J|A|C|S
ARTICLES
Published on Web 05/08/2007

Methyl Rotation Barriers in Proteins from ^{2}H Relaxation Data. Implications for Protein Structure

Yi Xue,[†] Maria S. Pavlova,[†] Yaroslav E. Ryabov,[†] Bernd Reif,[‡] and Nikolai R. Skrynnikov*,[†]

Contribution from the Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette Indiana 47907-2084, and Forschunginstitut für Molekulare Pharmakologie (FMP), Robert-Rössle-Strasse 10, 13125 Berlin, Germany

Received January 10, 2007; E-mail: nikolai@purdue.edu

Abstract: Side-chain ^{2}H and backbone ^{15}N relaxation data have been collected at multiple temperatures in the samples of the SH3 domain from α-spectrin. Combined analyses of the data allowed for determination of the temperature-dependent correlation times τ_f characterizing fast methyl motion. Molecular dynamics simulations confirmed that τ_f are dominated by methyl rotation; the corresponding activation energies approximate methyl rotation barriers. For 33 methyl groups in the α-spectrin SH3 domain the average barrier height was thus determined to be 2.8 ± 0.9 kcal/mol. This value is deemed representative of the "fluid" hydrophobic protein core where some barriers are increased and others are lowered because of the contacts with surrounding atoms, but there is no local order that could produce systematically higher (lower) barriers. For comparison, the MD simulation predicts the average barrier of 3.1 kcal/mol (calculated via the potential of mean force) or 3.4−3.5 kcal/mol (rigid barriers after appropriate averaging over multiple MD snapshots). The latter result prompted us to investigate rigid methyl rotation barriers in a series of NMR structures from the Protein Databank. In most cases the barriers proved to be higher than expected, 4−6 kcal/mol. To a certain degree, this is caused by tight packing of the side chains in the NMR structures and stems from the structure calculation procedure where the coordinates are first annealed toward the temperature of 0 K and then subjected to energy minimization. In several cases the barriers >10 kcal/mol are indicative of van der Waals violations. The notable exceptions are (i) the structures solved using the GROMOS force field where tight methyl packing is avoided (3.0−3.6 kcal/mol) and (ii) the structure solved by means of the dynamic ensemble refinement method (Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Nature **2005**, *433*, 128) (3.5 kcal/mol). These results demonstrate that methyl rotation barriers, derived from the experiments that are traditionally associated with studies of protein dynamics, can be also used in the context of structural work. This is particularly interesting in view of the recent efforts to incorporate dynamics data in the process of protein structure elucidation.

Introduction

It has been long recognized that dynamic parameters of protein side-chain methyls provide a useful probe for protein studies. Both solid- and solution-state NMR have provided a wealth of information on methyl dynamics, and the data from these two sources are believed to be generally consistent.[1,2] In particular, a good agreement between solid- and solution-state data has been noted when (i) the sample is that of a globular protein where most methyl-bearing side chains are sequestered inside a sufficiently rigid scaffold and (ii) the solid sample is generously hydrated and studied at or near room temperature.

NMR studies of methyl dynamics in proteins involve a number of spin probes such as ^{2}H,[3,4] ^{13}C,[5,6] and ^{1}H.[5,7] Typically, the dynamics observed at the methyl site is dominated by the rapid rotation of methyl about its 3-fold symmetry axis. Accordingly, fast motion that manifests itself in spin relaxation mainly represents methyl rotation. Of note, the fast motion correlation time, τ_f, varies by roughly a factor of 5 from residue to residue, which underscores diverse dynamic behaviors of methyl-bearing side chains.

For the experimental study we have chosen a 62-residue SH3 domain from chicken α-spectrin (α-spc SH3). The structure of this protein has been characterized by X-ray crystallography[8,9] as well as solution-[10] and solid-state NMR.[11] We collected the

[†] Purdue University.
[‡] Forschunginstitut für Molekulare Pharmakologie (FMP).

(1) Sparks, S. W.; Cole, H. B. R.; Torchia, D. A.; Young, P. E. *Chem. Scr.* **1989**, *29A*, 31−38.
(2) Reif, B.; Xue, Y.; Agarwal, V.; Pavlova, M. S.; Hologne, M.; Diehl, A.; Ryabov, Y. E.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2006**, *128*, 12354−12355.
(3) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 386−389.
(4) Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. *J. Am. Chem. Soc.* **1995**, *117*, 11536−11544.
(5) Nicholson, L. K.; Kay, L. E.; Baldisseri, D. M.; Arango, J.; Young, P. E.; Bax, A.; Torchia, D. A. *Biochemistry* **1992**, *31*, 5253−5263.
(6) Ishima, R.; Petkova, A. P.; Louis, J. M.; Torchia, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 6164−6171.
(7) Zhang, X.; Sui, X. G.; Yang, D. W. *J. Am. Chem. Soc.* **2006**, *128*, 5073−5081.
(8) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851−855.

backbone ^{15}N and side-chain methyl ^{2}H relaxation data in two different samples of α-spc SH3 at four different temperatures. Using the spectral density mapping approach,[12] we identified a single side chain affected by intermediate time-scale (~1ns) local dynamics. All other methyl sites proved to be compatible with the simple Lipari−Szabo model parametrized by two variables: τ_f and the associated order parameter, $(^1/_9)S_{axis}^2$. The values of τ_f obtained at different temperatures were used to produce Arrhenius plots and determine activation energies, E_a^f, for 35 distinct methyl sites. The case can be made that E_a^f values approximate with sufficient accuracy the energy barriers in rotating methyl groups, ΔV.

We suggest that the average barrier height of ca. 2.8 kcal/mol, found in the experimental study of α-spc SH3, can be typical for other proteins as well. Indeed, methyls are mostly found in the protein's hydrophobic core, which has a rather irregular, amorphous structure. There is no long-range order in the core and, therefore, no reason to expect systematically higher (or lower) barriers such as sometimes occur in the crystals of small model compounds.

To interpret the experimental results, we turned to molecular dynamics simulations. A 25-ns trajectory of α-spc SH3 was recorded in explicit solvent at the temperature 20 °C using the program CHARMM.[13] Using these data, methyl barriers ΔV_{rigid} were calculated by forcibly rotating each methyl group in the otherwise rigid structure (MD snapshot). The results were subsequently averaged over a series of snapshots to produce $\overline{\Delta V_{rigid}}$ (technically, the averaging is applied to the transition rates rather than directly to the barriers). From this treatment it becomes clear that the barrier heights are highly sensitive to dynamic fluctuations in the protein structure.[14]

NMR structures consist of multiple conformers and, to some degree, are representative of dynamic fluctuations that take place in proteins.[15] Starting from this premise, we decided to calculate $\overline{\Delta V_{rigid}}$ for a number of NMR ensembles from Protein Databank. As it turns out, several structures produce the average barriers of ca. 3.5 kcal/mol. This is in line with expectations, given the approximate character of the calculations. On the other hand there is a group of structures with abnormally high barriers, in excess of 10 kcal/mol. This is generally attributed to overly tight side-chain packing. At least in part, this effect is brought about by the structure calculation procedure where the coordinates are annealed toward the temperature of 0 K and then energy-minimized. The structures calculated under this protocol acquire certain low-temperature character—in particular, outside atoms are often positioned in between methyl protons, thus obstructing methyl rotation.

Two categories of NMR structures stand out. One is a group of structures solved with the help of GROMOS.[16] This force field treats methyls as expanded spheres; consequently, the intercalation effects are avoided and realistic barriers, 3.0−3.6 kcal/mol are obtained. Even more significantly, reasonable barriers are found in the structure solved by means of the dynamic ensemble refinement approach.[17] This approach seeks to accurately reproduce the motional variability of protein structure. It is anticipated that methyl rotation barriers ΔV can be incorporated in this method as one of the targets for structure refinement.

Background

^{2}H Relaxation in Methyls. Deuterium is a universally attractive probe because its spin evolution is dominated by a single (quadrupolar) interaction. Furthermore, it turns out that the quadrupolar tensor of ^{2}H in protein methyl sites practically does not change from one residue to another.[18] Five relaxation rates that can be measured for a single ^{2}H spin[19,20] are expressed through three spectral densities, $J(0)$, $J(\omega_D)$, and $J(2\omega_D)$.

Torchia and Szabo described the model where the methyl group executes random jumps between the three equivalent sites.[21,22] In the case when the overall tumbling of the protein is isotropic, this model produces biexponential correlation functions. The respective spectral densities are identical to Lipari−Szabo spectral densities and can be expressed as[23]

$$J(\omega) = (1 - (^1/_9)S_{axis}^2)\frac{\tau}{1+\omega^2\tau^2} + (^1/_9)S_{axis}^2\frac{\tau_{rot}}{1+\omega^2\tau_{rot}^2} \quad (1)$$

where $\tau^{-1} = \tau_f^{-1} + \tau_{rot}^{-1}$, S_{axis}^2 represents the order parameter of the methyl 3-fold symmetry axis, τ_{rot} is the correlation time of the overall tumbling, and, for this particular model, $\tau_f = 1/(3k)$, where k is the exchange rate for the nearest-neighbor jumps occurring in methyls. The simple model eq 1 containing two fitting parameters, τ_f and S_{axis}^2, has been termed LS-2 model.[24]

A more sophisticated theory has been developed by Edholm and Blomberg.[25] Their model considers diffusion of a methyl group in a potential $V(\phi) = \Delta V_\phi(1 - \cos 3\phi)/2$. The correlation function in this case features a steep drop (caused by rapid harmonic oscillations within the potential well), followed by a region where the decay is due to transitions between the potential wells. In what follows, we illustrate this kind of correlation function as obtained from our molecular dynamics simulations. It has been noted that this model is experimentally indistinguishable from the three-site jump model[26] and, therefore, can be ultimately reduced to eq 1.

Side-Chain Rotameric Jumps. In addition to the fast spinning of the methyl group and overall molecular tumbling, ^{2}H spectral densities can also be influenced by other forms of

(9) Chevelkov, V.; Faelber, K.; Diehl, A.; Heinemann, U.; Oschkinat, H.; Reif, B. *J. Biomol. NMR* **2005**, *31*, 295−310.
(10) Blanco, F. J.; Ortiz, A. R.; Serrano, L. *J. Biomol. NMR* **1997**, *9*, 347−357.
(11) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98−102.
(12) Peng, J. W.; Wagner, G. *J. Magn. Reson.* **1992**, *98*, 308−332.
(13) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(14) Chatfield, D. C.; Augsten, A.; D'Cunha, C. *J. Biomol. NMR* **2004**, *29*, 377−385.
(15) Bonvin, A. M. J. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.
(16) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS manual*; Biomos bv. Biomolecular Software: Groningen, The Netherlands, 1987.
(17) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−132.
(18) Mittermaier, A.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 10608−10613.
(19) Jacobsen, J. P.; Bildsøe, H. K.; Schaumburg, K. *J. Magn. Reson.* **1976**, *23*, 153−164.
(20) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439−6448.
(21) Torchia, D. A.; Szabo, A. *J. Magn. Reson.* **1982**, *49*, 107−121.
(22) Wallach, D. *J. Chem. Phys.* **1967**, *47*, 5258−5268.
(23) Kay, L. E.; Torchia, D. A. *J. Magn. Reson.* **1991**, *95*, 536−547.
(24) Skrynnikov, N. R.; Millet, O.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6449−6460.
(25) Edholm, O.; Blomberg, C. *Chem. Phys.* **1979**, *42*, 449−464.
(26) Batchelder, L. S.; Niu, C. H.; Torchia, D. A. *J. Am. Chem. Soc.* **1983**, *105*, 2228−2231.

internal motion. For longer side chains, there are fast axial fluctuations and 3-fold rotameric jumps involving torsional angles χ_i. Rotameric jumps in methyl-bearing side chains were first identified by line-shape analyses of ^{2}H solid-state NMR spectra.[3,27,28] Later, solution-state studies focusing on scalar and dipolar couplings confirmed that a significant fraction of protein side chains are conformationally disordered.[29−33]

The characteristic rates of the side-chain rotameric jumps, τ_χ, vary widely. In some cases, rotameric jumps do not influence ^{2}H relaxation rates, suggesting that $\tau_\chi \gg \tau_{rot}$. In other cases, the jumps occur in the low nanosecond range, $\tau_\chi \approx \tau_{rot}$, and can be directly identified in the relaxation data.[5,24,34] It turns out that a relatively small fraction, ca. 10%, of all side chains fall in this category. The data analysis can be adapted to account for the effect of these jumps.[24] Alternatively, the affected residues can simply be removed from consideration. Finally, there are residues where the jumps occur on a relatively fast time scale, $\tau_\chi \ll \tau_{rot}$. Fast jumps are responsible for low order parameters S_{axis}^2 that are quite common in leucine side chains; these jumps have been also observed in molecular dynamics simulations.[31,35,36] The presence of fast rotameric jumps can complicate the data analyses since τ_f can no longer be identified with methyl rotation alone. In what follows, we use the results of molecular dynamics simulations to probe the relationship between τ_f and the methyl rotation correlation time τ_ϕ.

Methyl Tunneling. Since methyl rotation involves relatively light particles (protons or deuterons), tunneling is possible. This effect has been a subject of many NMR studies conducted at low temperatures.[37] How important is methyl tunneling near room temperature? If the effect is significant, then molecular dynamics data should be corrected to reproduce correct τ_ϕ.[38]

The problem in question has been treated by Kowalewski and Liljefors.[39] The calculations employing two different models showed that Arrhenius activation energies for methyl rotameric jumps E_a^ϕ are nearly temperature-independent in the range 200−333 K. This is in good agreement with a large body of experimental evidence (see below). Most important, E_a^ϕ was found to be nearly identical to the height of the potential barrier which controls methyl rotation, ΔV (the barrier is defined as the difference between the maximum and the minimum of the potential energy). The deviations between the two quantities amount to less than 6% (less than 0.2 kcal/mol), which includes approximately 2% tunneling correction. Recent work of Latanowicz suggests that the tunneling corrections should disappear altogether near room temperature where our measurements have been carried out.[40,41]

Prior Data on Activation Energies and Energy Barriers. The activation energy for methyl reorientation in CH_3-CH_3 is 3.0 kcal/mol.[42] The same value, 3.0 kcal/mol, has been obtained for the barrier height in $CH_3-(CH_2)_2-CO-(CH_2)_2-CH_3$.[43] In a number of studies activation energies have been determined for individual amino acids, mainly by solid-state NMR.[26,27,44] Measurements have been also done on proteins selectively labeled with respect to certain types of amino acids.[27] The reported E_a^f values typically fall in the range 2−4 kcal/mol (one special situation is methionine where the barrier associated with the thioether bond is low). The solid-state measurements were usually performed in a wide-temperature range, in some cases spanning 200 K.[26,45] In essentially all cases, the Arrhenius plots are linear in the entire temperature range, consistent with the view of methyl rotation as an activated process.[41]

In certain crystals higher-than-expected activation energies have been found. The increased barriers were attributed to the crystal packing effects which obstruct methyl rotation. For example, E_a^f of 5.4 kcal/mol has been found in zwitterionic alanine. The crystallographic structure of this compound showed two methyl groups within 3.6 Å of each other, which is substantially less than the sum of van der Waals radii of methyls.[26] Later, calculations by Chatfield and co-workers confirmed that the increase in E_a^f caused by the crystalline environment is mainly due to the van der Waals term, with relatively modest contribution from electrostatic interaction.[38] In the absence of the crystal packing effects, the barriers of approximately 3 kcal/mol were predicted for alanines and leucines by DFT calculations in vacuo.[36,38]

Experimental Results

A single deuterium spin allows for measurement of five distinct relaxation rates.[19] Millet, Kay, and co-workers recently demonstrated that all these rates can be accurately measured in proteins, leading to self-consistent results.[20] After self-consistency of deuterium relaxation experiments has been established, one can take advantage of the redundancy that is present in the set of the rates. Since all five rates can be expressed through three spectral densities, $J(0)$, $J(\omega_D)$, and $J(2\omega_D)$, it is in principle sufficient to measure three rates out of five (preferably those for which higher sensitivity can be attained with the specific measurement scheme). We start by analyzing methyl ^{2}H $R_1(D_z)$, $R_2(D_x)$, and $R_{1zz}(D_z^2)$ rates measured by us in the sample of α-spc SH3 domain at 10 °C.[2]

The three rates were transformed into spectral densities,[12] and the results were fitted to the Lipari−Szabo LS-2 model, eq 1. In doing so, τ_f and S_f^2 were treated as fitting parameters, whereas τ_{rot} was fixed according to ^{15}N relaxation data measured on the same sample.[2] The analysis demonstrated that, aside from a single side chain, all methyl sites are in perfect compliance with the LS-2 model. Two top panels in Figure 1 show the typical spectral density profiles supplemented with the LS-2 best-fit curves. The sole exception is the side chain of Val 23 with its

(27) Keniry, M. A.; Kintanar, A.; Smith, R. L.; Gutowsky, H. S.; Oldfield, E. *Biochemistry* **1984**, *23*, 288−298.
(28) Colnago, L. A.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1987**, *26*, 847−854.
(29) Perez, C.; Lohr, F.; Ruterjans, H.; Schmidt, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 7081−7093.
(30) Mittermaier, A.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 6892−6903.
(31) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.
(32) Best, R. B.; Rutherford, T. J.; Freund, S. M. V.; Clarke, J. *Biochemistry* **2004**, *43*, 1145−1155.
(33) Tang, C.; Iwahara, J.; Clore, G. *J. Biomol. NMR* **2005**, *33*, 105−121.
(34) Millet, O.; Mittermaier, A.; Baker, D.; Kay, L. E. *J. Mol. Biol.* **2003**, *329*, 551−563.
(35) Best, R. B.; Clarke, J.; Karplus, M. *J. Mol. Biol.* **2005**, *349*, 185−203.
(36) Hu, H.; Hermans, J.; Lee, A. L. *J. Biomol. NMR* **2005**, *32*, 151−162.
(37) Horsewill, A. J. *Prog. NMR Spectrosc.* **1999**, *35*, 359−389.
(38) Chatfield, D. C.; Wong, S. E. *J. Phys. Chem. B* **2000**, *104*, 11342−11348.
(39) Kowalewski, J.; Liljefors, T. *Chem. Phys. Lett.* **1979**, *64*, 170−174.
(40) Latanowicz, L. *J. Phys. Chem. A* **2004**, *108*, 11172−11182.
(41) Latanowicz, L. *Concepts Magn. Reson. A* **2005**, *27A*, 38−53.
(42) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283−438.
(43) Horsewill, A. J.; Green, R. M.; Alsanoosi, A. M. Tunnelling spectroscopy and dynamics of CH3 in a homologous series of ketones. In *Springer Proceedings in Physics;* Springer: Berlin, 1987; Vol. 17, pp 28−39.
(44) Beshah, K.; Griffin, R. G. *J. Magn. Reson.* **1989**, *84*, 268−274.
(45) Copié, V.; McDermott, A. E.; Beshah, K.; Williams, J. C.; Spijker-Assink, M.; Gebhard, R.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **1994**, *33*, 3280−3286.

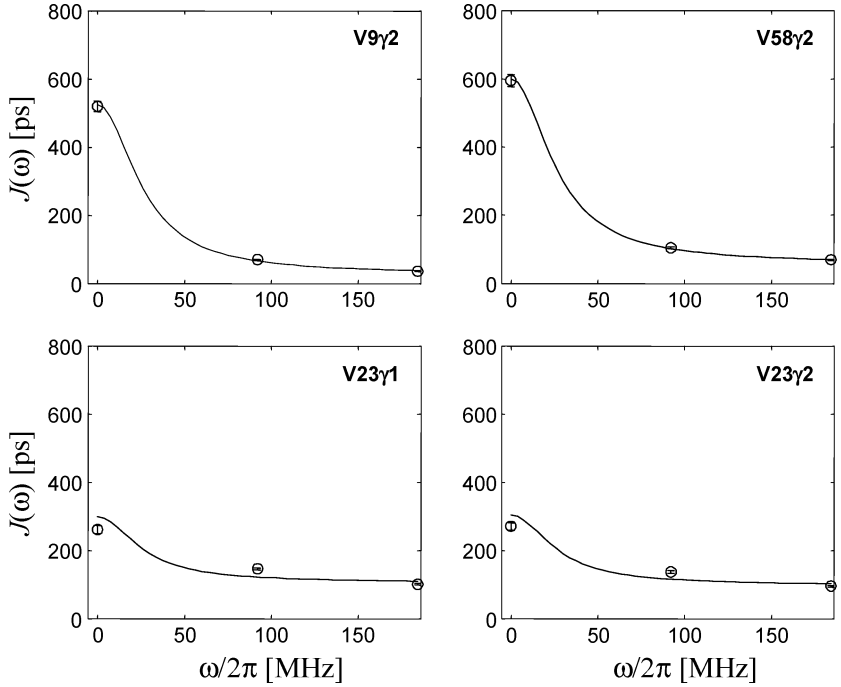

Figure 1. Spectral density profiles for selected methyl sites in α-spc SH3. Two top panels represent the typical quality of the data fitting using the LS-2 model (as determined on the basis of the fit residual).[2]

two methyl sites (two bottom panels in Figure 1). Previously, the conformational disorder in this side chain has been detected using X-ray crystallography[9] and solid-state NMR.[46] We have also demonstrated that the present set of relaxation data can be explained if one assumes that both methyl groups in Val 23 are affected, to the same degree, by χ_1 rotameric jumps on the time scale of several nanoseconds.[2] In the context of this study, however, we merely observe that the LS-2 model fails to fit the data for Val 23, and choose to exclude this residue from further consideration.

To obtain the activation energies associated with τ_f, we repeated the measurements of ^{2}H R_1 and R_2, as well as ^{15}N R_1 and R_2, at the field of 600 MHz and temperatures of 17, 24, and 30 °C. To ascertain that the separation of internal dynamics and overall rotation is achieved correctly, all measurements were duplicated using two different samples of α-spc SH3. The first sample was expressed using u-^{13}C glucose as a carbon source, permitting detection of all methyl sites. The second sample was prepared with 3-^{13}C pyruvate, limiting observations to Val, Leu, and Ile-γ sites. Ultimately, the data from both samples were available for 26 methyl sites, the data from the glucose-derived sample only were available for 7 sites, and the data from pyruvate sample only were at hand for the additional 2 sites (in the latter case, the glucose sample data were unusable because of the spectral overlap). Because of the difference in α-spc SH3 concentration and, to some extent, the difference in deuteration levels, the two samples turned out to be systematically different with respect to reorientational motion of the protein. Specifically, the glucose-derived sample produced τ_{rot} values that were ca. 10% higher throughout the series of measurements. This did not interfere, however, with the analyses of methyl dynamics because τ_{rot} in each case has been determined independently from ^{15}N data (see Materials and Methods).

The data for each methyl site at each temperature were fitted to the LS-2 model. Note that in doing so we set aside the R_{1zz} data measured at 10 °C and restricted the analyses to R_1 and R_2.[47] To estimate the errors in τ_f and S_{axis}^2, the Monte Carlo procedure was implemented as a part of the LS-2 fitting routine. The resulting τ_f values were used to produce the Arrhenius plots. Since the data from the two samples proved to be fully consistent, they were fitted collectively to obtain the activation energies, E_a^f. Three representative fits are shown in Figure 2 (the choice of residues is explained in the figure caption).

Methyl rotation barriers in proteins were a subject of recent studies based on the results of MD simulations.[14,35] The experimental component of these studies, however, was represented by the τ_f data. The activation energies, E_a^f, determined in this work provide much more suitable material for this type of analysis. Hence, we turn to MD simulations in attempt to re-examine the nature of methyl rotation barriers in proteins and to bridge the dynamic and structural aspects of the problem.

Comparison with MD Data

Correlation Times. Figure 3 shows the correlation functions for the methyl ^{2}H quadrupolar interaction (presumed axially symmetric, with the principal axis along carbon-proton bond). The correlation functions are extracted from the 25 ns MD trajectory of α-spc SH3; only the initial part of each function is presented in the plot. The three methyl sites selected for this figure are the same as in Figure 2.

The full correlation functions shown with solid lines in Figure 3 are, in fact, more complex than assumed in the LS-2 model. The MD-generated correlation function begins with a steep (sub-

(46) Hologne, M.; Faelber, K.; Diehl, A.; Reif, B. *J. Am. Chem. Soc.* **2005**, *127*, 11208−11209.

(47) Quadrupolar order relaxation experiments can suffer somewhat from strong ^{13}C-^{13}C scalar coupling effects in Leu side chains: Ollerenshaw, J. E.; Xue, Y.; Pavlova, M. S.; Kay, L. E.; Skrynnikov, N. R. Unpublished work.

(48) Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1992.

(49) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

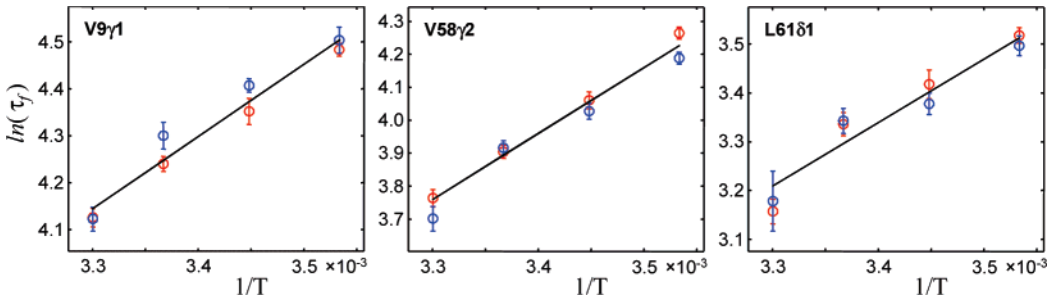

Figure 2. Arrhenius plots for three representative methyl groups in α-spc SH3. The values of τ_f are from the glucose- and pyruvate-derived samples (red and blue symbols, respectively). Linear fits were generated by minimizing the residual where the contribution from each point was weighted in inverse proportion to the error (i.e., to the size of the error bar).[48] The uncertainties in E_a^f were evaluated with the help of the jackknife procedure using reduced data sets consisting of 6 points. The results for all methyl groups were consequently sorted according to the magnitude of uncertainty in E_a^f; the three entries from the middle of the sorted list are presented in this figure. For these three methyls, left to right, the determined activation energies and the uncertainty intervals are 3.1 [2.7−3.6], 4.0 [3.5−4.4], and 2.6 [2.0−3.0] kcal/mol. The same type of treatment was applied to 9 methyl sites for which only the data from one of the two samples are available (not shown). In this case, the jackknife procedure involved reduced data sets consisting of three points. In what follows, these methyls are identified in the plots using distinct symbols. The values of τ_f presented in this figure are derived under the assumption that rotational diffusion of α-spc SH3 is isotropic (the same assumption is later used in the analysis of MD simulation data). This approximation is well justified since (i) as indicated by our ^{15}N relaxation data, the anisotropy of α-spc SH3 is modest, $D_{\parallel}/D_{\perp} = 1.22 \pm 0.02$, (ii) methyl relaxation rates are dominated by fast dynamics rather than the overall tumbling, and (iii) longer side chains tend to sample multiple conformations which leads to partial averaging of the anisotropic tumbling effects. To validate these assumptions, the analysis was repeated using the version of the LS-2 model which incorporates the anisotropic tumbling correlation function.[49] The rms deviation between the two sets of E_a^f values, derived with the use of isotropic and anisotropic models, was found to be 0.15 kcal/mol (well below the level of experimental uncertainty), and the bias did not exceed 0.05 kcal/mol.

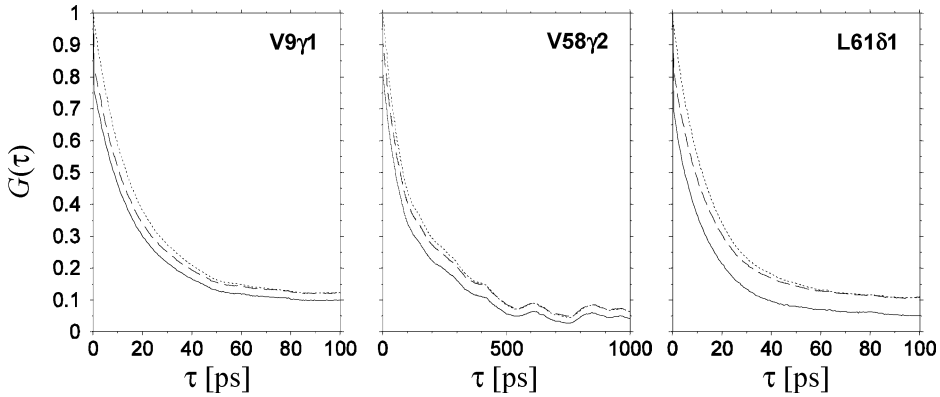

Figure 3. MD-based correlation functions for selected residues in α-spc SH3. Solid lines represent MD data that are used as is, with a standard processing scheme (see Materials and Methods). The correlation functions are computed according to $G(\tau) = \langle P_2(\cos\theta(\tau))\rangle$, where $P_2(x)$ is the second-order Legendre polynomial and $\theta(\tau)$ is the angle between the orientations of the methyl proton-carbon bond at two points in time separated by the interval τ. The rotational correlation time is taken to be $\tau_{rot} = 5.0$ ns (obtained from interpolation of the experimental data). Dashed lines represent MD data that are manipulated prior to calculation of $G(\tau)$. Specifically, the torsional angles $\phi(t)$ parametrizing methyl rotation are extracted from the MD data. The vector trajectories are then regenerated assuming that rotation of the (ideal tetrahedral) methyl group according to $\phi(t)$ is the only form of dynamics in the system. Dotted lines represent further simplification where the CH_3 rotation is reduced to a sequence of rotameric jumps. Specifically, all $\phi(t)$ values in the range from 0° to 120° are reset to 60°, all $\phi(t)$ values between 120° and 240° are reset to 180°, and all $\phi(t)$ values between 240° to 360° are reset to 300°.[38] The vector trajectories are then regenerated assuming that rotameric jumps is the only form of motion present. Small displacements of actual potential wells from the canonical values, 60°, 180°, and 300°, were found to be inconsequential in the context of these calculations.

ps time scale) drop which is mainly caused by harmonic axial fluctuations experienced by the methyl. This is followed by a period of decay with the time constant of 20−100 ps, which is dominated by methyl spinning. The remaining part of the correlation function, which falls outside the plotted region, decays mainly through the overall molecular rotation. This behavior has been quantitatively predicted in the model by Edholm and Blomberg.[25,50] The model-free parametrization of such correlation function has been introduced by Chatfield and co-workers.[51]

In addition to the full MD-based correlation function, Figure 3 also shows two reduced versions of $G(\tau)$. In one of them (dashed line) only the methyl rotation is retained. In the other one (dotted line) only the rotameric jumps between the three symmetric wells are included. Clearly, there is an appreciable difference between the actual correlation function and the reduced versions. Specifically, we are concerned about the difference between the full correlation function, which is relevant for the analyses of relaxation data, and the reduced correlation function, which is relevant for characterization of the methyl rotation barrier (continuous vs dotted line). The MD data provide a good opportunity to quantify this difference.

With this goal in mind, we employed the full correlation functions to simulate the ^{2}H R_1 and R_2 relaxation rates for all methyl sites in α-spc SH3. These simulated rates were subsequently fitted with the LS-2 model. As a result, the set of MD-based τ_f values has been obtained that mirrors the experimentally determined τ_f. Separately, we fitted the reduced correlation functions (dotted lines) with a combination of two exponentials and thus obtained the values of τ_ϕ. The comparison between τ_ϕ

(50) Daragan, V. A.; Mayo, K. H. *Prog. NMR Spectrosc.* **1997**, *31*, 63−105.
(51) Chatfield, D. C.; Augsten, A.; D'Cunha, C.; Wong, S. E. *J. Comput. Chem.* **2003**, *24*, 1052−1058.

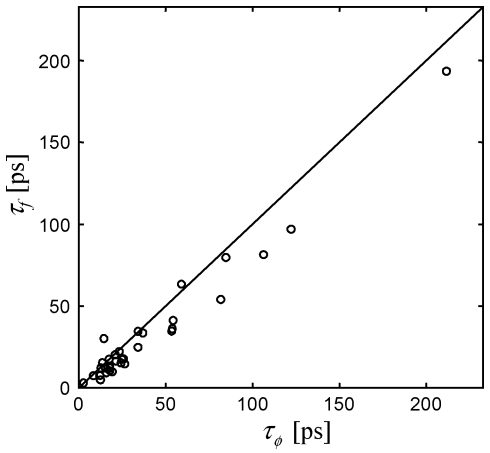

Figure 4. Comparison of the two methyl correlation times extracted from the MD trajectory for multiple methyl sites in the α-spc SH3 domain. To obtain τ_f, the full correlation function $G(\tau)$ was used to simulate ^{2}H R_1 and R_2 relaxation rates and these rates were subsequently fitted with the LS-2 model, eq 1. To obtain τ_ϕ, the reduced correlation function describing methyl rotameric jumps (see caption of Figure 3) was fitted with a biexponential curve. The correlation coefficient for the data in the plot is $r = 0.98$. Two points that fall above the diagonal corresponds to the residues affected by the nanosecond time scale local dynamics, $\tau_\chi \approx \tau_{rot}$. Recall that in the treatment of experimental data such points are excluded. A complementary plot, Figure S1, correlates the MD-based τ_f values with the experimentally determined τ_f values.

and τ_f is illustrated in Figure 4. As expected, τ_f values turn out systematically shorter, on average by 18%, since they absorb the effect of very fast fluctuations.

While τ_f itself may not be a very accurate measure of methyl rotation, the corresponding activation energy, E_a^f, provides a reasonable approximation for E_a^ϕ. To see why that is so, let us turn again to the analysis of full correlation functions, Figure 3. The harmonic motions responsible for the initial drop in $G(\tau)$ can be parametrized in terms of additional order parameter, S_{uf}^2, and correlation time, τ_{uf} (the subscript stands for 'ultrafast'). Since τ_{uf} is very short, less than 1 ps, this motional process is not relaxation-active. Hence, the spectral density corresponding to $G(\tau)$ can be rewritten as

$$J(\omega) = (S_{uf}^2 - (^1/_9)S_{axis}^2)\frac{\tau}{1+\omega^2\tau^2} + (^1/_9)S_{axis}^2\frac{\tau_{rot}}{1+\omega^2\tau_{rot}^2} \quad (2)$$

where $\tau^{-1} = \tau_f^{-1} + \tau_{rot}^{-1}$. The result in eq 2 can be viewed as "true" spectral density, which is somewhat different from the LS-2 model, eq 1.

First of all, we established that even the most comprehensive set of ^{2}H relaxation data[24] does not allow for separation of S_{uf}^2 and τ_f based on eq 2. This can be readily understood because the first term in eq 2 falls in the extreme narrowing limit, $\omega_D\tau \ll 1$, which makes the separation impossible. It is also important to keep in mind that S_{uf}^2 is *not* subsumed in the quadrupolar coupling constant (considering specifically the reported solid-[52] and solution-state[18] experiments).

To estimate the effect of S_{uf}^2 on the determined activation energies, we resorted to further simulations. In brief, eq 2 was used to simulate the ^{2}H R_1 and R_2 relaxation rates at four different temperatures, same as employed in the experimental study, and the results were then fitted by the LS-2 model and used to determine the activation energy. Not surprisingly, the activation energy was reproduced correctly when S_{uf}^2 was assumed to be temperature-independent, but it was compromised when S_{uf}^2 was allowed to vary with temperature.

To quantitate the temperature dependence of S_{uf}^2, we recorded three 1-ns long trajectories of α-spc SH3 at the temperatures of 10, 20, and 30 °C (see Materials and Methods). From these trajectories we estimated that, on average, S_{uf}^2 changes by 0.0009 units per Kelvin, consistent with the general notion that the temperature dependence of the fast-motion order parameters is rather subtle.[53] On the basis of this result we further estimated that the apparent activation energy differs from the target activation energy by ca. 7%.

While the results of MD-based analyses should be regarded with caution, the general observations made in this section are sufficiently intuitive: (i) Relaxation-based correlation times τ_f underestimate methyl rotation correlation times τ_ϕ. The reason for this is the presence of the ultrafast axial fluctuations, particularly those that involve the methyl group itself. (ii) Apparent activation energies E_a^f somewhat overestimate activation energies of methyl rotation E_a^ϕ. This happens because the amplitudes of ultrafast motions are temperature-dependent.

Energy Barriers. The conventional method of evaluating the barriers is via the potential of mean force, $V_{pmf}(\phi) = -RT \ln p(\phi)$, where $p(\phi)$ is the probability distribution for the torsional angle controlling methyl rotation, as obtained from molecular dynamics. The energy profiles generated in this manner for three methyl groups represented in Figure 2 are shown in Figure S2. The procedure for evaluating the barriers, ΔV_{pmf}, is the same as described by Chatfield and Wong.[38]

In addition, we calculate the so-called rigid barriers. In doing so a single structure (MD snapshot) is loaded in XPLOR[54] and a selected methyl group is rotated, while maintaining the rest of the structure fixed. The resulting energy profile is symmetrized with respect to three protons, $V_{rigid}(\phi) = (V_{rigid}(\phi) + V_{rigid}(\phi + 120°) + V_{rigid}(\phi + 240°))/3$, and the barrier ΔV_{rigid} is evaluated. The calculation is carried out for multiple snapshots producing a series of values $\Delta V_{rigid}^{(i)}$ (where index i enumerates the snapshots), and the results are averaged as follows:

$$\overline{\Delta V_{rigid}} = (-RT)\ln\left(k_0^{-1}\frac{1}{N}\sum_{i=1}^{N} k_0 \exp(-\Delta V_{rigid}^{(i)}/RT)\right) \quad (3)$$

As can be appreciated from eq 3, the averaging is applied to rates, rather than directly to barriers, in accordance with the principles of chemical kinetics. In what follows this procedure would be referred to as "log-averaging". For a large set of MD snapshots the result is representative of the equilibrium distribution of conformational species. (For a small set, $\overline{\Delta V_{rigid}}$ can be in principle calculated as energy-weighted average. However, the total energy associated with the ith snapshot only very weakly depends on the state of the specific methyl group and its environment, which makes this approach impractical).

In this study, we made no attempt to calculate adiabatic barriers. The definition of the adiabatic barrier can be rather

(52) Burnett, L. J.; Muller, B. H. *J. Chem. Phys.* **1971**, *55*, 5829−5831.
(53) Vugmeyster, L.; Trott, O.; McKnight, C. J.; Raleigh, D. P.; Palmer, A. G. *J. Mol. Biol.* **2002**, *320*, 841−854.
(54) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65−73.

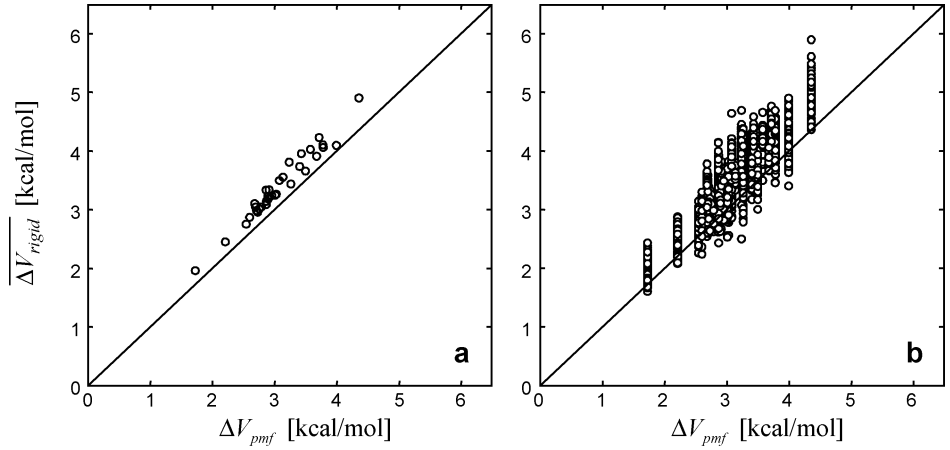

Figure 5. Methyl rotation barriers, ΔV_{pmf} vs $\overline{\Delta V_{rigid}}$, as derived from a 25-ns trajectory of the α-spc SH3 domain. Each point represents a specific methyl group: (a) $\overline{\Delta V_{rigid}}$ averaged over 10 000 MD snapshots that are uniformly distributed over the length of the trajectory (interval 2.5 ps); (b) $\overline{\Delta V_{rigid}}$ averaged over 20 MD snapshots which are extracted from the trajectory at random; the calculations are repeated for 100 randomly formed "ensembles".

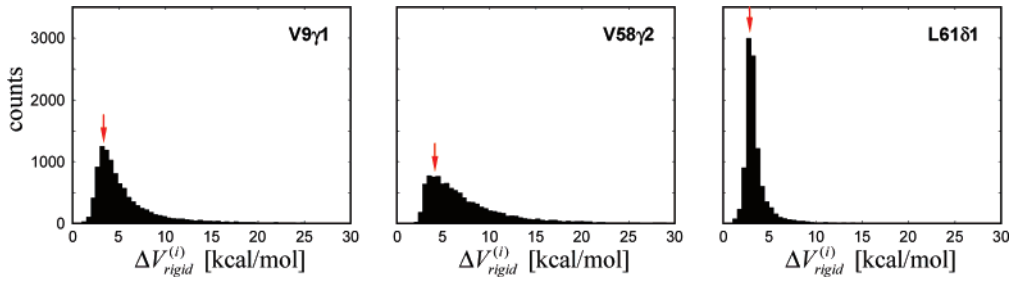

Figure 6. Histograms of $\Delta V^{(i)}_{rigid}$ for three selected methyl groups based on 10 000 snapshots from the MD trajectory of α-spc SH3. Indicated by the arrows are the corresponding log-averaged values $\overline{\Delta V_{rigid}}$.

arbitrary[14] since it is not a priori clear to what extent the environment is relaxed in response to methyl rotation. In fact, adiabatic barriers sometimes prove to be an inferior approximation.[38] Ultimately, we favor the log-averaged rigid barriers which provide a straightforward way of dealing with the ensembles of protein structures and do not involve any adjustable parameters (apart from the standard force field settings). As discussed in more detail below, log-averaged rigid barriers can be viewed as a structural concept, whereas adiabatic barriers gravitate toward the realm of molecular dynamics.

To assess the quality of the rigid barrier approximation, we compared $\overline{\Delta V_{rigid}}$ with ΔV_{pmf}. For this purpose, $\overline{\Delta V_{rigid}}$ was calculated using a set of 10 000 regularly spaced MD snapshots. The results of the comparison for 33 methyl groups in the α-spc SH3 domain are presented in Figure 5a. As it turns out, the two quantities are tightly correlated, $r = 0.98$, with $\overline{\Delta V_{rigid}}$ on average by 0.32 kcal/mol higher than ΔV_{pmf}. We suggest, therefore, that $\overline{\Delta V_{rigid}}$ provides a simple and reasonably accurate instrument for evaluating methyl rotation barriers.

One property of the rigid barriers is worth noting: the magnitude of $\overline{\Delta V_{rigid}}$ is primarily determined by the snapshots with low barriers $\Delta V^{(i)}_{rigid}$ at the expense of the snapshots with high barriers $\Delta V^{(i)}_{rigid}$. This is a trivial consequence of log-averaging. This property is illustrated in Figure 6, which shows the histograms of $\Delta V^{(i)}_{rigid}$ for three selected methyl groups. The arrows in this plot mark the corresponding values of $\overline{\Delta V_{rigid}}$. It can be readily appreciated from the plot that the averages are skewed. In general terms, it can be stated that $\overline{\Delta V_{rigid}}$ are disproportionately sensitive to the "excited" states (i.e., rare conformations where the barriers happen to be low).[14]

In view of these observations, the question arises as to whether $\overline{\Delta V_{rigid}}$ is still a robust parameter when calculated over a smaller set of structures. To test that, we prepared one hundred small sets, each consisting of 20 randomly selected MD snapshots. The size of 20 has been chosen because it is typical for protein structures solved by NMR. The values of $\overline{\Delta V_{rigid}}$ obtained from these calculations are presented in Figure 5b. As expected, substantial uncertainty emerges as the size of the ensemble is reduced. The mean values, however, remain reasonable. For instance, if the results are averaged over all methyl groups within each ensemble, the obtained $\langle\overline{\Delta V_{rigid}}\rangle$ are only 0.30−0.47 kcal/mol higher than ΔV_{pmf} (a modest and nearly constant offset). Thus, even for small dynamic ensembles $\langle\overline{\Delta V_{rigid}}\rangle$ proves to be a good indicator of the mean barrier height.

Comparison with Experimental Data. Figure 7 shows the comparison between the MD-derived barriers, ΔV_{pmf}, and the experimentally determined activation energies, E_a^f, for methyl groups in the α-spc SH3 structure. Clearly, there is no discernible correlation on per-site basis ($r = 0.48$). First, it should be noted that substantial experimental uncertainty makes it difficult to establish a correlation. In future studies, this problem can be alleviated by using a wider temperature range (in particular, by employing thermophilic proteins[55] and/or

(55) Butterwick, J. A.; Loria, J. P.; Astrof, N. S.; Kroenke, C. D.; Cole, R.; Rance, M.; Palmer, A. G. *J. Mol. Biol.* **2004**, *339*, 855−871.

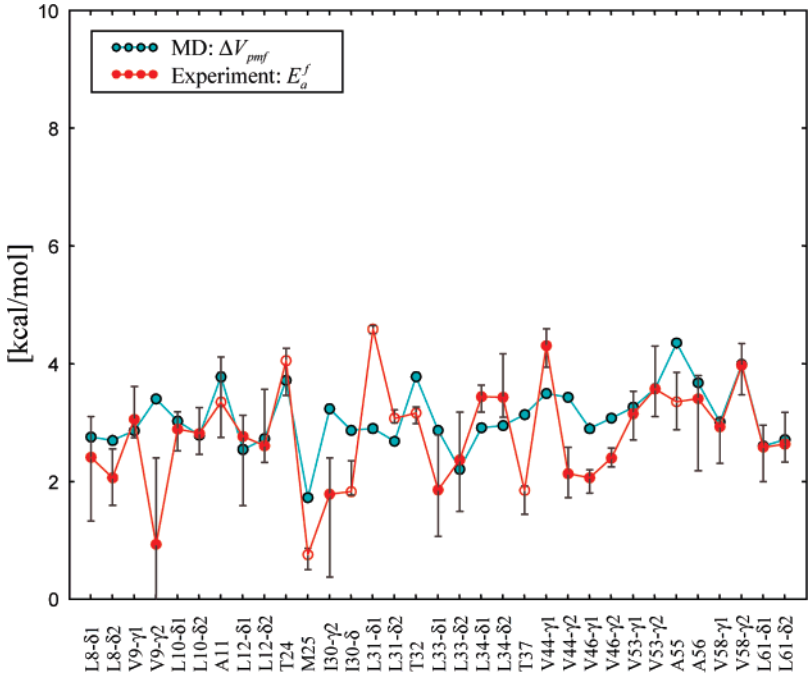

Figure 7. Methyl rotation barriers as characterized by the experimental data (red) and molecular dynamics simulations (turquoise). The E_a^f values based on the experimental data from only one of the two samples are indicated with open red circles. The average over all methyl groups is $\langle E_a^f \rangle = 2.8$ kcal/mol; considering the estimates made in the previous section, the activation energy of the methyl rotameric jumps may be slightly lower, ca. 2.6 kcal/mol.

conducting measurements in supercooled water[56]). The differences between E_a^f and E_a^ϕ discussed above may also contribute to the site-specific variations. Most important, however, is the apparent inability of molecular dynamics to deliver the accuracy required for per-residue comparison with the experimental data. In recent studies, little or no correlation has been observed between the simulated and experimental methyl τ_f and even between the experimental and simulated S_{axis}^2.[14,35,57]

While MD fails to reproduce the per-residue variations, Figure 7, it reproduces the average value reasonably well: $\langle E_a^f \rangle = 2.8$ kcal/mol versus $\langle \Delta V_{pmf} \rangle = 3.1$ kcal/mol. The perturbations caused by the environment apparently create some lower and some higher barriers, but on balance have little effect. Indeed, the protein hydrophobic core is loosely packed and there is no local order that could potentially lead to systematically lower or systematically higher barriers (such as, for example, in crystalline alanine). In what follows we assume that 2.5−3 kcal/mol is a typical barrier height not only in the α-spc SH3 domain, but in other proteins as well. While this hypothesis is somewhat speculative, it draws upon the general notion of the fluid hydrophobic core.[32,58]

Implications for Protein Structure

In the previous section we discussed the rigid barrier calculations in structural ensembles. It has been found that ΔV_{rigid}, after log-averaging over the members of the ensemble and then averaging over all methyl groups in the protein, produces a reasonable $\langle \overline{\Delta V_{rigid}} \rangle$ value. Predictably, this value is somewhat higher than the outcome of the rigorous MD analyses and the corresponding experimental result. This observation holds also for smaller ensembles that have the size of typical NMR structure.

It has been generally recognized that NMR structures encode some information about internal protein motions and, to a certain degree, can be viewed as dynamic ensembles.[15] Recently, the efforts were made to construct structural ensembles that would accurately reproduce dynamic properties of proteins.[17,59−61] From this perspective it is interesting to see whether the rigid barriers calculated in NMR ensembles match the expected values.

Figure 8a shows the $\overline{\Delta V_{rigid}}$ profile for the NMR structure of α-spc SH3 1AEY.[10] For the sake of comparison, the experimental E_a^f data are drawn in the same plot. Again, no significant correlation is observed between the two parameters on a per-residue basis ($r = 0.32$). For several methyl groups the barriers are clearly overestimated (note that in treating the ensembles of the MD conformers, Figure 5b, we have not encountered any barriers in excess of 6 kcal/mol). The average value, $\langle \overline{\Delta V_{rigid}} \rangle = 4.1$ kcal/mol, is also appreciably higher than what is expected on the basis of the molecular dynamics tests. Clearly, the quality of the input coordinates could affect the outcome of these calculations. To explore this possibility, the calculations were repeated using a high-resolution (1.5 Å) X-ray structure 1U06,[9] Figure 8b. The resulting $\langle \Delta V_{rigid} \rangle = 3.6$ kcal/mol is more in line with expectations, given that the rigid-barrier calculation on a single conformer should inevitably lead to slightly exaggerated barriers. Of note, the X-ray structure with lower resolution (1SHG,[8] 1.8 Å) gives rise to $\langle \Delta V_{rigid} \rangle =$

(56) Skalicky, J. J.; Mills, J. L.; Sharma, S.; Szyperski, T. *J. Am. Chem. Soc.* **2001**, *123*, 388−397.
(57) Prabhu, N. V.; Lee, A. L.; Wand, A. J.; Sharp, K. A. *Biochemistry* **2003**, *42*, 562−570.
(58) Wong, K. B.; Daggett, V. *Biochemistry* **1998**, *37*, 11182−11192.
(59) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977−8991.
(60) Kitao, A.; Wagner, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2064−2068.
(61) Best, R. B.; Lindorff-Larsen, K.; DePristo, M. A.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10901−10906.

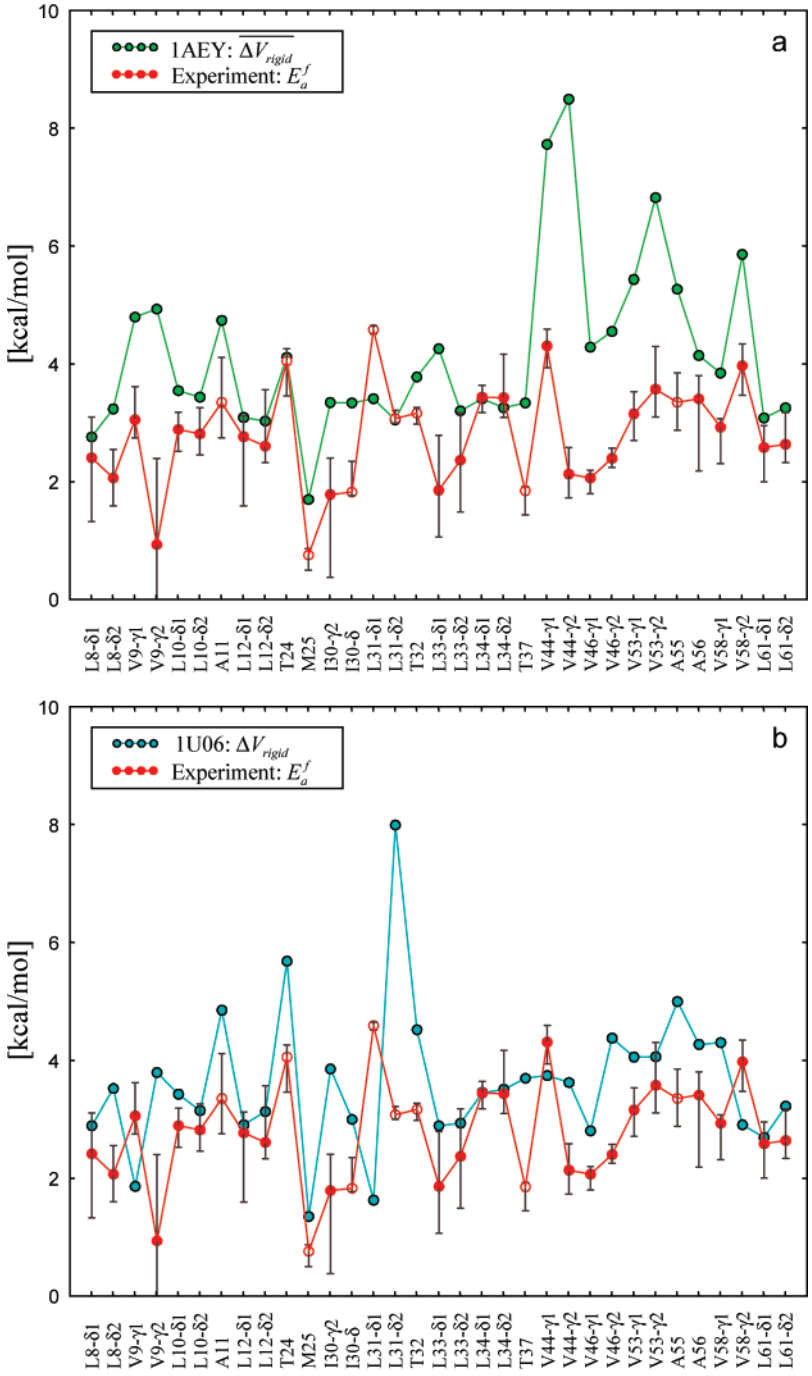

Figure 8. Rigid barriers and activation energies for methyl group rotation in α-spc SH3: $\overline{\Delta V_{\text{rigid}}}$ for NMR ensemble 1AEY consisting of 15 conformers (green), ΔV_{rigid} for X-ray structure 1U06, first chain (blue), and experimentally determined E_a^f (red). The X-ray structure was protonated prior to calculations and the proton coordinates were refined as described in Materials and Methods.

4.8 kcal/mol (data not shown). The observed trend suggests that rigid barriers provide a potential measure of structure quality.

To put these observations in a broader context we compiled a table where $\langle \Delta V_{\text{rigid}} \rangle$ and $\langle \overline{\Delta V_{\text{rigid}}} \rangle$ are determined for a number of X-ray and NMR structures, respectively. To facilitate the comparison, we have chosen a set of proteins for which both X-ray and NMR structures have been determined. In essence, NMR structures were used as is, whereas the crystallographic structures were protonated and the proton coordinates were refined prior to calculation of rigid barriers. The results for NMR structures are grouped into four categories according to the structure calculation software used. These include XPLOR/CNS, relying on the CHARMM force field,[13] DIANA and other programs that typically, but not always, invoke the AMBER force field,[62] GROMOS which uses the force field of the same name,[16] and, as a separate entry, a new dynamic ensemble refinement approach,[17] also using the CHARMM force field. This classification is, of course, tentative since a variety

(62) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(63) Simon, K.; Xu, J.; Kim, C.; Skrynnikov, N. R. *J. Biomol. NMR* **2005**, *33*, 83−93.

Table 1. $\langle \Delta V_{rigid} \rangle$ and $\langle \overline{\Delta V_{rigid}} \rangle$ (kcal/mol) for Selected X-ray and NMR Structures[a]

	X-ray[b]		XPLOR/CNS		DIANA/DYANA/ AMBER/DISCOVER		GROMOS		DER	
α-bungarotoxin	1HC9 (1.8 Å)	4.9	1BXP	11.8	1IKC	21.2	1KFH	3.2		
apolipoprotein	1AEP (2.7 Å)	14.3	1LS4	5.4						
barnase	1A2P (1.5 Å)	4.8	1BNR	6.3	1FW	6.7				
barstar	1B2S (1.8 Å)	3.9	1BTB	18.7						
bpti	1BPI (1.1 Å)	3.7			1PIT	3.5				
calmodulin N	1LIN[c] (2.0 Å)	6.5	1J7O	4.3	1AK8	8.9				
cdc42	1GRN (2.1 Å)	5.0	1CEE	4.9						
chemotaxis y	2CHF (1.8 Å)	12.3					1CEY[d]	3.6		
crambin	1EJG (0.5 Å)	4.2					1CCM	3.2		
cytochrome c_3	2CTH (1.7 Å)	4.3	1IT1	12.7	1A2I	5.0				
dihydrofolate reductase	3DFR (1.7 Å)	4.5	1AO8	3.9	1DIU	5.4				
helicase	1B79 (2.3 Å)	7.7			1JWE	3.9				
hipip	1B0Y (0.9 Å)	4.2			1HRR	4.7				
his phosphocarrier	1OPD (1.5 Å)	5.3	1J6T	12.5			1HDN	3.0		
interleukin	3IL8 (2.0 Å)	4.6	2IL8	5.0	1ILP	5.4				
lac repressor hpc			1CJG	4.3			1LQC	3.2		
lactoglobulin	1BEB (1.8 Å)	8.7	1CJ5	15.1						
profilin	1FIL (2.0 Å)	5.3	1PFL	8.8						
spectrin SH3[e]	1U06 (1.5 Å)	3.6			1AEY	4.1				
syntaxin	1EZ3 (1.9 Å)	5.2			1BR0	3.3				
ribonuclease	1M07 (1.8 Å)	4.3	1BC4	4.5						
ribonuclease Sa	1LNI (1.0 Å)	4.6			1C54[f]	5.4				
thioredoxin	1ERT (1.7 Å)	3.9	4TRX	4.6	1XOB	4.4				
ubiquitin	1UBQ (1.8 Å)	5.0	1D3Z	4.5	1G6J	4.5			1XQQ	3.5

[a] The set of proteins used in this table is mainly from the 100-protein database composed of X-ray/NMR pairs,[63] with several additions. Primary sequences of the proteins within the X-ray/NMR pairs are identical with the exception of (i) up to three point mutations (in one case, 1IT1, there are nine mutations) and (ii) several terminal residues. The procedure used for calculation of rigid barriers, $\langle \Delta V_{rigid} \rangle$ and $\langle \overline{\Delta V_{rigid}} \rangle$ for the X-ray and NMR structure, respectively, is described in Materials and Methods. [b] Crystallographic resolution reported in brackets. [c] C-terminal domain has been deleted from the structure 1LIN. [d] The structures were calculated with DIANA and then refined via restrained molecular dynamics and energy minimization under GROMOS. [e] For consistency with the previous analyses, Val 23 methyl sites were not included in calculating the average barriers in α-spc SH3. [f] The structures were calculated with DYANA and energy-minimized with GROMOS.

of different protocols and auxiliary programs have been used to calculate these structures.

It is convenient to begin the discussion of the results in Table 1 by considering the barriers calculated for the X-ray structures. We expect the barriers to be overestimated since, as demonstrated earlier, the barriers depend on the "excited" dynamic states of the protein and cannot be calculated accurately based on a "ground-state" structure. In this sense, a good guidance is presented by three ultrahigh-resolution structures, 1BPI, 1EJG, and 1LNI. The $\langle \Delta V_{rigid} \rangle$ values for these structures fall in the range 4.2−4.6 kcal/mol. It can be suggested, therefore, that for a single accurate set of coordinates the rigid barrier calculation procedure overestimates the true barrier height by ca. 1.5 kcal/mol.

At the same time, Table 1 contains a number of crystallographic structures with abnormally high $\langle \Delta V_{rigid} \rangle$ values. In each case it can be traced to steric hindrances, which ultimately reflect the limited resolution or specific defects of the structure. In the three lowest-resolution X-ray structures, 7% of all methyls show unrealistic barriers in excess of 20 kcal/mol. On the other hand, no such methyls occur in the three X-ray structures with the highest resolution.

In going from X-ray to NMR structures, we expect two opposite trends to take effect: (i) because of the generally lower quality of NMR structures there should be more steric conflicts leading to higher barriers and (ii) because each NMR structure represents the ensemble of many conformational states, the log-averaging procedure should lead to lower barriers. Both things considered, the majority of NMR structures fall in the range 4−6 kcal/mol, similar to crystallographic structures. There is, however, a substantial fraction of structures with very high $\langle \overline{\Delta V_{rigid}} \rangle$ values.

It can be expected that $\langle \overline{\Delta V_{rigid}} \rangle$ should correlate with conventional measures of structure quality, in particular those that refer to steric clashes. For instance, a popular structure validation program WHAT IF[64] determines a number of "bumps" (i.e., interatomic distances that are shorter than the sum of two van der Waals radii less 0.4 Å). It turns out that for three NMR structures in Table 1 with the lowest $\langle \overline{\Delta V_{rigid}} \rangle$ there are 3 bumps for every 1000 atoms. On the contrary, for three structures with the highest $\langle \overline{\Delta V_{rigid}} \rangle$ there are 112 bumps for every 1000 atoms.

One notable feature of $\langle \overline{\Delta V_{rigid}} \rangle$ is illustrated in Figure 9. Shown in this figure is the example of a methyl group with an external proton (red sphere) "tucked" in between of the methyl protons. This configuration is favorable from the point of view of the van der Waals interaction. However, a steric clash clearly occurs when the methyl rotates. The trend illustrated in Figure 9 is to some extent responsible for the systematic overestimation of the barriers, which is not necessarily reflected in the van der Waals violations.

By itself, the presence of conformations such as illustrated in Figure 9 is not unexpected. If these conformations were rare, their presence would be of little consequence and $\langle \overline{\Delta V_{rigid}} \rangle$ would be reasonably low. Consider, for example, Figure 5b. There the structural ensembles are formed from 20 randomly chosen MD snapshots and used for rigid barrier calculations. The typical $\langle \overline{\Delta V_{rigid}} \rangle$ in this case is 3.4 kcal/mol, which is much

(64) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52−56.

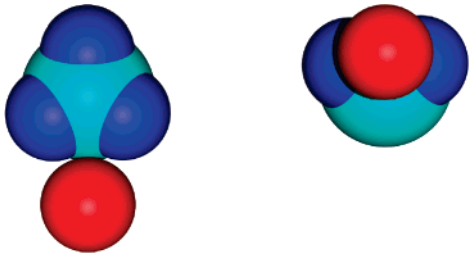

Figure 9. Steric hindrance to methyl rotation: Val 9 γ1 methyl (blue and cyan) and Lys 60 $H^{\gamma 1}$ (red) from model 5 in the coordinate set 1AEY of α-spc SH3. The line of view is along the Val 9 $C^{\beta}-C^{\gamma 1}$ bond (left panel) and perpendicular to it (right panel). This particular methyl−proton pair has been chosen for illustration purpose only; in fact, there are few examples of such neat packing. Nevertheless, the general trend whereupon the outside atom is partially inserted between methyl protons is evident in many NMR structures.

lower than most of the entries in Table 1 and closer to the experimental result.

It is clear, however, that conventional NMR structures are different from true dynamic ensembles. To see that it is the case, recall that the majority of the NMR structures are annealed toward the temperature of 0 K and then subjected to energy minimization. Under these circumstances, the conformations such as shown in Figure 9 tend to occur more frequently and, as a result, higher $\langle \overline{\Delta V_{\text{rigid}}} \rangle$ values are obtained (see Table 1). This observation underscores the somewhat ambivalent nature of an NMR structure. Although NMR data are collected at around room temperature, the structure calculation protocol makes it, at least in some respects, a low-temperature structure.

A notable exception from the general trend is a group of NMR structures refined by means of GROMOS, fifth column in Table 1. Those structures show the lowest barriers, 3.0−3.6 kcal/mol. This can be readily explained since in GROMOS methyl groups are treated in a united atom approximation, that is, as expanded spheres. Consequently, the packing effects, such as illustrated in Figure 9, do not occur during the GROMOS refinement (see Figure S3). When explicit protons are added to such a structure, the rigid barriers remain low, reflecting a healthy separation between each methyl and its surrounding atoms. It is an interesting observation that NMR structures calculated with GROMOS apparently provide the best model for dynamic methyl environment in the protein core.

Lately, the efforts have been underway to build NMR structures that would provide a faithful representation of protein dynamics.[17,60] In particular, Lindorff−Larsen, Vendruscolo, and co-workers developed the dynamic ensemble refinement (DER) approach where the dynamic NMR data, such as order parameters, are incorporated in structure calculations. Using this strategy, the authors re-generated coordinates of ubiquitin, 1XQQ. Our work provides a (sufficiently independent) check for the quality of this ensemble. As it turns out, the structure 1XQQ gives rise to a perfectly reasonable rigid barrier, 3.5 kcal/mol, which is substantially better than what has been obtained with other ubiquitin structures (see bottom row in Table 1 and Figure S4). In fact, this barrier height is about the same as found in our analyses of the collections of MD snapshots. Hence the structure 1XQQ can be indeed viewed as a miniature MD ensemble built under the control of diverse NMR data.

Concluding Remarks

Several efficient methods have been developed to predict backbone dynamics on the basis of protein structure.[65,66] In contrast, the search for structural determinants of side-chain dynamics had only limited success.[32,35,67,68] The correlations between methyl τ_f and structural parameters such as solvent exposure or local packing density proved to be tenuous at best. At the same time it has been demonstrated that very subtle structural changes caused by point mutations can alter τ_f.[34]

In this study we attempted to relate the activation energy associated with τ_f to a yet another structural parameter, ΔV_{rigid}. While we found no correlation on a per-residue basis, the average rigid barrier proved to be a meaningful measure of structure quality. It should be emphasized, however, that more experimental data from different proteins are needed before one can make a confident judgment about the barrier heights typically encountered in proteins.

A variety of NMR parameters have been used to assess the quality of structures. Traditionally, 'static' parameters have been used for this purpose such as NOE,[69] chemical shifts,[70] and dipolar couplings.[71] In this work we provide the example of structure validation using the experimental parameter which is essentially dynamic in origin, namely methyl rotation barrier. The experimental information about methyl barriers in proteins can help to generate better structural models representative of proteins' dynamic nature.

Materials and Methods

Sample Conditions. α-spc SH3 was expressed by growing cells in 100% D_2O, using a 3-[60%-^2H,^{13}C] pyruvate as the sole carbon source or, alternatively, in 50% D_2O with u-^{13}C glucose as a carbon source. The protein concentration in the two samples was 1.0 and 1.5 mM, respectively. The sample conditions were pH 3.5 (unbuffered), 90:10 H_2O:D_2O.

NMR Measurements. NMR spectra were recorded at 600 MHz Varian Inova spectrometer at the temperatures 10, 17, 24, and 30 °C. ^2H relaxation data were collected using the experiments by Kay and co-workers.[4,20] In these experiments the correction associated with the two-spin order H_zC_z is dealt with in the main pulse sequence[20] so that there is no need for a separate reference experiment. Each measurement sampled seven points on the relaxation curve and took 24 h to record. In the R_2 experiment, the deuterium spin lock with the field of 1.3 kHz was applied for the maximum of 15 ms. Measured relaxation rates (two samples, four temperatures) together with the results of the model-free analysis have been deposited in the Biological Magnetic Resonance Data Bank[72] (BMRB ID 15144). ^{15}N data were recorded using the updated version of the experiments by Farrow and Kay.[73,74] During the relaxation period, cross-correlations were suppressed by applying

(65) Sunada, S.; Go, N.; Koehl, P. *J. Chem. Phys.* **1996**, *104*, 4768−4775.
(66) Zhang, F. L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 12654−12655.
(67) Mittermaier, A.; Kay, L. E.; Forman-Kay, J. D. *J. Biomol. NMR* **1999**, *13*, 181−185.
(68) Ming, D. M.; Brüschweiler, R. *J. Biomol. NMR* **2004**, *29*, 363−368.
(69) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Saffrich, R.; Nilges, M. *Science* **1993**, *261*, 328−331.
(70) Williamson, M. P.; Kikuchi, J.; Asakura, T. *J. Mol. Biol.* **1995**, *247*, 541−546.
(71) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008−9012.
(72) Doreleijers, J. F.; Mading, S.; Maziuk, D.; Sojourner, K.; Yin, L.; Zhu, J.; Markley, J. L.; Ulrich, E. L. *J. Biomol. NMR* **2003**, *26*, 139−146.
(73) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.
(74) Korzhnev, D. M.; Skrynnikov, N. R.; Millet, O.; Torchia, D. A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10743−10753.

up to eight 180° pulses with the REBURP profile[75] which inverted $^1H^N$ signals without perturbing water. All spectra were processed using nmrPipe package; peak intensities were quantified using nlinLS.[76] The ^{15}N data set was reduced by excluding the data from the flexible N-terminus, residues 2−7, where heteronuclear NOEs were found to be less than 0.65 (10 °C). The retained data were analyzed by means of the program R2R1_diffusion[77] using the protein coordinates 1U06.[9]

Molecular Dynamics. MD simulations were performed using the program CHARMM[13] version 32b2 with the CHARMM22 force field and the modified TIP3P water potential.[78] The force constant associated with methyl torsional angle ϕ was revised as recommended by Chatfield[38] such that ΔV_ϕ was reduced from 3.6 to 2.6 kcal/mol for all positions except Ile-δ (default value 2.9 kcal/mol) and Met (1.7 kcal/mol). The empirical CMAP correction,[79] which has an effect of restraining backbone dynamics, has been in use as a default CHARMM option. The starting protein structure was built from the first chain of the X-ray structure 1U06[9] which misses the first six residues belonging to the flexible N-terminus. The cubic water box $56 \times 56 \times 56$ Å^3 was built around the protein (the size of the box was determined by adding 24 Å, or twice the nonbonded interaction cutoff, to the long dimension of the protein). In doing so, all crystallographic water molecules were retained. The box was equilibrated and then trimmed to a truncated octahedron. This system was set up with periodic boundary conditions. All bonds involving hydrogen atoms were constrained to their equilibrium values using the SHAKE algorithm.[80] For electrostatic interactions, the particle mesh Ewald scheme was used with a Gaussian width of 0.35 Å.[81] The real-space interactions were calculated with a cutoff 12 Å and a transition function between 10 and 12 Å; the nonbonded list was maintained to 14 Å and updated heuristically. The reciprocal-space contributions were calculated with the FFT algorithm using 60 grid points per box length and a sixth-order B-spline interpolation scheme. The equations of motion were integrated using the leapfrog scheme with the time step of 2 fs.

The system was energy-minimized via four sequential runs of 200 steps each with harmonic restraints using the force constants of 100, 10, 5, and 0 kcal/(mol·Å^2). It was subsequently heated from 43 to 293 K with the steps of 5 K/1 ps. In doing so, heavy backbone atoms were harmonically constrained with a force constant of 5*(particle mass) kcal/(mol·Å^2·amu). The equilibration consisted of a 20-ps MD simulation with 5*(particle mass) harmonic restraints, followed by an 80-ps MD simulation without harmonic restraints. The equilibration as well as the subsequent 25-ns production run was performed under NPT conditions (293 K, 1 atm). During the production run the coordinates were stored every 50 fs. The simulation required ca. 5 weeks on a GNU/Linux workstation with two 3-GHz dual-core Xeon processors. The backbone rmsd over the course of the trajectory is shown in Figure S5.

In addition to the 25-ns trajectory, short 1-ns trial trajectories have been recorded at different temperatures (integration step 1 fs, sampling rate 0.01 ps) and used to calculate ΔV_{pmf}. As it turned out, ΔV_{pmf} at 283 and 303 K are tightly correlated, $r = 0.95$, and register only a small systematic difference, 0.2 kcal/mol. Thus, the MD data point toward Arrhenius behavior, consistent with the experimental observations.

To derive the correlation functions shown in Figure 3, individual frames from the 25-ns trajectory were first superimposed onto the structure 1U06 by matching the positions of heavy backbone atoms belonging to the elements of secondary structure. From this pre-processed trajectory, methyl CH vectors and dihedral angles ϕ were extracted and used to calculate the correlation functions, $G(\tau)$. To improve the computational efficiency, the correlation functions were evaluated on a sparse grid: $\tau_1 = 0$, $\tau_2 = 0.5$ ps and further $\tau_{i+1} = \tau_i + 0.005\tau_i$, with τ_{i+1} rounded upward to $n\times0.5$ ps. The obtained correlation functions were symmetrized with respect to three methyl protons, multiplied by $\exp(-\tau/\tau_{\mathrm{rot}})$ and then used to predict the deuterium relaxation rates.[24,82] These simulated rates were subsequently analyzed by means of the LS-2 model, yielding τ_{f} values displayed in Figure 4.

Barrier Calculations. In preparing for the calculations of $\Delta V_{\mathrm{rigid}}$, all structures have been stripped of ligands, crystallographic water, prosthetic groups, etc. Subsequently, the structures were processed with python script, written in-house, to ensure that they can be loaded into XPLOR[54] for the purpose of the $\Delta V_{\mathrm{rigid}}$ calculations. When needed, a small number of missing heavy atoms (e.g., C-terminal oxygen) were added to the coordinates and their positions were optimized via 100 rounds of Powell minimization. For proton-less crystallographic structures as well as NMR structures solved with GROMOS, protons were added using the HBUILD facility of XPLOR and their coordinates were refined using 40 rounds of Powell minimization. For some NMR structures solved with software other than GROMOS a small number of protons were added (e.g., N-terminal protons), but no Powell minimization was applied. In the case of crystallographic structures, $\Delta V_{\mathrm{rigid}}$ were calculated using a single polypeptide chain from each coordinate set (when several chains were present, the first one was selected). In the case of NMR structures the entire ensemble has been used for calculations and then log-average was determined according to eq 3. The collection of scripts used for barrier calculations is available from the authors upon request.

In calculating $\Delta V_{\mathrm{rigid}}$ the cutoff for nonbonded interactions was set to 6.5 Å, with switching function localized between 6 and 6.5 Å, and the nonbonded list maintained to 7.5 Å. Using longer radius, as in MD simulations, did not cause any appreciable change in $\Delta V_{\mathrm{rigid}}$. The tests were also performed to assess the effect of water on calculations of rigid barriers. The effect proved to be small: for example, when crystallographic water was retained in the structure 1U06 the calculated barriers changed, on average, by 0.05 kcal/mol.

Acknowledgment. This study was funded in part by the NSF CAREER grant 044563 to N.R.S.

Supporting Information Available: Correlation between the experimentally determined and MD-based τ_{f}, histograms illustrating the determination of ΔV_{pmf}, united-atom representation of the methyl group in GROMOS, rigid barriers in the pair of ubiqutin structures (1UBQ and 1XQQ), backbone atom *rmsd* in the MD trajectory of α-spc SH3. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0702061

(75) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93−141.
(76) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(77) Mandel, A. M.; Akke, M.; Palmer, A. G. *J. Mol. Biol.* **1995**, *246*, 144−163.
(78) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(79) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36−L38.
(80) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(81) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(82) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 4272−4284.

Methyl rotation barriers in proteins from ^{2}H relaxation data. Implications for protein structure.

Yi Xue,[†] Maria S. Pavlova,[†] Yaroslav E. Ryabov,[†] Bernd Reif,[‡] Nikolai R. Skrynnikov[†]*

Supporting information

† Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907-2084, USA

‡ Forschunginstitut für Molekulare Pharmakologie (FMP), Robert-Rössle-Str. 10, 13125 Berlin, Germany

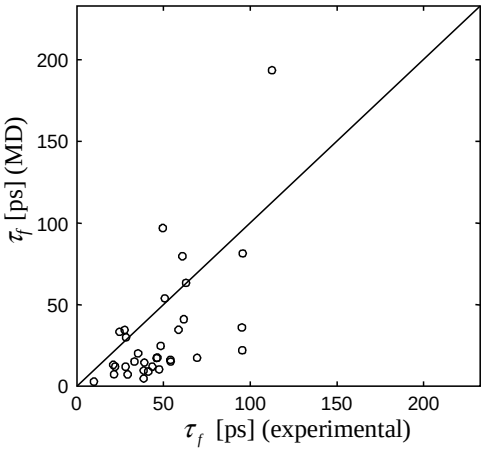

Fig. S1. Comparison of the MD-based and experimentally determined correlation times τ_f. The MD-based τ_f is the same as in Fig. 4; the nominal temperature of the MD simulation is 20 ºC. The experimentally determined τ_f are from the LS-2 fit of ^{2}H R_1 and R_2 data from the glucose-derived sample of α-spc SH3 at 24 ºC. The correlation coefficient for the data in the plot is r =0.64.

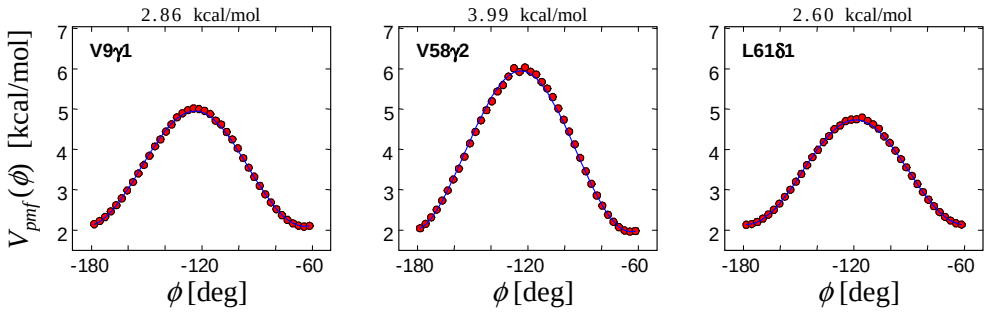

Fig. S2. Potential of mean force for methyl group rotation as obtained from a 25 ns MD trajectory of the α-spc SH3 domain. For the purpose of these calculations, the trajectory has been sampled at 0.05 ps. The results are fitted as described by Chatfield and Wong.[1] The selection of methyls is the same as in Fig. 2.

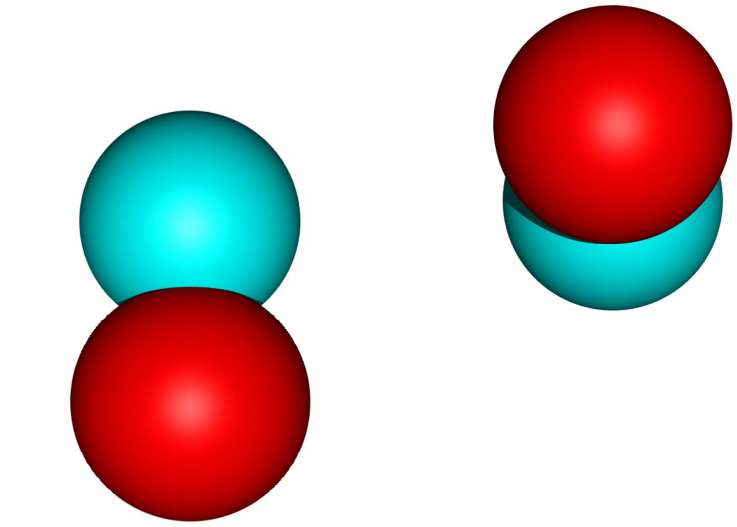

Fig. S3. Steric hindrance to methyl rotation as seen in the united-atom representation: Val 9 γ1 methyl (CH3, cyan) and Lys 60 γ methylene (CH2, red) from model 5 in the coordinate set 1AEY of α-spc SH3. The united atoms are centered at the respective carbons and their van der Waals radii are chosen according to the GROMOS force field parameter set.[2] Unlike in Fig. 9, the steric hindrance (red sphere) creates a van der Waals clash in this model. If the geometry is further refined, the clash will be relieved and the distance between the two carbons will increase, leaving more room for methyl rotation. This helps to explain why rotation barriers found in GROMOS-generated structures are relatively low, Tab. 1.

While united-atom force fields appear to produce better results with regard to methyl packing, they do not perform quite as well with regard to some other aspects of protein structure. For example, aromatic-aromatic interactions are better modeled with all-atom force fields.[3] In general, the improvement afforded by the use of GROMOS in this context is a result of a somewhat crude approximation.

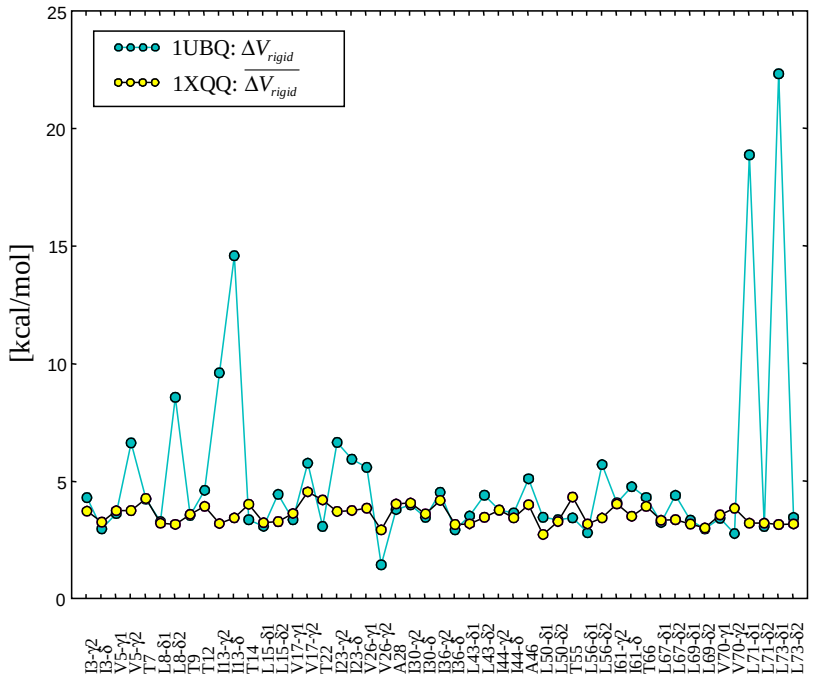

Fig. S4. Methyl rotation barriers calculated using the rigid barrier approximation for the X-ray coordinate set 1UBQ (cyan) and NMR coordinate set 1XQQ (yellow). The corresponding mean values are $< \Delta V_{rigid} >$ = 5.0 kcal/mol and $< \overline{\Delta V_{rigid}} >$ = 3.5 kcal/mol, respectively.

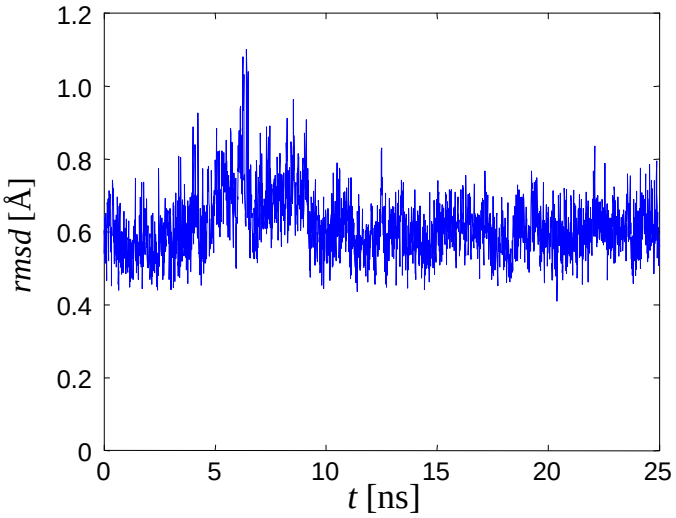

Fig. S5. Heavy backbone atoms *rmsd* relative to the first chain of the structure 1U06 evaluated over the time course of a 25-ns trajectory of the α-spc SH3.

References

(1) Chatfield, D. C.; Wong, S. E., *J. Phys. Chem. B* **2000,** *104*, 11342-11348.
(2) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., *J. Comput. Chem.* **2004,** *25*, 1656-1676.
(3) Mackerell, A. D., *J. Comput. Chem.* **2004,** *25*, 1584-1604.