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

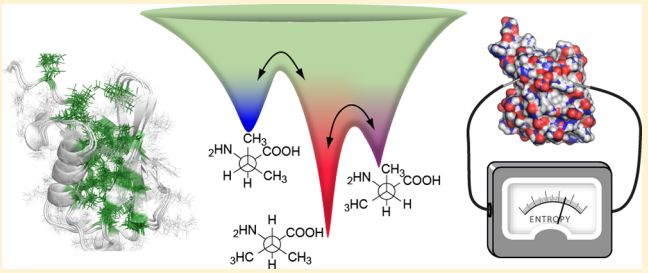
Microscopic Insights into the NMR Relaxation-Based Protein Conformational Entropy Meter

Vignesh Kasinath, Kim A. Sharp,* and A. Joshua Wand*

Graduate Group in Biochemistry and Molecular Biophysics and the Johnson Research Foundation and Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia 19104, United States

Supporting Information

ABSTRACT: Conformational entropy is a potentially important thermodynamic parameter contributing to protein function. Quantitative measures of conformational entropy are necessary for an understanding of its role but have been difficult to obtain. An empirical method that utilizes changes in conformational dynamics as a proxy for changes in conformational entropy has recently been introduced. Here we probe the microscopic origins of the link between conformational dynamics and conformational entropy using molecular dynamics simulations. Simulation of seven proteins gave an excellent correlation with measures of side-chain motion derived from NMR relaxation. The simulations show that the motion of methyl-bearing side chains are sufficiently coupled to that of other side chains to serve as excellent reporters of the overall side-chain conformational entropy. These results tend to validate the use of experimentally accessible measures of methyl motion—the NMR-derived generalized order parameters—as a proxy from which to derive changes in protein conformational entropy.



INTRODUCTION

The thermodynamic nature of the folded state of proteins governs their biological function in innumerable ways. Indeed, the many forces that govern the stability of the native state are complex and continue to be difficult to unravel and dissect through experiment, simulation, or theory.[1,2] This is particularly true for the residual conformational entropy of proteins, which is potentially a central component of the thermodynamics governing protein function.[3−5] Historically, it has been impossible to experimentally determine the contribution of residual protein entropy to fundamental protein activities, such as the binding of ligands, vital for protein function. Recent progress has illuminated the possibility of employing NMR relaxation methods to quantitatively determine the role of changes in conformational entropy in molecular recognition by proteins.[6] The approach rests on using fast internal protein dynamics as a proxy for conformational entropy.[7] The initial approach was to effectively take an inventory of the change in motion at a limited number of sites and interpret this within the context of a simple physical model, such as the harmonic oscillator[8] or diffusion within an infinite square well.[9] This raises several obvious issues, such as the effects of correlated motion, the operation of a more complex potential energy function, the completeness of the oscillator count, and so on.[7] More recently this issue was sidestepped by using an empirical calibration of the dynamical proxy for conformational entropy. Rather than attempt a model-dependent interpretation of an inventory of changes in local dynamics, an empirical scaling between the experimental measures of local dynamics and local disorder (entropy) was attempted.[10] The idea rests on the assumption that the experimentally accessible methyl-bearing side chains are numerous enough to provide good coverage of internal motion and are sufficiently coupled to neighboring side chains to report on the entropy.[6] With an appropriate experimental system, this leads to a rather simple relationship between what can be measured (protein motion and total binding entropy) or confidently calculated (solvent entropy) and what is desired (conformational entropy).[10] Thus by comprehensively measuring the change in motion of methyl-bearing side chains on the nanosecond time scale using classical NMR relaxation methods,[11−15] one could obtain a quantitative measure of the underlying change in conformational dynamics.[10] This approach was first introduced using the binding of calcium-saturated calmodulin to a series of calmodulin-binding domains.[10] The results demonstrated that changes in conformational entropy were important to ligand binding and also confirmed an earlier suggestion[16] that conformational entropy contributes to the tuning of the free energy of binding in this system.[10] Recently, a second example has appeared where Tzeng and Kalodimos used an impressive collection of mutants of the catabolite activator protein (CAP) to undertake a similar analysis of the thermodynamic origins of high affinity binding of DNA to the cAMP-activated protein.[17] The advantage of the CAP system is that the same binding ligand (DNA) was used, the structure does not change appreciably upon binding the DNA, and an array of structurally benign mutations remote from the binding interface were available. This eliminated most

Received: May 23, 2013
Published: September 6, 2013

   dx.doi.org/10.1021/ja405200u | J. Am. Chem. Soc. 2013, 135, 15092−15100

of the uncertainties that may have hampered the analysis of the calmodulin complexes described above. The resulting empirical calibration for CAP was most impressive and led to a convincing analysis of the role of conformational entropy in the binding of ligands to the CAP protein and its potential role in its allosteric regulation.[17]

Here we attempt to understand the microscopic origins of the apparent success of the empirical "entropy meter" based on a dynamical proxy measured through NMR relaxation methods. We employ extensive molecular dynamics (MD) simulations in explicit water to examine the ability of measures of fast motion of methyl groups in proteins to adequately represent the conformational entropy of the protein. In addition, we explore several issues, such as the degree to which correlated or conditional motion affects the analysis. In order to rely on such an analysis, it is required that the MD simulations faithfully represent those motions contributing to experimentally observables such as NMR relaxation. Although there have recently been significant advances in computational approaches, particularly with respect to long-time scale dynamics,[18,19] there has been very little benchmarking of MD simulations with experimental measures of methyl-bearing side-chain motion derived from NMR relaxation.[20] With one exception,[21] the correspondence between Lipari-Szabo (L-S) methyl group squared generalized order parameters derived from MD and experiment in the few cases examined has been mixed and insufficient to promote detailed analysis in this context. In a preamble to the main thrust of this paper, we compare L-S squared generalized order parameters of methyl group symmetry axes derived from deuterium or carbon relaxation methods. Using NAMD[22,23] simulations of seven proteins with the CHARMM27 force field,[24,25] we find excellent agreement with experiment, which provides a solid foundation for addressing the microscopic origins of the "entropy meter" established using NMR relaxation methods. Analysis of the simulations then reveals that the experimentally accessible motions of methyl-bearing side chains are sufficiently coupled to the motion of other side chains to serve as excellent reporters of the protein conformational entropy: The total side-chain conformational entropy can be accurately recapitulated using only measures of side-chain methyl motions. The conformational entropy varies mostly through changes in populations of rotameric states rather than by variation in the effective potential defining each rotameric well, although some well narrowing is seen at very rigid sites. Importantly, a statistical analysis reveals that correlated motion is sufficiently limited to have little impact on the relationship between the NMR measure of methyl dynamics and the derived entropy. These results validate the use of conformational dynamics derived experimentally from NMR relaxation as a proxy from which to obtain quantitative estimates of changes in protein conformational entropy.

METHODS

MD Simulations. MD simulations of the seven proteins listed in Table 1 were carried out with NAMD2,[23] using the CHARMM27[24] all-atom parameter set and the TIP3P[26] water potential. Hydrogen atoms were added to the PDB structures with VMD.[27] Using VMD[27] individual proteins were centered in a TIP3P water box such that protein atoms were at least 8 Å from the boundary except in the case of the two small proteins, ubiquitin and α_3D, where the solvent layer was at least 5 and 6 Å, respectively.

Simulations were performed using a time step of 2 fs. Bonds to hydrogen atoms were constrained using the SHAKE algorithm.[28] A switching distance of 10 Å and cutoff of 12 Å were used for the nonbonded interaction interactions, combined with a particle mesh Ewald summation with 1 Å grid spacing for long-range electrostatics. Simulations were run at constant temperature and pressure of 1 atm, controlled using the extended Langevin method. Simulation temperatures corresponded to those at which the NMR relaxation experiments were performed (Table 1). Following equilibration runs of at least 1 ns, several 60 ns data production runs were performed with every subsequent 60 ns simulation starting from the final coordinates of the earlier run but with different initial velocities. For three protein systems, ubiquitin and calmodulin-smMLCKp and calmodulin-nNOSp complexes, longer simulations were also run on the Anton supercomputer at the Pittsburgh Supercomputer Center using the same force field and simulation conditions except for a nonbond cutoff of 14 Å.

Table 1. Characteristics of the Protein Set Used for MD Simulations

proteins[a]	residues	PDB[b]	T (°C)[c]	waters[d]	length[e] (ns)
ALBP	131	1LIB	20	4008	112
a_3D	73	2A3D	30	2532	160
Cyt $c2$	116	1C2R	30	5116	120
CaM-smMLCKp	167	1CDL	35	5228	1280[f]
CaM-nNOSp	168	2O60	35	4957	1120[f]
HEWL	129	1LZA	35	4133	240
ubiquitin	76	1UBQ	25	2238	260[f]

[a]Abbreviations: ALBP, adipocyte lipid binding protein; Cyt c2, cytochrome c2; CaM-smMLCKp, calcium-saturated calmodulin (CaM) in complex with a peptide corresponding to the smooth muscle myosin light-chain kinase calmodulin-binding domain (smMLCKp); CaM-nNOSp, calcium-saturated calmodulin in complex with a peptide corresponding to the neuronal nitric oxide synthase calmodulin-binding domain (nNOSp); HEWL, hen egg white lysozyme. [b]PDB code of starting structure. [c]Temperature of simulation and NMR experiments. [d]Number of waters in the simulation. [e]Length of the simulation. [f]Utilized Anton (see Methods section).

Analysis of Internal Motion. The L-S[29] squared generalized order parameters (O^2) were calculated from MD simulations by overlaying snapshots of the protein from the trajectories using a standard rigid-body alignment (Cα) procedure. For each snapshot, the unit vector along the methyl symmetry axis was obtained in terms of its vector components in Cartesian axes, x, y, and z. The O^2 parameter for a given methyl is then calculated using[30]

$$O^2 = \frac{3}{2}[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle yz\rangle^2 + 2\langle xz\rangle^2] - \frac{1}{2} \quad (1)$$

where < > indicates the average over the trajectory. For comparison with experiment, the order parameter was derived from the measured order parameter for the methyl C−H bond by assuming free rotation around the symmetry axis and ideal tetrahedral geometry ($\theta = 109.5°$) by[31]

$$O^2_{obs} = O^2_{rot} \times O^2_{axis} = \frac{(3\cos^2\theta - 1)^2}{4} O^2_{axis} = 0.1108 \times O^2_{axis} \quad (2)$$

Order parameters from independent simulations on the same protein were first computed using eq 6, averaged, and then compared to those obtained experimentally using NMR relaxation through eq 7.

Calculation of Rotameric Entropy. Side chain χ angles for each amino acid were calculated from the MD trajectories and binned into one of three conformational states based on the dihedral angle: gauche$_+$ - [0°, 120°], trans - [120°, 240°]; and gauche$_-$ - [240°, 360°]. The rotamer probability distribution function (pdf) of all 3^{N_χ} rotamers of each side chain was constructed from the χ angle histograms, where N_χ is the number of side-chain χ angles. Only unique side-chain

dx.doi.org/10.1021/ja405200u | J. Am. Chem. Soc. 2013, 135, 15092−15100

Table 2. Correspondence between Simulated and Observed Motion of Methyl-Bearing Amino Acid Side Chains

proteins	exptl $<O^2_{axis}>$	MD $<O^2_{axis}>$	R^2	slope
ALBP	0.633	0.619	0.75	0.88
a_3D	0.451	0.571	0.76	1.25
Cyt *c*2	0.767	0.670	0.47	0.68
CaM-smMLCKp	0.583	0.562	0.62	0.83
CaM-nNOSp	0.534	0.560	0.50	0.77
HEWL	0.713	0.699	0.64	0.67
ubiquitin	0.664	0.629	0.85	0.86
average	0.620 ± 0.11	0.616 ± 0.05	0.65 ± 0.14	0.85 ± 0.13

torsion angles were considered (e.g., the isopropyl group contributes to one giving Leu 2 and Val 1 side-chain torsion angles). The side-chain rotamer entropy is

$$S/k_B = -\sum_i p_i \ln p_i \tag{3}$$

where k_B is the Boltzmann constant, p_i is the probability of the i^{th} rotamer. $S_b = S/k_B$ is the entropy given in dimensionless units, and the normalized entropy is then defined as S_b/N_χ.

Calculation of the Influence of Correlated Motions on Side-Chain Entropy. To quantify the effect of correlated motions of side-chains on their entropy, one-, two-, and three-dimensional (1D, 2D, and 3D) pdfs for every side-chain torsion angle, doublet, or triplet combination, respectively, were constructed from the trajectories. The torsion angles were again binned into gauche$_+$, gauche$_-$, and trans conformations. The 1D (i.e., uncorrelated) entropy, S_{1D} was obtained by summing $-P(\chi)\ln P(\chi)$ over each 1D pdf, $P(\chi)$, and then summing over all residues in the protein. The total side-chain entropy, including the effect of correlated motions of side-chains, was then estimated using the maximum information spanning tree (MIST) method.[32] A second-order MIST approximation was obtained by computing pairwise second-order mutual information terms (I_2) from the 1D and 2D pdfs as

$$I_2(j,k) = \sum_{l=1}^{3}\sum_{m=1}^{3} P(\chi_l^j, \chi_m^k) \ln\left(\frac{P(\chi_l^j, \chi_m^k)}{P(\chi_l^j)P(\chi_m^k)}\right) \tag{4}$$

where the 2D pdf $P(\chi_l^j,\chi_m^k)$ expresses the joint probability that side-chain torsion angles j, k are in the l'th and m'th conformations respectively. $P(\chi_l^j)$ and $P(\chi_m^k)$ are the corresponding 1D pdfs. Given all the I_2's, the second order MIST expansion is then obtained by selecting a spanning set of I_2's to maximize:

$$I_2^{MIST} = \sum_{i=2}^{n} I_2(j,i) \tag{5}$$

where n is the total number of side-chain torsion angles in the protein. To construct this spanning set, for each i, the index j is chosen from the set $\{1, 2, ..., i-1\}$, which gives the largest $I_2(i,j)$. Similarly, a third-order mutual information term (I_3):

$$I_2(i,j,k) = \sum_{o=1}^{3}\sum_{l=1}^{3}\sum_{m=1}^{3} P(\chi_o^i,\chi_l^j,\chi_m^k) \ln\left(\frac{P(\chi_o^i,\chi_l^j,\chi_m^k)}{P(\chi_o^i,\chi_l^j)P(\chi_m^k)}\right) \tag{6}$$

expresses the change in mutual information between i and j by including the k'th torsion angle. A third-order MIST approximation is built up analogously by choosing a spanning set of I_3's to maximize:

$$I_3^{MIST} = I_2(1,2) + \sum_{k=3}^{n} I_3(i,j,k) \tag{7}$$

where for each k, the indices i and j are chosen from the set $\{1, 2, ..., k-1\}$, which gives the largest $I_3(i,j,k)$. Then an upper bound estimate for the total entropy including correlations at order O is[32]

$$S^{total}_{MIST\text{-}O} = S_{1D} - I_O^{MIST} \quad (O = 2, 3) \tag{8}$$

RESULTS

Seven proteins ranging in size from 73 to 168 residues were selected for study (Table 1). Some have bound metals (calmodulin) or a prosthetic group (cytochrome c2). Two are high-affinity complexes of small peptides with calmodulin. All have extensive experimental reference sets for dynamics derived from site-resolved solution NMR relaxation studies. The methyl group L-S squared generalized order parameters (O^2_{axis}) were taken from published deuterium methyl relaxation studies: α_3D,[33] adipocyte lipid binding protein,[34] cytochrome c_2,[35] calcium-saturated calmodulin (CaM) complexes with peptides corresponding to the calmodulin-binding domain of the smooth muscle myosin light chain kinase (smMLCKp) and the neuronal nitric oxide synthase (nNOSp),[16] hen egg white lysozyme,[36] and ubiquitin.[37] The average dynamical properties of the proteins range from quite rigid (e.g., cytochrome c_2 and lysozyme) to quite internally dynamic (e.g., α_3D). The molecular coordinates for the proteins used in the simulations were derived from structures deposited in the Protein Data Bank: ubiquitin,[38] α_3D,[39] $(Ca^{2+})_4$-calmodulin-smMLCK,[40] $(Ca^{2+})_4$-calmodulin-nNOS (unpublished), hen egg white lysozyme,[41] adipocyte lipid binding protein,[42] and cytochrome c_2[43] (Table 1).

Comparison of Simulation and Experimental Methyl Side-Chain Dynamics. The primary goal of this work is to discover the microscopic origins of the apparent success in using a dynamical proxy employing motion of methyl-bearing side chains for the determination of conformational entropy of proteins. As a first step toward this goal, it is important to determine the reliability and accuracy of MD simulations in capturing the experimentally determined side-chain motion. Here we employ the CHARMM27 potential[24] implemented in the context of the NAMD molecular dynamics environment.[23] Trajectories were carried out in explicit water at the temperature at which the experimental measurements were made (Table 1). Methyl group symmetry axis L-S squared generalized order parameters (O^2_{axis}) were calculated from the trajectories using eq 6. Convergence of the simulations with regard to O^2_{axis} values was confirmed by examination of the full-time correlation function $C(t) = P_2(<\mu(t+T)\cdot\mu(t)>)$ for selected methyl groups spanning the range of observed O^2 values and by the variation in O^2_{axis} calculated from different batches of individual trajectories (Figure S1).

In detail, a methyl-by-methyl comparison reveals considerable variation between the calculated and simulated L-S O^2_{axis} parameters with both over- and underestimation being evident (Table 2 and Figure S2). Table 2 summarizes the statistical comparison of O^2_{axis} parameters obtained from simulation and by experiment. With the exception of cytochrome c_2, the correlation coefficients (R^2) are all above 0.5, with the highest value of 0.85 being achieved for ubiquitin. Importantly, there is

a good correlation between the calculated and observed average order parameters for each protein (Figure 1). This provides a solid foundation for exploring the microscopic basis for fast methyl-bearing side motion as a proxy for conformational entropy.

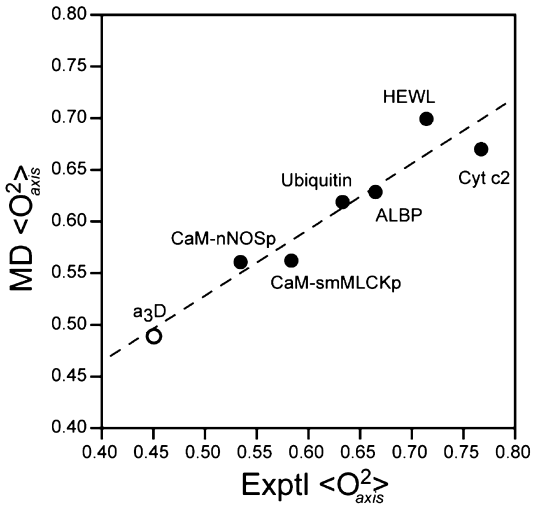


Figure 1. Comparison of the average experimentally determined L-S squared generalized order parameters of the methyl group symmetry axis ($<O^2_{axis}>$) with that calculated from the MD simulations. The experimental average includes all available data. With the exception of α_3D, the MD average includes all methyl groups. In the case of α_3D, the experimentally accessible sites are compared directly with the MD average of those sites to avoid an apparent artifact of limited experimental sampling. Linear regression yields an excellent correlation (R^2 = 0.92) with slope of 0.64 ± 0.09 and intercept of 0.21 ± 0.05. Forcing the fitted line through an intercept of zero yields a slope of 0.92 with a slightly lower R^2. Individual site-to-site correlation plots for each protein are provided in the Figure S2. The correlation coefficients and the slope values for each protein are summarized in Table 2.

Correlation of Methyl-Bearing Side Chain Motion and Rotamer Entropy. A fundamental aspect of the empirical "entropy meter" approach outlined by Marlow and co-workers is the need for a quantitative linkage between the motional averaging of the methyl group symmetry axis and the underlying conformational entropy.[10]

To examine this question, the rotamer population distributions of all the methyl containing amino acids were calculated for each of the protein simulations. The corresponding side-chain rotamer entropies were calculated as described by eq 3. The resulting entropies are plotted against the corresponding calculated methyl order parameters in Figure 2. There is an excellent linear correlation between the rotamer entropy (S_b) of each methyl containing amino acid when normalized by the number of side-chain torsion angles (N_χ) in that residue and the respective methyl side-chain order parameters calculated from the MD simulations. Moreover, the distributions from the different proteins overlay each other, strongly suggesting that the empirical scaling of motion to the corresponding entropy should be universal. Correlations of rotamer entropy and O^2_{axis} parameters for the individual proteins are provided in the Figure S3 and summarized in Table S1.

In principle, the entropic content represented by the O^2_{axis} parameter could arise not only from the distribution between rotamer wells but also from the distribution within each

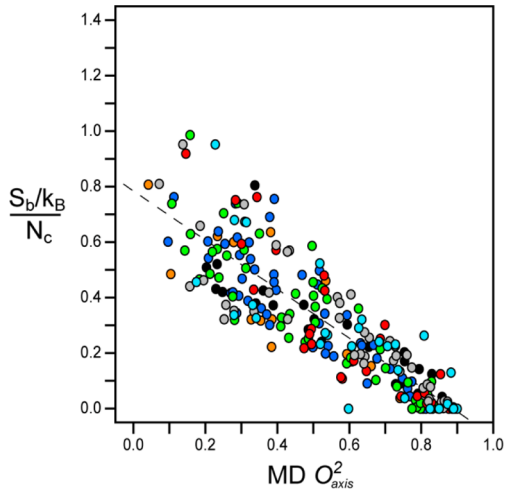


Figure 2. Correlation of the methyl rotamer entropy versus O^2_{axis} from MD simulations. Normalized entropy S_b/N_χ given for every side-chain methyl probe. The correlation was highly linear (R^2 of 0.77) with a slope of −0.88 ± 0.03 and an intercept of 0.78 ± 0.02. The different proteins are represented as follows (all solid circles): (black) ubiquitin, (orange) α_3D, (blue) calmodulin-smMLCKp, (turquoise) calmodulin-nNOSp, (red) hen egg white lysozyme, (light green) adipocyte lipid binding protein, and (gray) cytochrome c_2.

rotamer well and the associated vibrational entropy.[5,44,45] With this in mind, for each methyl group we examined the fine distribution of torsion angles using a resolution of 3°. The width of a rotameric well is characterized by the root-mean-square deviation (rmsd) of its χ-angles. Weighting the width of each well of a given torsion by its overall occupancy gives the mean width. This width is shown as a function of the calculated O^2_{axis} parameter for the corresponding residue in Figure 3. The within-well distributions of non-methionine rotamer angles is largely insensitive to values of O^2_{axis} below ~0.8. At higher O^2_{axis}

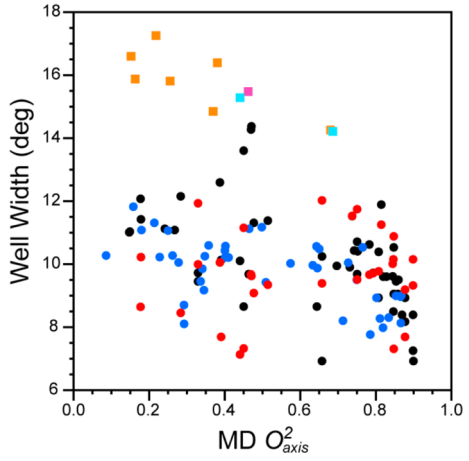


Figure 3. Correlation of the within-well width distribution of side-chain torsion angles with the directly connected methyl group symmetry axis L-S squared order parameters calculated from MD simulations. The width of the distribution is given by the rmsd protein and residue types indicated by solid circles (black) ubiquitin − (Leu, Ile, Val, Thr), (blue) calmodulin-smMLCKp − (Leu, Ile, Val, Thr), and (red) hen egg white lysozyme − (Ile, Leu, Val, Thr). Methionines are indicated by solid square symbols with (orange) from the CaM complexes, (pink) from ubiquitin, and (turquoise) from hen egg white lysozyme.

dx.doi.org/10.1021/ja405200u | J. Am. Chem. Soc. 2013, 135, 15092−15100

values some narrowing of the within-well distribution is indicated. Met methyl groups behave somewhat differently: Within-well widths are modestly correlated with O^2_{axis} (R^2 = 0.30), suggesting that the energy well governing the directly connected torsion angle of the Met methyl group is somewhat plastic.

Given the rarity of Met residues, the generality and origin of this remains to be established. Overall, the high correlation between the normalized rotameric entropy of methyl containing amino acids and the O^2_{axis} parameters of the attached methyl groups indicates that methyl probes serve as excellent reporters of the conformational entropy of methyl containing amino acids in proteins. Two conclusions can be drawn. First, since the residues of seven quite different proteins have indistinguishable correlations between rotamer entropy S_b and the O^2_{axis} parameter, there appears to be a universal relationship between these two quantities. Second, the O^2_{axis} parameter appears to be determined primarily by between-well rotamer transitions.

Though there is significant variation in the normalized rotamer entropy associated with a specific value of O^2_{axis} on a per-residue basis, particularly at lower order parameters (Figures 2 and S3), this variation is much smaller when considering the average behavior over the entire protein. The average normalized rotamer entropy shows an excellent correlation (R^2 = 0.91) with the average O^2_{axis} parameter (Figure 4). This suggests that the NMR O^2_{axis} parameter and the normalized rotamer entropy are effectively transforms of each other and follow a simple protein-independent linear relationship. Thus, though our current ability to accurately simulate individual experimental order parameters is limited (Table 2 and Figure S2), the accuracy of the corresponding derived entropy greatly improves when one uses averages over the entire molecule (Figures 4 and S4) and one is not trying to simulate order parameters but has known values from NMR.

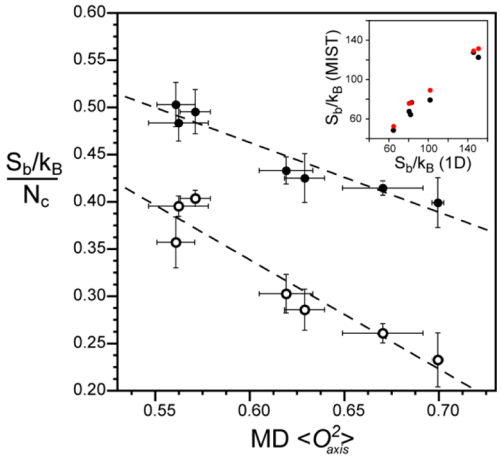


Figure 4. Dynamic proxy of methyl groups is an excellent reporter of both methyl and total side-chain rotameric entropy. (open circle) The normalized methyl rotameric entropy for each protein is calculated as the summation of S_b for individual methyl-bearing amino acids divided by the number of associated rotamer angles (N_χ). (black solid circle) The total rotameric entropy for each protein is calculated as the summation of S_b for all residues and is normalized by the respective total number of rotamer angles (N_χ). The average methyl O^2_{axis} parameter for all methyl-bearing residues including Ala is that obtained from MD simulations. A very high linear correlation is observed for both methyl-side-chain rotamer entropy [slope = −1.16 ± 0.17, R^2 = 0.90] and total rotamer entropy [slope = −0.74 ± 0.10, R^2 = 0.91]. The inset shows the correlation of the uncorrected entropy with the entropy corrected for correlated motions using (red solid circle) second-order and (black solid circle) third-order MIST estimates of inter-residue correlations. Both correlations are highly linear with a slope of −0.89 (R^2 = 0.99) and slope = −0.83 (R^2 = 0.96) for second- and third-order MIST calculations, respectively.

The Essence of the "Entropy Meter". A critical assumption in the use of the dynamical proxy of methyl group motion for conformational entropy of the entire protein is that the motion of the methyl group is sufficiently coupled to its surroundings to faithfully report on local disorder.[10] To examine this central issue, we calculated the total side-chain rotamer entropy from the probability distributions for all side-chain torsion angles. The linear correlation between the dynamical proxy (methyl group dynamics) and total conformational entropy is excellent (Figure 4). Thus averaged over a protein, the methyl order parameters report well on the total side-chain entropy. This can be expressed using the remarkably simple equation:

$$\frac{\sum S_b}{\sum N_\chi} = 0.91 - 0.74\langle O^2_{axis}\rangle \quad \text{MD-uncompensated} \tag{9}$$

which, as noted, is uncompensated for correlated states. Thus far the rotamer entropy has been calculated from entirely intraside-chain rotamer pdfs. Correlations between χ angles within each residue are accounted for by the exhaustive enumeration of individual side-chain torsion angle combinations (see Methods section). However, the effect of inter-residue correlations between torsional motions is absent. Such correlations will result in lowering of the derived protein conformational entropy. In order to quantify the effect of such correlations, doublet and triplet torsion angle distributions of side-chains in all the proteins were accumulated and analyzed. The contributions of inter-residue correlations to the total entropy were calculated using the MIST formulation of Tidor and co-workers.[32,46] The proportional degree of reduction in total conformational entropy is very similar across the set of proteins studied. Both second- and third-order corrections for inter-residue correlated motion gave resulting entropies that showed excellent linear correlations with the uncorrected entropy (inset, Figure 4). Moreover, the correction from third-order (17%) is not much higher than for second order (11%) indicating that the bulk of the correction due to all orders of correlation is captured by these lower order, calculable effects. This leads to a simple correction to eq 1 so that one can obtain an estimate of the total side-chain entropy of a protein, including both methyl- and nonmethyl-containing residues, intra- and inter residue correlation effects from the following equation:

$$S^{tot}_{sc} = 0.83 \sum N_\chi (0.91 - 0.74\langle O^2_{axis}\rangle) \quad \text{MD} \tag{10}$$

which is corrected for inter-residue correlations.

In summary, the side-chain rotamer entropy and methyl order parameters are closely related measures of the dynamics of methyl-bearing side chains. In addition, the major determinants of side-chain entropy changes are shifts in rotamer populations, which are visible to NMR. Moreover, when averaged over the whole protein, methyl order parameters report with identical fidelity on all side-chain motions. Thus NMR measurements of average methyl order

parameter values can be leveraged into estimates of total side-chain entropy.

Reformulation of the "Entropy Meter". The original construction of the conformational "entropy meter" based on the dynamical proxy of fast methyl group motion was admittedly crude.[6,10] This was especially true on how to project the measured motion of methyl groups across the protein. A simple residue-weighting scheme was used. Briefly, the approach rested on relating what could be easily measured (total binding entropy via isothermal titration calorimetry) or calculated (solvent entropy via changes in accessible surface area) to the dynamical proxy. The approach assumed that additional sources of entropy, such as rotational-translational entropy or undocumented entropy were constant across a family of protein−ligand interactions. It was further assumed that measured changes in methyl-group motion were linearly related to changes in local motion (disorder) that not only reflected the methyl-bearing side chain but also that of the surrounding side chains. Linear correlation between side-chain motion and entropy was suggested by simple models[8,9,47] and by the MD simulations presented here. These considerations led to the following formulation of the "entropy meter":[10]

$$(\Delta S_{tot} - \Delta S_{solv}) = m[(n_{res}^{prot})\langle \Delta O_{axis}^2 \rangle^{prot} + (n_{res}^{lig})\langle \Delta O_{axis}^2 \rangle^{lig}] + (\Delta S_{RT} + \Delta S_{other}) \quad (11)$$

where ΔS_{tot}, ΔS_{solv}, ΔS_{conf}, ΔS_{RT}, and ΔS_{other} are the changes in total system, solvent, conformational, and rotational-translational entropies and otherwise undocumented entropy, respectively. As noted in detail elsewhere,[6] the two protein systems where this approach was taken give decidedly different conversion constants (m). Though it may indeed be possible that different proteins have qualitatively different dynamics, it does seem puzzling.[6] We resolve this below.

The ability to recapitulate semiquantitatively experimentally derived order parameters via MD trajectories allows for the rapid exploration of the generality of slight reformulations of the experimental "entropy meter." The computational results described here suggest a modest refinement. Due to the excellent correlation between measures of angular disorder (O_{axis}^2) and the conformational entropy normalized by the number of involved torsion angles (Figure 4), we reanalyzed the empirical calibration of conformational entropy from the study on calmodulin and its binding partners[10] and cAMP activated CAP and its mutants binding to DNA.[17] Now the NMR order parameter derived proxy is given by

$$N_\chi^{prot}\left[\langle O_{axis}^2\rangle_{complex}^{prot} - \langle O_{axis}^2\rangle_{free}^{prot}\right] + N_\chi^{lig}\left[\langle O_{axis}^2\rangle_{complex}^{lig} - \langle O_{axis}^2\rangle_{free}^{lig}\right]_{experiment} \quad (12)$$

where N_χ^{prot} and N_χ^{lig} are the total number of side-chain torsion angles in the protein and ligand, respectively. Using this formulation of the dynamical proxy, a single linear calibration line applies to both the CaM and CAP protein complexes (Figure 5). This assumes that the difference in changes in rotational-translational and 'other' entropy between the two systems is negligible (see eq 11).

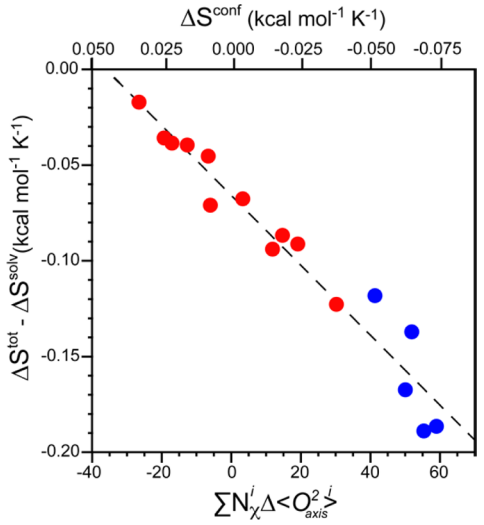


Figure 5. Calibration of the experimental dynamic proxy for protein conformational entropy. The points corresponding to both wild-type and mutants of CAP binding to DNA (red)[17] and CaM binding to its partners (blue)[10] follow the same empirical relationship when correlated using the average change in methyl order parameter weighted by the total number of torsion angles. An excellent linear correlation is seen (R^2 = 0.95) suggesting a common relationship between the conformational entropy of proteins and their respective change in dynamics as represented by the methyl groups. A slope of −0.0018 ± 0.0001 kcal mol^{-1} K^{-1} $\sum N_\chi^{-1}$ provides a conversion between the dynamical proxy and conformational entropy, as illustrated by the top abscissa axis.

DISCUSSION

Here we have employed MD simulations to investigate the validity of several assumptions that are fundamental to the idea of using dynamical information gathered from NMR relaxation in methyl groups of amino acid side chains as a quantitative proxy for conformational entropy. Since the pioneering introduction of MD simulations for comprehensive studies of internal protein motion several decades ago,[48] the capabilities and accuracy of such simulations have expanded greatly with increases in computational power and significant improvements in various aspects of the force fields employed.[19] Perhaps surprisingly, the accuracy of MD simulations has not been extensively tested in the ps−ns time regime against the sizable database of side-chain order parameters obtained by solution NMR relaxation methods.[18−20] Using the NAMD implementation of the CHARMM27 potential in a largely "out of the box" fashion, reasonable "site-to-site" agreement with experiment. These results compare favorably with evaluations of various AMBER potentials employing calbindin and ubiquitin as test proteins.[49,50] More importantly, we find excellent on average agreement between experimental L-S squared generalized order parameters and those extracted from MD trajectories. There exists a large site-to-site variance between experiment and simulation that remains to be explained and corrected. However, the overall agreement with experiment is sufficiently good to promote use of simulations to illuminate critical aspects of the use of the dynamical proxy as a means to quantify conformational entropy.

A central feature of the dynamical proxy to act as a measure of the distribution of states is the notion that motion between states can be linearly related to the underlying conformational entropy.[7] This was initially promoted by examination of a number of simple motional models.[8,9,47] The MD studies of the seven protein systems described here fully reinforce this expectation. There is a robust linear relationship observed

dx.doi.org/10.1021/ja405200u | J. Am. Chem. Soc. 2013, 135, 15092−15100

between the calculated rotamer entropy of each methyl-bearing side chain, normalized by the number of associated χ torsion angles, and its respective methyl order parameter for each protein. This is perhaps not unexpected given that the motion of terminal methyl groups arises from motion about all torsions connecting it to the backbone. Indeed, this type of motion is mirrored in the product of the corresponding generalized order parameters.[29,51] This indicates that methyl groups are excellent reporters of their own side-chain conformational entropy. Importantly, this relationship is uniform across all the proteins studied here suggesting that it is general (Figure 2). Further evidence for this generality comes from a recent and parallel analysis of apo HIV protease simulations.[52] Glass et al.[52] also find a linear dependence of methyl group entropy on O^2 over the range 0.1−0.8, with a slope, when expressed in dimensionless units, very similar to that of Figure 2. It is important to note in this regard that the motions detected by the NMR relaxation discussed here are restricted to time scales faster than overall tumbling of the macromolecule,[29] which is on the order of 10 ns or less for the proteins examined in this study. Interestingly, the MD simulations exceed this by an order of magnitude. Thus, the high correlation suggests that the ensemble of states that is experimentally accessible accurately reflects that sampled on longer time scales. It is important to note that all methyl-bearing side chains, except for Thr, show extensive rotamer interconversion during the trajectories. In principle, it is possible that longer time scale motion may interconvert states that are not similarly sampled in the time regime relevant to the generalized order parameter obtained by classical NMR relaxation.[53] Fortunately, the high empirical correlation found (Figure 5) suggests that the possible distortion of the relationship between local entropy and the generalized order parameter due to slow rotamer interconversion is small.

In the "entropy meter" treatment of Marlow and coworkers,[10] it is assumed that the motion of the experimentally accessible methyl group not only reflects the motional disorder of the entire methyl-bearing side chain but that it also linearly reflects motion (disorder) of the surrounding nonmethyl-bearing side chains. This coupling of motion is clearly evident in the MD simulations where a remarkably robust linear relationship exists between the total normalized rotamer entropy of a protein and the appropriately weighted change in effective amplitude (i.e., the angular disorder represented by the L-S O^2_{axis} parameter) of fast methyl group motion. The normalization by N_χ provides an unbiased view of the relationship between rotamer entropy, expressed now as an intensive quantity, and methyl dynamics for different length side-chains e.g. methionine versus valine, etc. Furthermore, this linear relationship is valid for the entire range of order parameters, i.e., from 0 to 1 irrespective of the type of methyl side chain, i.e., the relative contribution to entropy from a residue with a given order parameter depends only on $<O^2_{axis}>$ and N_χ not the type of side chain.

The microscopic origins of this rotamer entropy and its subsequent linear relationship to methyl side-chain order parameters are intrinsically related to the population distribution between the different energy states of the side-chain, i.e., different rotamer wells. The relationship is largely insensitive to the distributions within a rotamer well, except for perhaps methionine (Figure 3). In principle, the total side-chain conformational entropy arises from the combined effect of both transitions between rotamer wells ('conformational entropy') and distributions within a rotamer well ('vibrational entropy').[5] However, differences in entropy that distinguish more dynamic or less dynamic residues or that distinguish more dynamic proteins, such as α_3D from less dynamic ones such as HEWL arise almost entirely from the conformational component. This insensitivity of the torsional distributions within rotamer wells is attributable to the fact that the energies involved in stabilization of protein structure and in binding ligands are rather modest. They are apparently sufficient to modulate softer modes of motion, such as transitions between rotamers, but not enough to greatly affect vibrational contributions, such as motion within torsion wells nor indeed the much stiffer vibrational contributions from bond angle and bond stretch motions.

A logical consequence of the coupling of motion is the need to consider how correlation might affect the quantification of conformational entropy via a dynamical proxy. Although exhaustive enumeration of rotamers within a side chain is computational feasible and was carried out explicitly, complete inter-residue correlations are not computationally tractable but were estimated here using the MIST algorithm.[32] These calculations indicate that only a small linear scaling of the uncorrected entropy is required. This picture of rather restricted, local effects of correlation is consistent with other recent studies of conformational entropy in proteins.[54−56] Given the generally linear relationship revealed by simulations, the experimental "entropy meter" should represent the true conformational entropy change as the effects of correlated motion are encapsulated within the empirical calibration, assuming the absence of significant solvent−protein coupling (see eq 3). As emphasized above, the conformational entropy calculated from MD derived $<O^2>$ values alone would be insufficiently accurate. In contrast, the empirical entropy meter based on the measured methyl-group dynamical proxy would more accurately relate side-chain motion to the underlying conformational entropy and does so through a remarkably simple relationship:

$$\Delta S^{tot}_{sc}/k_B = 0.92 \sum N_\chi \Delta\langle O^2_{axis}\rangle \quad \text{experiment} \tag{13}$$

which is derived from the slope of the correlation plot of Figure 5 converted to units of k_B.

CONCLUSIONS

Using MD simulations as a guide, we have resolved the apparent discrepancy between the applications of a dynamical proxy for conformational entropy to the binding of ligands by calmodulin and the catabolite activation protein. A simple normalization based on the number of degrees of freedom (torsion angles) is sufficient. The MD simulations also suggest that the "entropy meter" derived from this approach may be generally applicable. Further experiments are required to confirm this. Nevertheless, the apparent robustness of the dynamical proxy for conformational entropy from the perspective of both simulation and experiment reinforces the conclusion that conformational entropy can indeed play a significant and potentially determining role in associations involving proteins.[6] Looking forward, these results promote a more exhaustive examination of the quantitative role of conformational entropy in the free energy governing protein−ligand associations. This may be particularly important in the context the interaction of proteins with unnatural man-

made pharmaceuticals where the influence of conformational entropy remains largely unknown.[57]

■ ASSOCIATED CONTENT

Supporting Information

Convergence of simulation trajectories. Site-by-site comparison of simulated methyl side-chain order parameter to both experimental order parameters and methyl side-chain entropy. Summary of statistics for the effect of correlated motion on side-chain entropy calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

wand@mail.med.upenn.edu
sharpk@mail.med.upenn.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Supported by NIH grant GM102447 a grant from the Mathers Foundation and the National Resource for Biomedical Supercomputing and the Pittsburgh Supercomputing Center through NIH award RC2GM093307 to CMU through the NRBSC.

■ REFERENCES

(1) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.
(2) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. *Annu. Rev. Biophys.* **2012**, *41*, 585−609.
(3) Cooper, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2740−2741.
(4) Cooper, A.; Dryden, D. T. F. *Eur. Biophys. J. Biophys. Lett.* **1984**, *11*, 103−109.
(5) Karplus, M.; Ichiye, T.; Pettitt, B. M. *Biophys. J.* **1987**, *52*, 1083−1085.
(6) Wand, A. J. *Cur. Opin. Struct. Biol.* **2013**, *23*, 75−81.
(7) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672−1699.
(8) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647−2650.
(9) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369−382.
(10) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. *Nat. Chem. Biol.* **2010**, *6*, 352−358.
(11) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Formankay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.
(12) Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. *J. Am. Chem. Soc.* **1995**, *117*, 11536−11544.
(13) Ishima, R.; Petkova, A. P.; Louis, J. M.; Torchia, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 6164−6171.
(14) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439−6448.
(15) Tugarinov, V.; Kay, L. E. *Biochemistry* **2005**, *44*, 15970−15977.
(16) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325−329.
(17) Tzeng, S.-R.; Kalodimos, C. *Nature* **2012**, *488*, 236−240.
(18) 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. B.; Wriggers, W. *Science* **2010**, *330*, 341−346.
(19) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H. F.; Shaw, D. E. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.
(20) Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(21) Scouras, A. D.; Daggett, V. *Protein Sci.* **2011**, *20*, 341−352.
(22) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kale, L. V.; Skeel, R. D.; Schulten, K. *Int. J. Supercomputer Appl. High Perform. Comput.* **1996**, *10*, 251−268.
(23) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(24) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(27) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(28) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Chem. Phys.* **1977**, *23*, 327−341.
(29) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(30) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.
(31) Nicholson, L. K.; Kay, L. E.; Baldisseri, D. M.; Arango, J.; Young, P. E.; Bax, A.; Torchia, D. A. *Biochemistry* **1992**, *31*, 5253−5263.
(32) King, B. M.; Tidor, B. *Bioinformatics* **2009**, *25*, 1165−1172.
(33) Walsh, S. T. R.; Lee, A. L.; DeGrado, W. F.; Wand, A. J. *Biochemistry* **2001**, *40*, 9560−9569.
(34) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C. H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *Biochemistry* **1998**, *37*, 7965−7980.
(35) Flynn, P. F.; Urbauer, R. J. B.; Zhang, H.; Lee, A. L.; Wand, A. J. *Biochemistry* **2001**, *40*, 6559−6569.
(36) Moorman, V. R.; Valentine, K. G.; Wand, A. J. *Protein Sci.* **2012**, *21*, 1066−1073.
(37) Lee, A. L.; Flynn, P. F.; Wand, A. J. *J. Am. Chem. Soc.* **1999**, *121*, 2891−2902.
(38) Vijaykumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(39) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486−5491.
(40) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1992**, *257*, 1251−1255.
(41) Maenaka, K.; Matsushima, M.; Song, H.; Sunada, F.; Watanabe, K.; Kumagai, I. *J. Mol. Biol.* **1995**, *247*, 281−293.
(42) Xu, Z. H.; Bernlohr, D. A.; Banaszak, L. J. *J. Biol. Chem.* **1993**, *268*, 7874−7884.
(43) Benning, M. M.; Wesenberg, G.; Caffrey, M. S.; Bartsch, R. G.; Meyer, T. E.; Cusanovich, M. A.; Rayment, I.; Holden, H. M. *J. Mol. Biol.* **1991**, *220*, 673−685.
(44) Chen, W.; Gilson, M. K.; Webb, S. P.; Potter, M. J. *J. Chem. Theory Comput.* **2010**, *6*, 3540−3557.
(45) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609−1618.
(46) King, B. M.; Silver, N. W.; Tidor, B. *J. Phys. Chem. B* **2012**, *116*, 2891−2904.
(47) Lee, A. L.; Sharp, K. A.; Kranz, J. K.; Song, X. J.; Wand, A. J. *Biochemistry* **2002**, *41*, 13814−13825.
(48) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−590.
(49) Long, D.; Li, D. W.; Walter, K. F.; Griesinger, C.; Bruschweiler, R. *Biophys. J.* **2011**, *101*, 910−915.
(50) Showalter, S. A.; Johnson, E.; Rance, M.; Bruschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146−14147.
(51) Wittebort, R. J.; Szabo, A. *J. Chem. Phys.* **1978**, *69*, 1722−1736.
(52) Glass, D. C.; Krishnan, M.; Smith, J. C.; Baudry, J. *J. Phys. Chem. B* **2013**, *117*, 3127−3134.

(53) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.
(54) Li, D. W.; Showalter, S. A.; Brüschweiler, R. *J. Phys. Chem. B* **2010**, *114*, 16036−16044.
(55) DuBay, K. H.; Geissler, P. L. *J. Mol. Biol.* **2009**, *391*, 484−497.
(56) Das, A.; Chakrabarti, J.; Ghosh, M. *Biophys. J.* **2013**, *104*, 1274−1284.
(57) Marshall, G. R. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 3−8.