Published in final edited form as:

Proteins. 2015 May ; 83(5): 922–930. doi:10.1002/prot.24789.

On the relationship between NMR-derived amide order parameters and protein backbone entropy changes

Kim A. Sharp*, Evan O'Brien, Vignesh Kasinath, 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

Abstract

Molecular dynamics simulations are used to analyze the relationship between NMR-derived squared generalized order parameters of amide NH groups and backbone entropy. Amide order parameters (O^2_{NH}) are largely determined by the secondary structure and average values appear unrelated to the overall flexibility of the protein. However, analysis of the more flexible subset ($O^2_{NH} < 0.8$) shows that these report both on the local flexibility of the protein and on a different component of the conformational entropy than that reported by the side chain methyl axis order parameters, O^2_{axis} . A calibration curve for backbone entropy vs. O^2_{NH} is developed which accounts for both correlations between amide group motions of different residues, and correlations between backbone and side chain motions. This calibration curve can be used with experimental values of O^2_{NH} changes obtained by NMR relaxation measurements to extract backbone entropy changes, e.g. upon ligand binding. In conjunction with our previous calibration for side chain entropy derived from measured O^2_{axis} values this provides a prescription for determination of the total protein conformational entropy changes from NMR relaxation measurements.

Keywords

protein motion; backbone motion; conformational entropy; molecular dynamics; NMR relaxation

Introduction

Entropy (S) is one of the two thermodynamic quantities that determine the affinity in protein-ligand binding reactions, the other term being the enthalpy (H). The various degrees of freedom that contribute to the entropy change upon binding have been identified and a rigorous statistical mechanical formalism that relates them to the binding affinity has been developed.^{1,2} These degrees of freedom comprise the solvent conformations, the loss of translational and rotational freedom of the ligand with respect to the protein, the internal conformations of the ligand, and the internal conformations of the protein. The net entropy

*Correspondence to: A. J. Wand (wand@mail.med.upenn.edu) or K. A. Sharp (sharpk@mail.med.upenn.edu), Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104-6059. Contact information: Professor A. J. Wand, Department of Biochemistry & Biophysics, University of Pennsylvania, 905 Stellar-Chance Laboratories, 422 Curie Blvd, Philadelphia, Pennsylvania 19104-6059, telephone: 215-573-7288, facsimile: 215-573-7288, wand@mail.med.upenn.edu

change, S , results from changes in conformational distributions. The protein conformational entropy term is of particular interest, as it appears to play a controlling role in some binding reactions.^{3–6} To date, however, it has proven very difficult to obtain accurate estimates of the protein contribution from either experimental methods or theoretical calculations. Though calorimetry can provide measurements of the total entropy change there is no direct way to measure individual terms contributing to S . Furthermore protein molecules contain many degrees of freedom, and thus, even in a well-folded and stable state it potentially has a large residual conformational entropy.^{7–9} Significant contributions to S could result from small, hard to quantify, changes in the protein's conformational entropy reservoir upon binding. However, a combination of NMR measurements and molecular dynamics simulations has proven to be particularly synergistic in elucidating changes in protein conformational entropy. We focus here on the measurement, by NMR relaxation methods, of the Lipari and Szabo¹⁰ square of the generalized order parameter of the peptide NH group (O^2_{NH}) and that of the side chain methyl axis (O^2_{axis}). NMR-derived order parameters provide site specific measures of the degree of motion of the NH and methyl group symmetry axis on the ps-ns timescale, respectively. These order parameters are a predominant experimental source of information on protein dynamics at the atomic level. To relate these order parameters to the protein conformational entropy a number of difficulties must be overcome.¹¹ A primary issue from an analytical perspective is defining the functional relationship between the order parameter and the conformational entropy of the corresponding group or residue. A second task is accounting for correlations between motions. Order parameters report directly on the motion of individual groups, but they say nothing about the correlation in motions between the groups, which act to decrease the net entropy of the protein.

Early attempts to relate order parameters to the conformational entropy used models of amide or methyl group motion in a variety of simplified effective potentials, such as harmonic and step potentials.^{8,12–15} Given a model potential, calculation of the conformational probability distribution function (pdf) through statistical mechanics is straightforward. The pdf then provides both S and O^2 , again via standard statistical mechanical relations. Although simplistic, these models were quite informative in revealing the general inverse relationship between order parameter and entropy. Indeed these types of models were able to provide insight into the origin of the temperature dependence of order parameters and the corresponding heat capacity^{13,16–18} and reasonable estimates of the entropy change in binding reactions even when applied in a rather simplistic fashion.³ One conclusion from this work is that most of these effective potentials give very similar *changes* in entropy for a given change in O^2 over the range 0.3–0.7, in spite of qualitative differences in the form of the potential. This suggested that a robust general relationship could exist between O^2 and S for the actual internal motions of a protein, although these motions undoubtedly occur over a more complex energy landscape. In principle this idea could be substantiated by molecular dynamics (MD) simulations where the conformational probability distributions that determine both S and O^2 can be extracted directly. Initially, this approach was hampered due to the inability of MD simulations to satisfactorily reproduce experimental values of O^2 . This was particularly acute for methyl order parameters, where correlations between calculated and measured O^2_{axis} parameters showed R^2 values typically in the 0.3 –

0.4 range.^{15,19,20} While better reproduced because they mostly cluster in a narrow range of 0.85 – 0.95 in secondary structure regions, amide order parameters were also rather poorly reproduced in the critical flexible regions (loops and coils), which often undergo the most change in a binding reaction. Advances in simulation algorithms and longer simulations permitted by faster hardware have substantially improved the agreement with measured O^2_{axis} values. Correlation coefficients of $R^2 = 0.7$ or better are possible for many proteins.^{21–23} In principle, these more accurate simulations now allow one to better estimate the quantitative relationship between O^2 and S .

In relating O^2 to S , MD simulations can also address the contribution due to correlation effects through analysis of two-dimensional and higher dimension conformational pdf's. Correlation effects, however, require considerably more simulation than determination of average quantities such as O^2 , or even individual group entropies. The difficulty is that sampling requirements increase exponentially as one moves to higher order pdf's. However, there is a growing body of evidence that correlation effects in protein motions are rather local.²⁴ This simplifies the task of computing correlation effects, and also means that the reduction in total entropy due to correlations is rather modest,^{25–27} although there are exceptions.²⁸ At the same time, better methods for estimating the effects of correlations from MD simulations have been developed.^{29–33}

Building on these advances in simulation and theory we recently resolved a robust relationship between O^2_{axis} and the total protein side chain entropy.²³ Analysis of MD simulations of seven globular proteins spanning a range of flexibilities revealed several key features of protein dynamics. First, changes in side chain conformational entropy result almost entirely from shifts in rotamer populations (conformational effects), with little contribution from within rotamer well (vibrational) entropy changes.²³ Second, O^2_{axis} values report directly on rotamer population shifts not only from the residue bearing the methyl group but also for neighboring non-methyl bearing residues, which are generally difficult to characterize in this context using NMR relaxation methods. Third, entropy changes scale almost linearly with the number of side chain torsion (χ) angles. The consequence is that measurement of the change in each O^2_{axis} averaged over the entire protein ($\langle O^2_{axis} \rangle$), along with knowledge of the total number of χ angles can be used to extract a quantitative estimate of the change in total side chain entropy S_{SC} including the effect of correlations between side chain motions.

We apply here an analogous strategy to develop a calibration between measured O^2_{NH} values and the protein backbone entropy, accounting for correlations in motions between backbone groups. Then, in order to obtain an NMR-measurement derived estimate of the total protein conformational entropy, we combine the side chain and backbone entropy calibrations, including the effect of correlations between backbone and side chain motions.

Methods

Molecular Dynamics Simulations

Molecular dynamics simulations of the seven proteins listed in Table 1 were carried out with NAMD³⁴ using the CHARMM27³⁵ all-atom parameter set and the TIP3P³⁶ water

potential. Hydrogen atoms were added to the crystallographic structures with VMD³⁷ and 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 smaller proteins, ubiquitin and α₃D, where the solvent layer was at least 6 Å. All simulations were performed using a time step of 2 fs. Bonds to hydrogen atoms were constrained using the SHAKE algorithm.³⁸ A switching distance of 10 Å and cut-off of 12 Å were used for the non-bonded interaction interactions, combined with a Particle Mesh Ewald (PME) summation with a 1 Å grid spacing for long-range electrostatics. Simulations were run at constant temperature and pressure of 1 Atm and 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, calmodulin-smMLCKp complex and calmodulin-nNOSp complex, 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 non-bond cutoff of 14 Å.

Calculation of order parameters

The Lipari-Szabo squared generalized order parameters (O^2)¹⁰ were calculated from MD simulations by overlaying snapshots of the protein from the trajectories using a standard rigid-body alignment procedure on the C_α atoms. For each snapshot, the unit vector along the methyl symmetry axis or the amide NH bond was obtained in terms of its vector components in Cartesian axes, x, y, and z. The O^2 parameter for a given methyl axis or amide group is then calculated using³⁹

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

where $\langle \rangle$ indicates the average over the trajectory.

Calculation of backbone entropy

Backbone Φ and ψ angles for each amino acid were calculated from the MD trajectories and binned in increments of 18° (20 bins per angle). For each amide group, the two dimensional (2D) histogram of its flanking ϕ and ψ angles was accumulated and used to obtain the corresponding 2D joint probability distribution function, $P(\phi, \psi)$. Variations in the ω angle are typically small, and its possible contribution to entropy changes was omitted here. The backbone entropy for each residue was calculated as

$$S_{nh} = -k_B \sum_i \sum_j p(\psi_i, \phi_j) \ln p(\psi_i, \phi_j) \quad (2)$$

where k_B is the Boltzmann constant and the sum is over all conformational bins i and j . For simplicity, all entropies are expressed in dimensionless units, i.e. normalized by k_B , hereon.

Calculation of the effect of inter-group correlated motions on entropy

The effects of correlated motions between ϕ and ψ flanking each amide are fully accounted for by the use of the two-dimensional pdf $P(\phi, \psi)$ in equation 2. To quantify the effect of correlations between other backbone torsion angles, one dimensional (1D) pdfs for all backbone torsion angles were constructed from the trajectories. Similarly 2D pdfs for all other backbone torsion angle pairs, except the flanking pairs, were constructed from the trajectories. The effect of correlated motions was then estimated using the maximum information spanning tree (MIST) method³¹ as follows. The first order (no inter-angle correlation) entropy estimate was obtained as

$$S_k^{1D} = - \sum_{l=1}^{20} P(\alpha_l^k) \ln P(\alpha_l^k) \quad (3)$$

where α_l^k indicates the k 'th ϕ or ψ angle in the l 'th conformation, and S_k^{1D} was summed over all backbone torsion angles to give S_{NH}^{1D} . Then the second order MIST approximation was obtained by computing pair-wise second order mutual information terms (I_2) from the one-dimensional and two-dimensional pdfs as⁴⁰

$$I_2(j, k) = \sum_{l=1}^{20} \sum_{m=1}^{20} P(\alpha_l^j, \alpha_m^k) \ln \frac{P(\alpha_l^j, \alpha_m^k)}{P(\alpha_l^j) P(\alpha_m^k)} \quad (4)$$

From all the I_2 terms the second order MIST expansion is then obtained by selecting a spanning set of I_2 's to maximize

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

where n is the total number of backbone torsion angles. To construct this spanning set, for each i , the index $j(i)$ which gives the largest $I_2(j, i)$ is chosen from the set $j = \{1, 2, \dots, i-1\}$. Then an upper bound estimate for the entropy of the backbone including correlations at second order is

$$S_{bb}^{MIST-2} = S_{nh}^{1D} - I_2^{MIST} \quad (6)$$

To estimate the effect of correlations between backbone and side chain torsion angles, the analogous second order MIST approximation was constructed using 2D pdfs for every backbone ϕ/ψ and side chain χ angle pair, $P(\alpha^i, \chi^j)$. Pdf's for χ angles were constructed as described previously, using a (g^+ , g^- , trans) rotamer classification scheme.²³ Uncertainties in the calculated entropy values were estimated by dividing each 60 ns set of samples into four batches and computing the variance in corresponding quantities between batches. Uncertainties in slope values for entropy vs. order parameter fits were calculated from the regression analysis.⁴¹

Results and Discussion

Table 1 lists the proteins used in this analysis. The seven proteins span a significant range of flexibility or dynamic disorder, as judged by the mean methyl axis order parameter $\langle O^2_{axis} \rangle$, which varies from 0.489 in α_3D to 0.699 in hen egg white lysozyme (HEWL). As we have shown in recent work,²³ MD simulations can now reproduce methyl order parameter values fairly accurately for this set of proteins, with Pearson coefficients of correlation (R^2) greater than 0.6, on a per residue basis. The correlation between measured and simulated O^2_{axis} values averaged over the protein ($\langle O^2_{axis} \rangle$) is even better with $R^2 = 0.92$ for this set of proteins. It is more difficult to assess the accuracy of MD simulations of amide order parameters due to the smaller dynamic range. The majority of measured amide order parameters cluster in a narrow range of 0.8–0.95, usually in the helix and sheet regions, with lower values typically in the turn and loop regions. This overall qualitative behavior is generally seen experimentally.⁴² O^2_{NH} is greater than 0.8 for most residues in helix and sheet regions. Lower values occur in loop regions and at the initiation and termination of helices. The MD simulation protocol used here matches the order parameter values accurately in the secondary structure regions, as well as reproducing the trend of lower order parameters in the loop regions. Similar quality results are obtained for the other six proteins. Although prediction of order parameters is not the focus of this work, the ability to reproduce measured amide order parameters, particularly the pattern of high and low values, allows us to examine the relationship between backbone dynamics and entropy with some confidence.

Since the majority of amide order parameters cluster in a narrow range, if one looks at the average value over an entire protein, $\langle O^2_{NH} \rangle$, it apparently reveals little about the overall flexibility of the protein. There is almost no correlation with the average methyl axis order parameter, $\langle O^2_{axis} \rangle$, which is a reliable measure of protein flexibility (Table 1, Fig. 1; $R^2 = 0.29$). The average value of O^2_{NH} for a protein is dominated by the rigidity of its secondary structure. However, changes in protein backbone dynamics and flexibility are more likely to be detected in the regions that are inherently more flexible, such as loop regions and helix termini, i.e. for those amino acids with lower amide order parameters. To identify this set of order parameters the frequency distribution of O^2_{NH} values over the set of seven proteins was determined (Fig. 2). The values cluster at high values (0.8–0.95), with a long tail down to 0.1. For the purpose of this analysis, a tail region was defined as the lower 30th percentile of values, equivalent to a cutoff of O^2_{NH} less than 0.8. This set shows a larger dynamic range of values than the overall average (Table 1). More interestingly, there is now an excellent correlation between average backbone flexibility in these regions, as measured by $\langle O^2_{NH} \rangle$ for O^2_{NH} less than 0.8, and the overall side chain flexibility, as measured by $\langle O^2_{axis} \rangle$ averaged over all side chain methyl groups. A linear fit gives a slope of 1.05 with $R^2 = 0.94$ (Fig. 1). This ability of backbone order parameters to reveal differences in dynamics between proteins is masked unless the more rigid 70% of amide order parameters is excluded from the analysis.

We next examined whether the information on dynamics provided by amide order parameters with $O^2_{NH} < 0.8$ and that provided by O^2_{axis} is redundant or not. The correlation between the sets with $O^2_{NH} < 0.8$ and O^2_{axis} was analyzed at the residue level. First, every

amide which had $O^2_{NH} < 0.8$ and was a neighbor of a methyl bearing side chain in either direction of the polypeptide chain was identified in the seven proteins, yielding a total of 62 examples. The correlation between the two order parameters for this set was then calculated. Since the side chain lies about equidistant from two flanking amide groups, we also examined the correlation between O^2_{axis} and the average O^2_{NH} for its two flanking amides. Both analyses lead to the same conclusions. Figure 3a shows a plot of O^2_{axis} versus the average of the flanking amides (O^2_{NH*}), identified by residue type. The ‘control’ is given by the data for Ala residues, where the methyl group is rigidly attached to the backbone. For this residue type, the amide and methyl axis order parameters are essentially the same: Linear least squares fit gives $R^2 = 0.9$, with a slope of 0.92 ± 0.1 . For all the other residues, almost all methyl axis values lie on or below the $x = y$ line, as one might expect. With the additional bond distance and torsional degree(s) of freedom separating backbone and methyl group, the latter group has increased possibility to be disordered or flexible. One could also imagine situations where the methyl group is pinned by its surroundings and has less motion than the backbone but Figure 3a indicates that this is rare. Thus the amide O^2_{NH} value effectively provides an upper bound for a neighboring O^2_{axis} . Beyond this, however, there is no correlation. Figure 3a shows that the methyl values span the entire range of the lower triangle defined by $O^2_{axis} < O^2_{NH}$. This implies that the ratio O^2_{axis}/O^2_{NH} should be similarly distributed between 0 and 1. Figure 3b shows that this is indeed the case. Importantly, these observations indicate that backbone and side chain order parameters are sensing different and effectively non-overlapping components of protein dynamics i.e. they are largely decoupled.

Recently we established a quantitative connection between methyl axis order parameters and side chain rotamer entropies.²³ Since the methyl axis order parameters and amide order parameters from flexible backbone regions are probing different components of the protein motion, a separate relationship between amide order parameters and backbone conformational entropy, if any, must be established. Our overall strategy is similar to that used successfully for side chains. Significant backbone conformational changes come primarily from torsional motions around the ϕ and ψ angles. Thus, the probability distribution functions (pdfs) for these angles were extracted from the MD simulations and then the corresponding entropies obtained from the standard Boltzmann expression for entropy (Eq. 1). These entropies were then compared with the corresponding O^2_{NH} values. Given the rigidity of the peptide group changes in flanking ψ and ϕ angles are presumably highly correlated. This was accounted for explicitly through use of the two dimensional (2D) pdf's constructed from the flanking ψ and ϕ of each amide, $P(\psi, \phi)$. Torsion angles were binned in $\delta\psi$ and $\delta\phi$ increments of 18° requiring 20×20 dimension pdf's. The results are shown in Figure 4. Entropies calculated using bin sizes of 25×25 or 30×30 compared to 20×20 were almost identical, with linear fit correlation coefficients of $R^2 = 0.995$ and $R^2 = 0.994$ respectively, and slopes of 0.99 and 1.02 respectively. An average variation of 2% or less in calculated entropy with bin size indicates an appropriate choice of bin size—sufficiently narrow to capture the torsion angle distributions, but wide enough to contain statistically robust sampling.^{43,44} The distributions of S_{NH} vs O^2_{NH} for each peptide group from the seven proteins are largely indistinguishable. For a given value of amide order parameter, the backbone entropies cluster in a restricted band. This is similar to the

dependence of side chain entropy on O_{axis}^2 .²³ Unlike with side chain methyl groups, however, the relationship between O_{NH}^2 and S is not linear. In fact the overall shape of the curve strongly resembles that expected from motion in a stiff, single well potential, as predicted from a number of different theoretical models.^{8,12,13,15} Indeed, the S/k_B vs. O^2 curve from a model harmonic well potential fits the data quite well (Fig. 4). Previous theoretical work shows that for motion in a harmonic or near harmonic potential the entropy should to a good approximation vary linearly with $\ln(1-O^2)$.^{15,17} Data from our MD simulations bears this out (Fig. 4). A linear least squares fit for the data analyzed at the individual amide level gives

$$S_{NH}/k_B = 0.97 \ln(1 - O_{NH}^2) + 3.7 \quad (7)$$

with the entropy given in dimensionless units. A correlation of $R^2 = 0.75$ is obtained. A much better correlation ($R^2 = 0.89$) with similar slope of 1.11 ± 0.17 is obtained if S_{NH} and $\ln(1 - O_{NH}^2)$ for O_{NH}^2 less than 0.8 are first averaged over each of the seven proteins and the resulting averages fitted (Fig. 4), giving:

$$\langle S_{NH} \rangle / k_B = 1.11 \langle \ln(1 - O_{NH}^2) \rangle + 3.8 \quad (8)$$

Figure 4 shows that for amide groups, as with methyl bearing side chains, the NMR derived order parameter is expected to correspond, on average, to a well-defined entropy. Though there is some dispersion in entropy for the value of O_{NH}^2 on a per residue basis these deviations are largely averaged out upon summing the contributions from all residues in the determination of the total entropy of a protein, which is of primary interest. In summary, the calibration line or NMR-based “entropy meter” given by Eq. 8 relates the backbone conformational entropy to a measurable dynamic parameter, namely the NMR derived generalized order parameter of the amide group, O_{NH}^2 .

The calibrations in Eqs. 7 and 8 explicitly account for the large degree of correlation in fluctuations of flanking ψ , ϕ angles due to the rigidity of the peptide group that manifested in, for example, crankshaft motions. Such correlations have the effect of reducing the entropy relative to that obtained using just 1D pdfs (i.e. assuming no correlations). Two other sources of correlated motions are possible: 1) Between any sets of non-flanking backbone torsion angles, i.e. inter-residue motion correlations. 2) Between backbone and side chain torsional motions. We examine each of these in turn using the MIST method at second order.³¹ In principle a full analysis of correlation effects would require accumulation of pdf's of dimension 3 and higher, up to $2N_{res}$ dimensions, the total number of torsions. However, even at third order for a moderate size protein like calmodulin with about 300 backbone torsion angles there would be ~ 27 million three dimensional pdf's of 8000 terms each. This combinatorial explosion renders a brute force approach impractical. However, the MIST method at second order is very tractable and, importantly, gives an *upper* bound to the entropy. Figure 5 shows the results of this analysis for backbone-backbone correlations. To confirm that the simulations were of sufficient length to give converged estimates of the entropy including correlation effects, calculations were performed for two proteins,

ubiquitin and CaM-smMLCKp for simulations lengths of 20ns, 40ns, 60ns and 80ns. For ubiquitin, the total entropy, including backbone entropy, side chain entropy, and correlations within and between these two components, varied by about 1% for simulations of 40ns or more. Similarly, for CaM-smMLCKp the corresponding variation was about 2%. We note that total entropy will converge better in these type of calculations than individual residue contributions, because of the effective averaging over the entire protein.

Considering the dependence of the total backbone entropy on the number of residues, with just the flanking (ψ, ϕ) correlations included, there is a remarkably good proportionality. A linear fit to the data constrained to pass through the origin gives a slope of $S_{BB} = 1.93N_{res}$ with $R^2 = 0.88$. Inclusion of inter-residue correlations also gives a good fit, with a somewhat smaller slope of $S_{BB} = 1.69N_{res}$ with $R^2 = 0.90$. Correlations reduce the entropy by a factor of 0.87 ± 0.07 or 13%. Analysis of the data with and without the flanking angle correlations shows about an 11% effect on the entropy, also proportional to the number of residues. In other words, the reduction in backbone entropy from correlation in the two flanking (ψ, ϕ) angle motions due to the peptide group rigidity is about the same as that from all other backbone torsion angle correlations combined.

Figure 5 also shows the effect of correlations between backbone and side chain torsional motions. The total conformational entropy of side chains plus backbone, excluding any correlation between them is again proportional to the number of residues, with $S_{BB} + S_{SC} = 2.7N_{res}$ with $R^2 = 0.95$. Inclusion of backbone-side chain correlations gives $S_{BB} + S_{SC} = 2.45N_{res}$ with $R^2 = 0.87$, reducing the entropy by a factor of 0.91 ± 0.06 or just 9%. The small degree of backbone-side chain correlation confirms the conclusion drawn from Figure 3 that amide and methyl axis order parameters report on distinct, largely independent components of the protein dynamics.

We previously found that the total side chain entropy, including all correlation effects up to third order, was to a good approximation also proportional to the number of residues.²³ At least for the test set of seven proteins used here, this shows that when considering conformational entropy contributions of a protein in aggregate, the backbone entropy, the side chain entropy, the correlations between backbone groups, the correlations between side chains, and the correlations between backbone and side chains all scale *linearly* with the size of the protein. It is noted that the test set of proteins is diverse in terms of size, folding motif, secondary structure composition, overall flexibility, and the presence (or absence) of various types of cofactors. A plausible mechanism to generate proportionality is if each residue interacts dynamically with a limited region of its surroundings; simply put, with a fixed number of neighbors. Residues would then be blind to the size of protein they are in and so make, on average, a constant per-residue contribution to the entropy. Spatially limited correlation effects of this kind are consistent with recent theoretical^{25,27} and experimental²⁴ investigations of this issue.

Impact on the Experimental “Entropy Meter”

Though molecular dynamics simulations have made great strides in accuracy over the past decade it remains very desirable to have experimental confirmation. By the same token, the

semi-quantitative nature of current simulations provide a powerful means to efficiently explore potential targets for experimental investigation and to suggest experimental strategies. This is particularly true for recent efforts to employ NMR-based measures of internal motion as a proxy for the residual conformational entropy of proteins.^{11,23} The present results suggest that one can combine measurements of backbone and methyl-bearing side chain motion in a rather simple way to obtain estimates of the total conformational entropy of the protein. To avoid a number of issues, this NMR-based “entropy meter” is constructed in the context of changes in motion that are interpreted as changes in entropy.^{5,23} Combining our previous calibration for side chains²³ with Equation (8) for the backbone, our prescription for extracting the total change in conformational entropy of a protein resulting from a change of state (e.g. binding of a ligand) from the NMR relaxation data is

$$\Delta S/k_B = 0.91 \{0.87N_{\text{res}} 1.11 \langle \delta \ln(1 - O_{NH}^2) \rangle + 0.83N_{\chi} \cdot -0.74 \langle \delta O_{axis}^2 \rangle\} \quad (9a)$$

$$\Delta S/k_B = 0.88N_{\text{res}} \langle \delta \ln(1 - O_{NH}^2) \rangle - 0.56N_{\chi} \langle \delta O_{axis}^2 \rangle \quad (9b)$$

where N_{res} is the total number of residues and N_{χ} is the total number of side chain chi angles. $\langle \delta O_{axis}^2 \rangle$ is the measured average change in methyl axis order parameters, $\langle \delta \ln(1 - O_{NH}^2) \rangle$ is the corresponding quantity for measured changes in the amide order parameter. The factors of 0.83, 0.87 and 0.91 account for the reduction in entropy due to correlations of side chain to side chain, backbone to backbone, and backbone to side chain torsional motions, respectively. The constants 1.11 and -0.74 are derived from the dependence of entropy on order parameter extracted from the molecular dynamics simulations. The constants in Equation 9a have been left separate, rather than multiplied together, to emphasize the inclusion of all three sources of correlations. They have been collapsed in Equation 9b into two multiplicative factors, 0.88 ± 0.25 for the amide data and -0.56 ± 0.3 for the methyl data, to correspond to what would be in principle be measurable by experiment. In addition, the dependence on N_{χ} , the total number of side chain χ -angles from methyl bearing and non-methyl bearing residues alike, signifies that the calibration accounts for the fact that changes in O_{axis}^2 report on entropy changes from all side chains.²³ These results promote an experimental avenue to confirm the validity of the approach and empirically determine the coefficients relating the dynamical proxy to conformational entropy. Such studies are in progress and will be reported elsewhere.

To put the entropy changes described by equation 9b in context, we compare to recent estimates by Baxa et al. They also used a combination of molecular dynamics simulations, NMR data and correlation effects calculated using the mutual information expansion.^{32,33} Baxa et al. considered the backbone and side chain entropy loss upon folding from an ensemble of unfolded states, and found that the total entropy loss was considerably smaller than previous estimates, with a loss of $S/k_B = 1.17$ to 1.83 from the backbone, and considerably less from the side chains, corresponding to a loss of about 3–7-fold in number of conformations from the backbone, and only about 1.4-fold from the side chains. Since the conformational entropy of the unfolded state is large, these reduced estimates for S_{fold}

implies that the residual entropy of the folded state is large, in agreement with previous studies.^{8,9,11} It is within this large residual entropy reservoir that the changes described by Equation 9 occur. Taking, for example, a group whose order parameter changes from say 0.4 to 0.8, a range frequently found for both methyl axis and NH groups, equation 9b gives a decrease in S/k_B of -0.25 and -1.0 respectively, corresponding to decreases in the number of conformations of about 1.3-fold and 3-fold respectively.

Acknowledgments

Supported by NIH grants GM102447, T32 GM008275 and the National Resource for Biomedical Supercomputing and the Pittsburgh Supercomputing Center through NIH award RC2GM093307 to CMU through the NRBSC.

References

1. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J.* 1997; 72:1047–1069. [PubMed: 9138555]
2. Luo H, Sharp KA. On the calculation of absolute binding free energies. *Proc Natl Acad Sci (USA).* 2002; 99:10399–10404. [PubMed: 12149474]
3. Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. *Nature.* 2007; 448:325–329. [PubMed: 17637663]
4. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat Struct Biol.* 2000; 7:72–77. [PubMed: 10625431]
5. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol.* 2010; 6:352–358. [PubMed: 20383153]
6. Tzeng HS, Kalodimos CG. Protein activity regulation by conformational entropy. *Nature.* 2012; 488:236–240. [PubMed: 22801505]
7. Karplus M, Ichiye T, Pettitt BM. Configurational entropy of native proteins. *Biophys J.* 1987; 52:1083–1085. [PubMed: 3427197]
8. Li Z, Raychaudhuri S, Wand AJ. Insights into the local residual entropy of proteins provided by NMR relaxation. *Prot Sci.* 1996; 5:2647–2650.
9. Trbovic N, Cho J, Abel R, Friesner R, Rance M, Palmer A. Protein side-chain dynamics and residual conformational entropy. *J Am Chem Soc.* 2009; 131:615–622. [PubMed: 19105660]
10. Lipari GL, Szabo A. Model-free approach to the interpretation of NMR relaxation in macromolecules. I. Theory and range of validity. *J Am Chem Soc.* 1982; 104:4546–4559.
11. Wand AJ. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Cur Opin Struct Biol.* 2013; 23:75–81.
12. Akke M, Bruschweiler R, Palmer A. NMR order parameters and free energy. *J Am Chem Soc.* 1993; 115:9832–9833.
13. Yang D, Kay L. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J Mol Biol.* 1996; 263:368–382.
14. Lee AL, Sharp KA, Kranz JK, Song X, Wand AJ. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry.* 2002; 41:13814–13825. [PubMed: 12427045]
15. Prabhu N, Lee A, Wand JW, Sharp KA. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. *Biochemistry.* 2003; 42:562–570. [PubMed: 12525185]
16. Yang D, Mok Y, Forman-Kay J, Farrow N, Kay L. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. *J Mol Biol.* 1997; 272:790–804. [PubMed: 9368658]

17. Vugmeyster L, Trott O, McKnight C, Raleigh D, Palmer A. Temperature-dependent dynamics of the villin headpiece helical subdomain, an unusually small thermostable protein. *J Mol Biol.* 2002; 320:841–854. [PubMed: 12095260]
18. Song X-J, Flynn PF, Sharp KA, Wand AJ. Temperature dependence of fast dynamics in proteins. *Biophys J.* 2007; 92:L43–45. [PubMed: 17218465]
19. Best RB, Clarke J, Karplus M. The origin of protein sidechain order parameter distributions. *J Am Chem Soc.* 2004; 126:7734–7735. [PubMed: 15212494]
20. Prabhu NV, Zhu P-J, Sharp KA. Implementation and testing of stable, fast Implicit solvation in molecular dynamics using the smooth-permittivity finite difference poisson-Boltzmann method. *J Comp Chem.* 2004; 25:2049–2064. [PubMed: 15481091]
21. Ming DM, Bruschweiler R. Prediction of methyl-side chain dynamics in proteins. *J Biomol NMR.* 2004; 29:363–368. [PubMed: 15213434]
22. Li DW, Bruschweiler R. Certification of molecular dynamics trajectories with NMR chemical shifts. *J Phys Chem Lett.* 2010; 1:246–248.
23. Kasinath V, Sharp KA, Wand AJ. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *J Amer Chem Soc.* 2013; 135:15092–15100. [PubMed: 24007504]
24. Fu Y, Kasinath V, Moorman V, Nucci N, Hilser V, Wand AJ. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc.* 2012; 134:8543–8550. [PubMed: 22452540]
25. Li D-W, Showalter S, Bruschweiler R. Entropy localization in proteins. *J Phys Chem B.* 2010; 114:16036–16044. [PubMed: 21077640]
26. Li D, Meng D, Bruschweiler R. Short-range coherence of internal protein dynamics revealed by high-precision in silico study. *JACS.* 2009; 131:14610–14611.
27. Das A, Chakrabarti J, Ghosh M. Conformational contribution to thermodynamics of binding in protein-peptide complexes through microscopic simulation. *Biophys J.* 2013; 104:1274–1284. [PubMed: 23528087]
28. Fenley A, Killian B, Hnizdo V, Fedorowicz A, Sharp D, Gilson M. Correlation as a determinant of configurational entropy in supramolecular and protein systems. *J Phys Chem B.* 2014; 118:6447–6455. [PubMed: 24702693]
29. Prompers J, Bruschweiler R. Thermodynamic interpretation of NM relaxation parameters in proteins in the presence of motional correlations. *J Phys Chem B.* 2000; 104:11416–11424.
30. Killian BJ, Kravitz J, Gilson MK. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J Chem Phys.* 2007; 127:024107. [PubMed: 17640119]
31. King BM, Silver N, Tidor B. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *J Phys Chem B.* 2012; 116:2891–2904. [PubMed: 22229789]
32. Baxa MC, Haddadian EJ, Jha AK, Freed KF, Sosnick TR. Context and force field dependence of the loss of protein backbone entropy upon folding using realistic denatured and native state ensembles. *J Am Chem Soc.* 2012; 134:15929–15936. [PubMed: 22928488]
33. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc Natl Acad Sci (USA).* 2014; 111:15396–15401. [PubMed: 25313044]
34. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comp Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
35. Brooks B, Brooks C, Mackerell A, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner A, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R, Post C, Pu J, Schaefer M, Tidor B, Venable R, Woodcock H, Wu X, Yang W, York D, Karplus M. CHARMM: The biomolecular simulation program. *J Comp Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983; 79:926.
37. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *J Molec Graphics.* 1996; 14:33–38.

38. Ryckaert JP, Ciccotti G, Berendsen HJC. Integration of the Cartesian equation of motions of a system with constraints: molecular dynamics of n-alkanes. *Numerical J Comp Chem.* 1977; 23:327–341.
39. Chatfield D, Szabo A, Brooks BR. Molecular dynamics of S. nuclease: Comparison of simulation with ¹⁵N and ¹³O relaxation data. *J Am Chem Soc.* 1998; 120:5301–5311.
40. Matsuda H. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Phys Rev E.* 2000; 62:3096–3102.
41. Mendenhall, W.; Scheaffer, RL. *Mathematical Statistics with Applications.* Duxbury Press; North Scituate, Ma: 1973.
42. Jarymowycz VA, JSM. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev.* 2006; 106:1624–1671. [PubMed: 16683748]
43. Edholm O, Berendsen H. Entropy estimation from simulations of non-diffusive systems. *Mol Phys.* 1984; 51:1011–1028.
44. Baron R, Hunenberger PH, McCammon JA. Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: Correction terms and convergence properties. *J Chem Theory Comput.* 2009; 5:3150–3160. [PubMed: 20011626]

Figure 1.

Relationship between average methyl axis and average amide squared generalized order parameters calculated from molecular dynamics simulations of the seven proteins in Table 1. Solid squares: Average over all methyls and all amides. Solid circles: Average over all methyls and all amides with $O^2_{NH} < 0.8$. Dashed lines correspond to the linear least squares fits.

Figure 2. Frequency distribution of amide N-H bond squared generalized order parameters derived from molecular dynamics simulations of the seven proteins listed in Table 1.

Figure 3.

Relationship between the methyl symmetry axis squared generalized order parameter (O^2_{axis}) and the average of the squared generalized order parameters of the two flanking amide groups ($O^2_{NH^*}$) for the set where $O^2_{NH} < 0.8$ (Panel a). Data were obtained from MD simulations of the seven proteins in Table 1. The dashed line $y = x$ is a guide. The frequency distribution for the O^2 -ratio defined as $O^2_{axis}/O^2_{NH^*}$ is shown in Panel b.

Figure 4.

Relationship between backbone entropy and the amide N-H squared generalized order parameter obtained from molecular dynamics simulations of the seven test proteins. Entropy is calculated from the two-dimensional probability distribution of the ϕ and ψ torsion angles flanking each amide group. Individual amide N-H groups are shown as small circles. Also shown are the averages over the set of $O_{NH}^2 < 0.8$ for each protein (squares). In some cases, the estimated precision (error bars) is less than the size of the symbol. The solid line is a linear fit to the average values.

Figure 5.

Backbone and side chain entropy of the seven proteins listed in Table 1 calculated from MD trajectories. a) Backbone conformational entropy without (solid squares) and with (solid circles) correction for correlations. b) Backbone conformational entropy plus side chain conformational entropy without (solid squares) and with (solid circles) correction for correlations. In some cases, the estimated precision (error bars) is less than the size of the symbol. Dashed lines correspond to the least squares fits to a linear dependence on number of residues with zero intercept.

Table 1

Characteristics of the test protein set used for analysis

Protein ^a (PDB ID) ^b	N _{res} ^c	T (°C) ^d	$\langle O^2_{NH} \rangle^e$	$\langle O^2_{NH} \rangle^f$
Ubiquitin (1UBQ)	76	25	0.629	0.833
a ₃ D (2A3D)	73	30	0.489	0.824
CaM-smMLCKp (1CDL)	167	35	0.562	0.832
CaM-nNOSp (2O60)	168	35	0.560	0.797
HEWL (1LZA)	129	35	0.699	0.842
ALBP (1LIB)	131	20	0.619	0.825
Cyt c2 (1C2R)	116	30	0.670	0.831

^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); see Frederick et al.³ for definition of the peptide sequences; HEWL, hen egg white lysozyme

^b PDB code of starting structure

^c Number of residues

^d Temperature of NMR experiments and simulations

^e All residues

^f Only residues with O^2_{NH} less than 0.8