion angles are in the g+ conformation, with the exception of Val 89 that is in the trans state. Indeed, correct placement of Val side-chain rotamers in structures solved by crystallography generally requires a resolution of ~2 Å or better that can be challenging in applications involving very high molecular weight proteins. This emphasizes the importance of NMR-based approaches for studies of side-chain orientation in supramolecular systems and the complementarity between solution NMR and X-ray diffraction.

dx.doi.org/10.1021/ja2014532 |*J. Am. Chem. Soc.* 2011, 133, 8272–8281

CONCLUDING REMARKS

We have presented a new method for quantifying rotamer distributions of Val χ_1 dihedral angles from methyl $^{13}C^{\gamma 1,\gamma 2}$ chemical shift values. Central to the approach is a minimization procedure in which a set of 20 χ_1 angles is selected for each Val by comparing averages of the chemical shifts calculated directly from these χ_1 angles using DFT derived surfaces with the experimentally measured $^{13}C^{\gamma 1}$ and $^{13}C^{\gamma 2}$ values. Rotamer populations are then calculated from the best 'distribution'. The utility of the approach has been assessed by comparing the rotamer distributions obtained via chemical shifts with those generated from three-bond scalar couplings that were measured for Val residues from six proteins. Overall, the rotamer populations ($p_i \in [0{:}1]$) can be obtained to within an uncertainty of <0.2. A computer program that calculates distributions of χ_1 rotamers from input Val methyl chemical shift data and lists the fractional populations of each of the three rotameric states is available upon request or can be downloaded from the web (www.smb.ucl.ac.uk/hansen/).

The method presented here is based only on methyl chemical shifts. These data are easily acquired with high accuracy, even in applications involving proteins with molecular weights in the hundreds of kDa, such as for the $\alpha_7\alpha_7$ 1/2 proteasome, where other parameters cannot be measured. As such, this methodology opens up the possibility for obtaining side-chain rotatmeric distributions—and hence, side-chain dynamics information that is sensitive to a range of motional time-scales extending from picoseconds to hundreds of microseconds. A second important application, not considered here, relates to studies of side-chain structure and dynamics in excited proteins states. Relaxation dispersion experiments have been developed for measuring methyl ^{13}C chemical shifts of 'invisible' protein states,[19,20] and these chemical shifts can then be used to obtain information about rotamer distributions in methyl-containing side-chains in cases where, as for supramolecules, scalar coupling-based experiments are not feasible. The methodology described adds to a growing list of approaches for quantifying the structure and dynamics of proteins in solution from measured chemical shifts.

MATERIALS AND METHODS

Experimental Measurements. A U−^{15}N,^{13}C-labeled sample of maltose binding protein (1.0 mM in protein, 2 mM β-cyclodextrin, 20 mM sodium phosphate pH 7.2, 3 mM NaN_3, 100 μM EDTA, 0.1 mg/mL Pefabloc, 1 μg/μL pepstatin, and 10% D_2O) was prepared as described previously,[45] and the published Val $^{13}C^{\gamma 1,\gamma 2}$ methyl assignments were verified using standard triple resonance spectroscopy.[46−49] Stereospecific assignments of the Val prochiral methyl groups were obtained via the method of Neri.[25] Three-bond scalar couplings for quantifying Val χ_1 rotamers were measured using quantitative *J*-based pulse schemes described in detail in the literature.[34,50,51] Experiments were recorded on a 600 MHz spectrometer equipped with a room-temperature triple resonance probehead, 37 °C. Values of the measured three-bond scalar couplings, $\{^3J(C^{\gamma 1},N)$, $^3J(C^{\gamma 2},N)$, $^3J(C^{\gamma 1},CO)$, $^3J(C^{\gamma 2},CO)\}$, are listed in Supporting Information. Data for the other proteins used in this study used to compare distributions from chemical shifts and scalar couplings were obtained from the literature and are also listed in Supporting Information. Chemical shifts and RDC values for the 1/2 proteasome were taken from the literature.[18,52] Stereospecific assignments of the Val prochiral methyl groups of $\alpha_7\alpha_7$, obtained previously by the method of Neri,[25] were checked using a sample generated with precursor that labels ($^{13}CH_3$) only in the *Pro-S* Leu,Val methyl positions.[26,27]

Generation of Ensembles. Each Val χ_1 dihedral angle was modeled in terms of a set of 20 angles that is intended to take into account the range of motional properties that might be expected for this residue. It is assumed that all of the 20 'states' interconvert rapidly so that the effective chemical shift is given by an average over shifts calculated for all 20 conformers characterizing a single ensemble. The 20 χ_1 angles were chosen randomly from Gaussian distributions centered about $\chi_1 = \{63°, 174°, 296°\}$ and with a standard deviation of 9°, as observed for χ_1 angles in a set of high-resolution X-ray structures of proteins.[30,31] This was accomplished in the following way: For each χ_1 angle a random integer between 1 and 3 was generated. If the number 1 appeared, the χ_1 angle was chosen to be in the g+ state; if 2(3) appeared, a χ_1 angle in the t(g−) state was selected. In a second step the χ_1 angle was selected randomly from the appropriate Gaussian profile centered at 63° (g+), 174° (t), or 296° (g−), 9° standard deviation. As described in the text, the size of the ensemble was chosen to be 20 since with this number it is possible construct a random coil distribution with populations, $\overline{p}_{g+} \approx 0.05$, $\overline{p}_t \approx 0.75$, $\overline{p}_{g-} \approx 0.20$, as predicted from the average of all χ_1 angles calculated from a database of high-resolution X-ray structures.[30,31] In addition, the ensemble size is sufficiently small so that it is possible to achieve rapid convergence using the genetic algorithm described below.

In order to calculate the 'best' ensemble from a set of four three-bond scalar couplings $\{^3J(C^{\gamma 1},N)$, $^3J(C^{\gamma 2},N)$, $^3J(C^{\gamma 1},CO)$, $^3J(C^{\gamma 2},CO)\}$ or from $^{13}C^{\gamma 1,\gamma 2}$ chemical shifts, an energy function was constructed, comprising two contributions. The first term, E_{MF}, 'forces' dihedral angles to lie within the expected distributions for Val (based on X-ray data[30,31] and illustrated by the green vertical bars in Figure 1b), while the second term ensures that the χ_1 distribution satisfies either the experimental scalar coupling (E_J) or chemical shift data (E_δ). We write the net energy, *E*, that is to be minimized as,

$$E = E_{MF} + E_J \tag{6.1}$$

or

$$E = E_{MF} + E_\delta \tag{6.2}$$

depending on whether a distribution is selected that satisfies the input scalar couplings (eq 6.1) or chemical shifts (eq 6.2). In eqs 6,

$$E_{MF} = \sum_i -\log \sum_r \overline{p}_r \exp\left(-\frac{(\chi_i - \overline{\chi}_r)^2}{2\sigma_r^2}\right) \tag{7}$$

$$E_J = k_J \sum_j (J_{exp,j} - J_{calc,j})^2 \left(1 - \exp\left[-\frac{(J_{exp,j} - J_{calc,j})^2}{2\sigma_{exp,j}^2}\right]\right) \tag{8}$$

$$E_\delta = k_\delta \sum_{i=1,2} (\delta_{calc}(^{13}C^{\gamma i}) - \delta_{exp}(^{13}C^{\gamma i}))^2 \tag{9}$$

where $\overline{p}_r$ (r = g+, t, g−) is the average population of the three rotamer states predicted from the X-ray database,[31] ($\overline{p}_{g+} = 0.06$, $\overline{p}_t = 0.74$, $\overline{p}_{g-} = 0.20$) and σ_r is the rmsd of χ_1 within each of the three states (9°). Note that the E_{MF} bias ensures that slightly noncanonical orientations of χ_1 can be obtained if required by the experimental data, however, with a penalty. The sum in eq 8 is over the four three-bond scalar couplings, *j*, listed above, k_J is the force constant for weighting the E_J term (relative to E_{MF}; a value of 100 Hz^{-2} is used, calculated as described below), $J_{exp,j}$ is the *j*th experimentally measured scalar coupling value ($j \in \{^3J(C^{\gamma 1},N)$, $^3J(C^{\gamma 2},N)$, $^3J(C^{\gamma 1},CO)$, $^3J(C^{\gamma 2},CO)\}$), $J_{calc,j}$ is the *j*th scalar coupling calculated from the χ_1 ensemble using Karplus-curve parameters determined previously by Bax and co-workers,[34] and $\sigma_{exp,j}$

dx.doi.org/10.1021/ja2014532 |J. Am. Chem. Soc. 2011, 133, 8272–8281

is the experimental uncertainty of $J_{\mathrm{exp},j}$ or 0.2 Hz, whichever is the largest. The form of the energy function in eq 8 is such that the uncertainty of experimental data is taken into account (second term on the right-hand side) to produce a nearly flat-bottom potential.[53,54] Finally, in eq 9 $\delta_{\mathrm{calc}}(^{13}C^{\gamma i})$ is given by eq 3, and a force constant of $k_\delta = 32$ ppm^{-2} is used (see below).

Finding the 'Best' Ensemble. In order to minimize E of eq 6 we have used a genetic algorithm.[55] Two hundred ensembles, each consisting of 20 χ_1 angles, were generated randomly as described above, and an energy, E, was calculated for each (generation number 0). The ensembles were sorted according to energy $\{E(0), E(1), ..., E(199)\}$, and a so-called fitness (probability) $p_E(i)$ was calculated for each ensemble according to:

$$p_E(i) = \frac{1}{Z}\left(\frac{E(0)}{2E(i) - E(0)}\right) \tag{10}$$

$$Z = \sum_i \frac{E(0)}{2E(i) - E(0)} \tag{11}$$

Each pair of ensembles of the next generation (generation number 1) is created from two 'parent-ensembles' from the previous generation (generation 0), where the two parent-ensembles are picked on the basis of their fitness probabilities, p_E, such that the probability of being picked increases with fitness. With a probability of 25% the two parent-ensembles are kept and passed on to the next generation, whereas in 75% of the cases the parents are mixed (produce 'offspring'). Reproduction occurs in the following manner: An integer between 1 and 20 is chosen, and if $\{\chi^{\mathrm{p1}}{}_j\}_{j\le 20}$ and $\{\chi^{\mathrm{p2}}{}_j\}_{j\le 20}$ are the two parent-ensembles and an integer m is chosen, then the two offspring-ensembles are $\{\chi^{\mathrm{b1}}{}_j\}_{j\le 20} = \{\chi^{\mathrm{p1}}{}_1, ..., \chi^{\mathrm{p1}}{}_m, \chi^{\mathrm{p2}}{}_{m+1}, ..., \chi^{\mathrm{p2}}{}_{20}\}$ and $\{\chi^{\mathrm{b2}}{}_j\}_{j\le 20} = \{\chi^{\mathrm{p2}}{}_1, ..., \chi^{\mathrm{p2}}{}_m, \chi^{\mathrm{p1}}{}_{m+1}, ..., \chi^{\mathrm{p1}}{}_{20}\}$. Subsequently, mutations are introduced into the offspring with a probability of 2.5%; for 2.5% of the newly generated offspring-ensembles one of the 20 χ_1 angles is removed, and a new value that is generated randomly as described above is inserted. New generations are produced until no further improvement in the energy of the best ensemble is obtained (or until a maximum of 100 generations is reached), and the lowest-energy ensemble of the final generation is the 'best' solution and used in further analyses. This procedure is repeated five times so that the final ensemble consists of 20 × 5 = 100 χ_1 angles. The procedure described above is summarized succinctly as:

```
for run=1:5
  generate 200 random ensembles each of 20 members
      for generation=0:100
            assign energies;
            sort according to energy;
            generate offspring;
            mutate offspring;
      end
  take the lowest energy ensemble and save
end
```

Determination of Force Constant k_J and k_δ. The value of the force constant k_J was determined by a cross-validation approach.[56] Only three of the four three-bond scalar couplings were used, and the 'best' ensemble was calculated for a given value of k_J, as described above. The fourth scalar coupling was subsequently back-calculated from the generated ensemble and compared with the experimentally measured coupling constant. This procedure was repeated four times so as to systematically exclude each of the four scalar couplings once. An average rmsd between the back-calculated scalar couplings and the experimentally measured values was obtained and used as a cross validation criterion to determine k_J,

$$\langle \mathrm{rmsd} \rangle = \frac{1}{4}\sum_j \sqrt{\frac{1}{N}\sum_{\mathrm{Val}} (J_{\mathrm{bc},j} - J_{\mathrm{exp},j})^2} \tag{12}$$

The sum in eq 12 (Val) is over all the Val residues in the set of six proteins used in the present study, $J_{\mathrm{bc},j}$ is the scalar coupling back-calculated from the best ensemble that, in turn, is derived from the three other scalar couplings. A plot of $\langle \mathrm{rmsd} \rangle$ as a function of k_J is shown in Supporting Information, and the value of k_J that produces a minimum in this curve ($k_J \approx 100$ Hz^{-2}) is used in all further calculations. As described in Results and Discussion a 'J-reference-ensemble' was generated for each Val residue using $k_J = 100$ Hz^{-2} (and with all four scalar couplings as input); this reference ensemble was employed, among other things, to optimize k_δ, as described below.

The value of the force constant k_δ was obtained by a cross-validation method as well.[56] For a given value of k_δ the 'best' ensemble satisfying the input Val $^{13}C^{\gamma 1,\gamma 2}$ chemical shift data (eq 6.2) for a given residue was obtained as described above. Ensembles calculated in this manner as a function of k_δ were compared with the J-reference ensembles that were obtained for each Val solely from scalar coupling data. A figure plotting the rmsd between populations calculated from chemical shifts, $p_\delta(r)$, and from scalar couplings, $p_J(r)$, as a function of k_δ is shown in Supporting Information, illustrating that a force constant of k_δ = 32 ppm^{-2} provides the best agreement between the two methods. This value was used in further analyses.

■ ASSOCIATED CONTENT

Supporting Information. Two figures showing 'best' values for k_J and k_δ; description of DFT calculations; tables listing chemical shifts and three-bond methyl scalar couplings for Val residues in the six proteins used; tables of $\{p_{g+}, p_t, p_{g-}\}$ from chemical shifts or scalar couplings for each of the proteins considered; and the complete ref 11. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
d.hansen@ucl.ac.uk; kay@pound.med.utoronto.ca

■ ACKNOWLEDGMENT

We thank Dr. Ad Bax (NIH), Professor James Chou (Harvard Medical School), and Professor Rieko Ishima (University of Pittsburgh) for providing us with some of the experimentally measured three-bond scalar coupling and chemical shift data used here. This work was supported by the Biotechnology and Biological Sciences Research Council, BBSRC (D.F.H.) and by a grant from the Canadian Institutes of Health Research, CIHR (LEK). L.E.K. holds a Canada Research Chair in Biochemistry. D.F.H. is a BBSRC David Phillips Fellow.

■ REFERENCES

(1) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.
(2) Kern, D.; Eisenmesser, E. Z.; Wolf-Watz, M. *Methods Enzymol.* **2005**, *394*, 507–524.
(3) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.

(4) Lange, O. F.; Lakomek, N. A.; Farès, C.; Schröder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471–1475.
(5) Palmer, A. G.; Grey, M. J.; Wang, C. Y. *Methods Enzymol.* **2005**, *394*, 430–465.
(6) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.
(7) McConnell, H. M. *J. Chem. Phys.* **1958**, *28*, 430–431.
(8) De Gortari, I.; Portella, G.; Salvatella, X.; Bajaj, V. S.; van der Wel, P. C. A.; Yates, J. R.; Segall, M. D.; Pickard, C. J.; Payne, M. C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2010**, *132*, 5993–6000.
(9) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213–223.
(10) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215–240.
(11) Shen, Y.; *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685–4690.
(12) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.
(13) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Biomol. NMR* **2008**, *41*, 113–120.
(14) Vallurupalli, P.; Hansen, D. F.; Kay, L. E. *Proc. Natl. Acad. Sci. U. S.A.* **2008**, *105*, 11766–11771.
(15) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312–1316.
(16) Berjanskii, M. V.; Wishart, D. S. *J. Biomol. NMR* **2008**, *40*, 31–48.
(17) Ollerenshaw, J.; Tugarinov, V.; Kay, L. *Magn. Reson. Chem.* **2003**, *41*, 843–852.
(18) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445*, 618–22.
(19) Lundstrom, P.; Vallurupalli, P.; Religa, T. L.; Dahlquist, F. W.; Kay, L. E. *J. Biomol. NMR* **2007**, *38*, 79–88.
(20) Baldwin, A. J.; Religa, T. L.; Hansen, D. F.; Bouvignies, G.; Kay, L. E. *J. Am. Chem. Soc.* **2010**, *132*, 10992–10995.
(21) London, R. E.; Wingad, B. D.; Mueller, G. A. *J. Am. Chem. Soc.* **2008**, *130*, 11097–11105.
(22) Hansen, D. F.; Neudecker, P.; Kay, L. E. *J. Am. Chem. Soc.* **2010**, *132*, 7589–7591.
(23) Hansen, D. F.; Neudecker, P.; Vallurupalli, P.; Mulder, F. A. A.; Kay, L. E. *J. Am. Chem. Soc.* **2010**, *132*, 42–43.
(24) Mulder, F. A. A. *ChemBioChem* **2009**, *10*, 1477–1479.
(25) Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wüthrich, K. *Biochemistry* **1989**, *28*, 7510–7516.
(26) Ruschak, A. M.; Velyvis, A.; Kay, L. E. *J. Biomol. NMR* **2010**, *48*, 129–135.
(27) Gans, P.; Hamelin, O.; Sounier, R.; Ayala, I.; Durá, M. A.; Amero, C. D.; Noirclerc-Savoye, M.; Franzetti, B.; Plevin, M. J.; Boisbouvier, J. *Angew. Chem., Int. Ed.* **2010**, *49*, 1958–1962.
(28) Förster, A.; Masters, E. I.; Whitby, F. G.; Robinson, H.; Hill, C. P. *Mol. Cell* **2005**, *18*, 589–599.
(29) Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. *Science* **1995**, *268*, 533–539.
(30) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins* **2000**, *40*, 389–408.
(31) Butterfoss, G.; Richardson, J.; Hermans, J. *Acta Crystallogr., Sect. D* **2005**, *61*, 88–98.
(32) Tonelli, A.; Schilling, F.; Bovey, F. *J. Am. Chem. Soc.* **1984**, *106*, 1157–1158.
(33) Tonelli, A.; Schilling, F. *Macromolecules* **1981**, *14*, 74–76.
(34) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959–8966.
(35) Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. *Biochemistry* **1996**, *35*, 361–368.
(36) Millet, O.; Mittermaier, A.; Baker, D.; Kay, L. E. *J. Mol. Biol.* **2003**, *329*, 551–563.
(37) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.
(38) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894–13895.
(39) Mulder, F. A. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578–590.
(40) Baumeister, W.; Walz, J.; Zühl, F.; Seemüller, E. *Cell* **1998**, *92*, 367–380.
(41) Goldberg, A. L. *Nature* **2003**, *426*, 895–899.
(42) Pickart, C. M.; Cohen, R. E. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 177–187.
(43) Religa, T. L.; Sprangers, R.; Kay, L. E. *Science* **2010**, *328*, 98–102.
(44) Tugarinov, V.; Sprangers, R.; Kay, L. E. *J. Am. Chem. Soc.* **2007**, *129*, 1743–1750.
(45) Gardner, K. H.; Zhang, X.; Gehring, K.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 11738–11748.
(46) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 225–231.
(47) Kay, L. E.; Xu, G.-Y.; Singer, A. U.; Muhandiram, D. R.; Forman-Kay, J. D. *J. Magn. Res. Ser. B* **1993**, *101*, 333–337.
(48) Gardner, K. H.; Konrat, R.; Rosen, M. K.; Kay, L. E. *J. Biomol. NMR* **1996**, *8*, 351–356.
(49) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 6291–6293.
(50) Vuister, G.; Wang, A.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 5334–5335.
(51) Grzesiek, S.; Vuister, G. W.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 487–493.
(52) Sprangers, R.; Kay, L. E. *J. Am. Chem. Soc.* **2007**, *129*, 12668–12669.
(53) Hansen, D. F.; Led, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1738–1743.
(54) Donaldson, L. W.; Skrynnikov, N. R.; Choy, W. Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 9843–9847.
(55) Kim, S.; Orendt, A. M.; Ferraro, M. B.; Facelli, J. C. *J. Comput. Chem.* **2009**, *30*, 1973–1985.
(56) Clore, G.; Garrett, D. *J. Am. Chem. Soc.* **1999**, *121*, 9008–9012.