J|A|C|S COMMUNICATIONS

Published on Web 05/07/2009

A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters

Da-Wei Li and Rafael Brüschweiler*

Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received March 27, 2009; E-mail: bruschweiler@magnet.fsu.edu

The configurational entropy of proteins is a key factor in the thermodynamic stability of protein states and its changes as a function of temperature, ligand binding, and other perturbations. NMR spectroscopy provides unique access to the configurational entropy by the conversion of spin-relaxation-derived dynamics information into local entropies.[1] Analytical relationships between S^2 order parameters[2] and entropy have been proposed[1,3] on the basis of various assumptions in order to characterize the thermodynamic properties of both the backbone and side chains.[4] However, it is not a priori clear how the total configurational entropy can be determined from site-specific contributions of individual spin pairs (e.g., $^{15}N-^{1}H$ or $^{13}C-^{1}H$) because of the possibility of over- or undercounting.[5]

A quantitative approach to the translation of S^2 order parameters into the total configurational entropy is feasible on the basis of accurate reference configurational entropies and order parameters determined from protein molecular dynamics (MD) simulations. Recently, we have shown how configurational entropies of proteins can be accurately determined in dihedral angle space from a structural ensemble, such as an MD trajectory.[6] In this approach, for each dihedral angle φ_j, the von Mises kernel estimation of its probability distribution $p_j(\varphi_j)$ is determined as an additive contribution to the total configurational entropy, S:

$$S = -k_B \sum_{j=1}^{N} \int_0^{2\pi} p_j(\varphi_j) \log[p_j(\varphi_j)]\, d\varphi_j \qquad (1)$$

Second-order correlation effects, which turn out to be small, can be assessed by an expansion method [see the Supporting Information (SI)].[7]

Here we present amino acid-specific relationships between the configurational entropies and S^2 values by applying eq 1 to a set of long MD trajectories of three proteins: ubiquitin, calbindin D_{9k}, and bovine pancreatic trypsin inhibitor (BPTI). For each amino acid type, we identify representative spin pairs whose averaged S^2 values most accurately reflect the configurational entropy.

All of the MD simulations were performed using the AMBER 9 package[8] with the AMBER99SB force field,[9] which was shown previously to accurately reproduce the native-state dynamics of ubiquitin and calbindin D_{9k}.[10] SHAKE[11] was employed to constrain all bonds involving hydrogen atoms, and a time step of 2 fs was used. For the ubiquitin and calbindin D_{9k} simulations, the proteins were embedded in a cubic box with SPC/E water models, and long-range electronic interactions were handled using the PME method[12] at a (real space) cutoff of 8 Å. For BPTI, the generalized Born solvation model was employed.[13] The starting coordinates were taken from X-ray crystal structures [Protein Data Bank entries 1UBQ (ubiquitin), 3ICB (calbindin D_{9k}), and 6PTI (BPTI)], and the simulations were run for 600, 200, and 600 ns, respectively, after application of standard minimization and heating protocols.

In the following, we have assumed that the motional behavior observed for each amino acid type in different proteins and environments is statistically representative of the motional behavior of the same amino acid type in different protein states. Generalized S^2 order parameters for all of the N−H, C−H, and C−CH_3 bonds were determined using the iRED method[14] and averaged over subtrajectories with lengths of 100 and 4 ns. The total configurational entropy was calculated for the side chain of each amino acid type for different simulation lengths and correlated with the average order parameter S^2_{NMR} for the same side chain (eq 2):

$$S = k_B M\left[A + Bf(1 - S^2_{NMR})\right] \qquad (2)$$

where M denotes the number of side-chain dihedral angles, A and B are fit parameters, and $f(x) = x$ or $\log(x)$ (to base e) (see Table 1). Similarly, the backbone entropy contribution from each peptide bond was determined from the corresponding (ψ_{i-1}, φ_i) dihedral angle pair distribution ($M = 2$) and related to its N−H order parameter by eq 2.

For each amino acid type, different combinations of side-chain N−H, C−H, and C−CH_3 S^2 values were tested to ensure that their arithmetic average inserted in eq 2 accurately reproduced the reference entropy obtained from eq 1. For most side chains with $M = 1$, 2, or 3 dihedral angles, measurable $1 - S^2$ values for bond vectors at the end of the side chain accurately probe the configurational entropy. Figure 1 shows amino acid-specific correlations between S and $1 - S^2$ determined from the three MD trajectories. These relationships turned out to be single-valued to a good approximation, with the best parametrizations indicated by straight lines. The corresponding fit parameters A and B are given in Table 1, together with the correlation coefficients R and average errors.

Table 1. Amino Acid-Specific Parametrizations of Side-Chain and Backbone Entropies versus S^2_{NMR} According to Equation 2

amino acid[c]	no. of data points	error/M (k_B)	R[d]	M	A[e]	B[e]
VST[a]	30	0.13	0.93	1	2.19	1.32
IL[a]	34	0.09	0.96	2	1.95	1.55
M[a,g]	4	0.02	0.98	3	2.73	0.77
N[a,f]	7	0.06	0.99	2	2.06	2.08
Q[a,f]	11	0.17	0.93	3	2.16	1.60
FHY[a,g]	18	0.10	0.96	2	2.07	1.51
P[a]	10	0.05	0.89	1	1.90	1.15
K[a]	21	0.06	0.93	4	2.20	1.22
R[a]	10	0.07	0.98	5	2.22	1.23
D[b]	11	0.12	0.93	2	3.69	0.44
E[b]	21	0.09	0.99	3	3.66	0.64
backbone[b]	206	0.15	0.88	2	3.42	0.50

[a] Using $f(x) = x$ in eq 2. [b] Using $f(x) = \log(x)$ in eq 2. [c] Internuclear vectors used for the averaging of S^2 in eq 2: V: $C_\beta-C_{\gamma1}$, $C_\beta-C_{\gamma2}$. S: $C_\beta-H_{\beta2}$, $C_\beta-H_{\beta3}$. T: $C_\beta-C_{\gamma2}$. I: $C_{\gamma1}-C_\delta$. L: $C_\gamma-C_{\delta1}$, $C_\gamma-C_{\delta2}$. M: $S_\delta-C_\varepsilon$. N: $N_{\delta2}-H_{\delta21}$, $N_{\delta2}-H_{\delta22}$. Q: $N_{\varepsilon2}-H_{\varepsilon21}$, $N_{\varepsilon2}-H_{\varepsilon22}$. F: $C_{\delta1}-H_{\delta1}$. H: $C_{\delta2}-H_{\delta2}$. Y: $C_{\delta1}-H_{\delta1}$. P: $C_\gamma-H_{\gamma2}$, $C_\gamma-H_{\gamma3}$. K: $C_\beta-H_{\beta2}$, $C_\beta-H_{\beta3}$, $C_\gamma-H_{\gamma2}$, $C_\gamma-H_{\gamma3}$, $C_\delta-H_{\delta2}$, $C_\delta-H_{\delta3}$, $C_\varepsilon-H_{\varepsilon2}$, $C_\varepsilon-H_{\varepsilon3}$. R: $C_\beta-H_{\beta2}$, $C_\beta-H_{\beta3}$, $C_\gamma-H_{\gamma2}$, $C_\gamma-H_{\gamma3}$, $C_\delta-H_{\delta2}$, $C_\delta-H_{\delta3}$, $N_\varepsilon-H_\varepsilon$. D: $C_\beta-H_{\beta2}$, $C_\beta-H_{\beta3}$. E: $C_\gamma-H_{\gamma2}$, $C_\gamma-H_{\gamma3}$. Backbone: $N-H^N$. [d] Pearson's correlation coefficient. [e] Uncertainties in the fit parameters are given in the SI. [f] For Asn and Gln, $f(x) = \log(x)$ in eq 2 gave similarly good results (see the SI). [g] For Met and His, the precision was low, since only a few data points were available.

10.1021/ja902477s CCC: $40.75

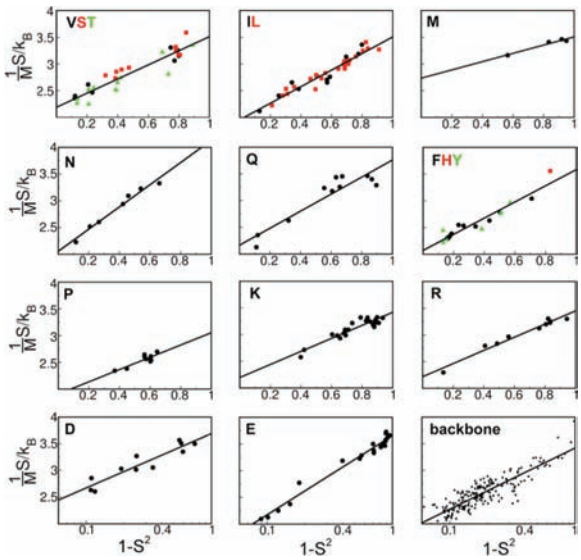

Figure 1. Amino acid-specific correlations (one-letter amino acid abbreviations are given in the upper-left corners) between representative S^2 order parameters and the configurational entropy S (eq 1) determined from MD simulations for ubiquitin, calbindin D_{9k}, and BPTI averaged over 100 ns segments. The straight lines show the relationships of eq 2 using the fit parameters given in Table 1. Data points that belong to different amino acid types are given in different colors. For the three bottom panels (D, E, and backbone), the entropy has a logarithmic dependence on $1 - S^2$.

Amino acids whose side chains exhibit similar S-versus-S^2 relationships can be clustered in separate groups (Figure 1): Val, Ser, Thr (VST), Ile, Leu (IL), and Phe, His, Tyr (FHY). The Asp and Glu amino acid side chains and the protein backbone differ from the other side chains by showing a significantly better correlation when $\log(1 - S^2)$ was used in eq 2 instead of $(1 - S^2)$.

For the determination of relative entropies, such as the entropy difference between two different protein states, the offset A in eq 2 is inconsequential. From this it can be directly seen that the S-versus-S^2 relationship is side-chain-specific. For example, the two methyl groups of Leu ($M = 2$) contribute $\Delta S = -3.10 k_B \Delta S^2_{avg}$, and those in Val ($M = 1$) contribute only $\Delta S = -1.32 k_B \Delta S^2_{avg}$, where ΔS^2_{avg} denotes the change in the C−CH_3 S^2 value averaged over the two methyl groups. This illustrates that the S^2-to-entropy conversion significantly depends on side-chain length and topology.

For the protein backbone, an optimal parametrization was achieved with the $\log(1 - S^2)$ dependence, which yielded for each N−H pair an entropy difference $\Delta S = 2k_B\{0.5 \log[(1 - S^2_{II})/(1 - S^2_{I})]\}$, where S^2_I and S^2_{II} belong to protein states I and II, respectively. This relationship turns out to be identical to the one reported previously for an isolated N−H vector in an axially symmetric potential,[1] with alternative parametrizations yielding very similar results.[3]

For most side chains, eq 2 requires an average S^2 value from a minimal number of selected N−H, C−H, or C−CH_3 bond vectors in order to give an accurate entropy for the whole side chain. The long side chains of Lys and Arg are exceptions:[15] their motional modes can be quite complex, requiring dynamic information from a larger number of bond vectors along the side chain for an accurate conversion into entropic contributions (see Table 1 and the SI).

Because of the additivity of backbone and side-chain entropic contributions, the combination of eqs 1 and 2 provides a quantitative relationship between NMR-derived S^2 order parameters and the total configurational entropy change without issues arising from over- or undercounting. It thus overcomes an important limitation of previous estimates of entropies from NMR S^2 parameters and hence should further the utility of this kind of analysis.

A potential source of error arises from the disparate motional time scale ranges probed by spin-relaxation-derived S^2 (ps to ns) and the entropy. In the presence of dynamic processes that affect S^2 with correlation times shorter than that for molecular tumbling, eq 2 provides only a lower limit for the entropy. To explore the impact of the time-scale disparity between S^2 order parameters and entropies, S^2 values averaged over 4 ns segments were correlated against the entropies averaged over the longer time windows of 100 ns (see the SI). The quality of the correlations decreased only slightly, reflecting the fact that most of the internal correlation times observed in the present MD simulations were shorter than ~10 ns. Alternatively, the time-scale issue can be addressed, in principle, by using order parameters from residual dipolar coupling (RDC) measurements, which are sensitive to motions up to the submillisecond range.[16]

In the past, most NMR studies have focused on backbone N−H and methyl side-chain dynamics,[17] while comparatively little is known about the motional properties of other side-chain types.[18] The S^2-to-entropy dictionary introduced here may motivate the application and further development of NMR experiments geared toward the routine experimental characterization of a diverse range of side chains whose dynamics and thermodynamic properties are expected to considerably broaden our understanding of proteins.

Acknowledgment. This work was supported by the NSF (Grant MCB-0621482).

Supporting Information Available: Discussion of eq 1 and the effect of second-order correlations; uncertainties in the fit parameters A and B; depiction of the dihedral angles and dipolar vectors used in eq 2; a figure and table analogous to Figure 1 and Table 1 but including second-order correlations; and figures and tables analogous to Figure 1 and Table 1 for different lengths of trajectory segments. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Akke, M.; Brüschweiler, R.; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832.
(2) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(3) (a) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369. (b) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647.
(4) (a) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624. (b) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672.
(5) Prompers, J. J.; Brüschweiler, R. *J. Phys. Chem. B.* **2000**, *104*, 11416.
(6) Li, D.-W.; Brüschweiler, R. *Phys. Rev. Lett.* **2009**, *102*, 118108.
(7) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107.
(8) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
(9) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(10) (a) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961. (b) Showalter, S. A.; Johnson, E.; Rance, M.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146.
(11) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Chem.* **1977**, *23*, 327.
(12) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(13) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
(14) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(15) Trbovic, N.; Cho, J. H.; Abel, R.; Friesner, R. A.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2009**, *131*, 615.
(16) (a) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020. (b) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098. (c) Lakomek, N. A.; Walter, K. F. A.; Fares, C.; Lange, O. F.; de Groot, B. L.; Grubmuller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41*, 139.
(17) (a) Palmer, A. G. *Chem. Rev* **2004**, *104*, 3623. (b) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224. (c) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325. (d) Agarwal, V.; Xue, Y.; Reif, B.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2008**, *130*, 16611.
(18) (a) Yang, D. W.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. *J. Mol. Biol.* **1998**, *276*, 939. (b) Namanja, A. T.; Peng, T.; Zintsmaster, J. S.; Elson, A. C.; Shakour, M. G.; Peng, J. W. *Structure* **2007**, *15*, 313. (c) Boyer, J. A.; Lee, A. L. *Biochemistry* **2008**, *47*, 4876. (d) Jin, C. W.; Prompers, J. J.; Brüschweiler, R. *J. Biomol. NMR* **2003**, *26*, 241. (e) Paquin, R.; Ferrage, F.; Mulder, F. A.; Akke, M.; Bodenhausen, G. *J. Am. Chem. Soc.* **2008**, *130*, 15805.

JA902477S