J|A|C|S
COMMUNICATIONS

Published on Web 10/04/2002

Contact Model for the Prediction of NMR N−H Order Parameters in Globular Proteins

Fengli Zhang and Rafael Brüschweiler*

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts 01610

Received July 24, 2002

Heteronuclear NMR relaxation of ^{15}N-labeled proteins represents a rich source of dynamic and thermodynamic information with atomic resolution.[1,2] For the assessment of ns and sub-ns time scale processes longitudinal T_1, transverse T_2, and heteronuclear $\{^1H\}-^{15}N$ nuclear Overhauser (NOE) relaxation parameters are traditionally interpreted in terms of model-free parameters.[3] In this approach the spatial aspects of reorientational motion of $N-H^N$ vectors are described by a generalized order parameter, S^2, and the temporal aspects by an internal and an overall tumbling correlation time. In a recent study, Goodman et al. have statistically analyzed, for a database containing 20 proteins, the influence of various factors on backbone $N-H^N$ S^2 order parameters, such as secondary structure, amino-acid type, and side-chain volume.[4] While many of these properties exhibit non-negligible correlations with respect to S^2, no single factor could be identified that dominates the S^2 behavior.

An analytical relationship is presented here for the estimation of NMR S^2 order parameters of $N-H^N$ vectors of the protein backbone from high-resolution protein structures. It relates S^2 of the $N-H^N$ vector of amino acid i to close contacts experienced by the H^N atom and the carbonyl oxygen of the preceding amino acid $i-1$ with heavy atoms k:

$$S_i^2 = \tanh(0.8\sum_k(\exp(-r^O_{i-1,k}/1\ \text{Å})) + 0.8(\exp(-r^H_{i,k}/1\ \text{Å}))) + b \quad (1)$$

where $r^O_{i-1,k}$ is the distance between the carbonyl oxygen of amino acid $i-1$ to heavy atom k and $r^H_{i,k}$ is the distance between the amide proton H^N and heavy atom k. The parameter b is set to -0.1, which takes into account that order parameters of rigid protein regions typically lie around 0.9. The sum ranges over all heavy atoms k that do not belong to amino acids i and $i-1$. Equation 1 was not obtained in a deductive manner; it rather emerged as an optimized parametrization from a series of empirical attempts.

Equation 1 is illustrated for proteins whose backbone S^2 order parameters have been measured and whose 3D structures have been determined both by X-ray crystallography and NMR. First, S^2 values were predicted for the following monomeric proteins without ligands: interleukin-4, hen-egg white lysozyme, and human ubiquitin. As a representative example, experimental[5] and predicted $1-S^2$ values are plotted for interleukin-4 in Figure 1. The order parameters predicted from the X-ray structure[6] agree well with the experimental values. For the NMR structure-based prediction, the first 10 NMR structures[7] in the PDB were used. For individual NMR structures, the predicted S^2 values can significantly vary, especially for the mobile regions, as is reflected in the standard deviations indicated in Figure 1c. After ensemble-averaging the predicted values compare well with those of the experiment.

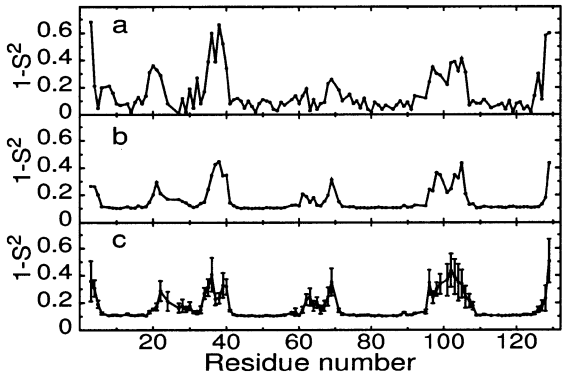

Figure 1. Backbone $N-H^N$ $1-S^2$ values of interleukin-4 plotted as a function of the amino acid number. (a) Experimental values.[5] (b) Values predicted from the X-ray structure[6] using eq 1. (c) Values predicted from the first 10 NMR structures[7] deposited in the PDB using eq 1.

Table 1. Pearson's and Spearman's Correlation Coefficients between Experimental and Predicted $N-H^N$ S^2 Order Parameters Using Eq 1

	PDB entries	X-ray		NMR	
protein[a]	X-ray/NMR	r^b	$r_s{}^c$	r^b	$r_s{}^c$
interleukin-4	1HIK[6]/1CYL[7]	0.81	0.67	0.75	0.62
lysozyme	4LZT[14]/1E8L[15]	0.67	0.71	0.63	0.69
ubiquitin	1UBQ[17]/1D3Z[18]	0.96	0.76	0.95	0.77
calmodulin	4CLN[10]/2BBN[11]	0.49	0.39	0.75	0.69
HIV protease	1MES[12]/1BVE[13]	0.67	0.52	0.59	0.56

[a] Experimental S^2 values are taken from refs 5, 16, 19, 9, 20 for interleukin-4, lysozyme, ubiquitin, calmodulin, and HIV protease, respectively, deposited in the Indiana Protein Dynamics Database (http://pooh.chem.indiana.edu/IDD/).[4] [b] Linear correlation coefficient (Pearson's r).[8] [c] Spearman rank-order correlation coefficient r_s.[8]

Analogous figures for lysozyme and ubiquitin are given in the Supporting Information.

Predicted and experimental S^2 values are quantitatively compared using Pearson's linear and Spearman's rank-order correlation coefficients,[8] which are compiled in Table 1. For the three proteins the two correlation coefficients are similar with values calculated from the X-ray structures only marginally higher than for the corresponding NMR structures.

The method is also applicable to proteins with bound ions as is illustrated for calmodulin. The four Ca^{2+} ions are treated for the S^2 prediction in the same way as all other heavy atoms of the protein. The prediction from the NMR structures[11] is in good agreement with the experimental S^2 values[9] (Figure 2), which includes the correct prediction of high mobility in the central linker region that connects the two domains. In contrast, from the X-ray structure[10] low mobility in the linker region is predicted (Figure 2b) because in the crystal it adopts a helical structure whose peptide planes exhibit numerous close contacts. Thus, the S^2 prediction using eq

* To whom correspondence should be addressed. E-mail: bruschweiler@nmr.clarku.edu.

10.1021/ja027847a CCC: $22.00

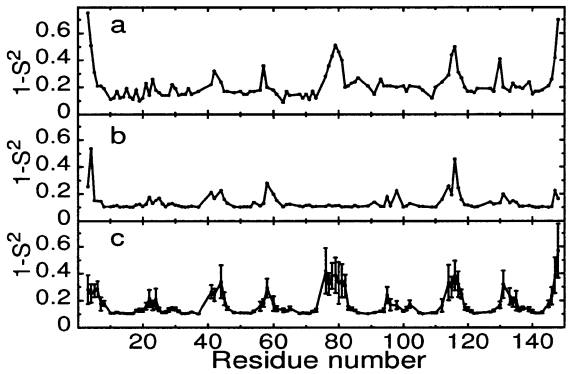

Figure 2. Backbone $N-H^N$ $1 - S^2$ values of calmodulin. (a) Experimental values.[9] (b) Values predicted from the X-ray structure.[10] (c) Values predicted from the first 10 NMR structures[11] deposited in the PDB.

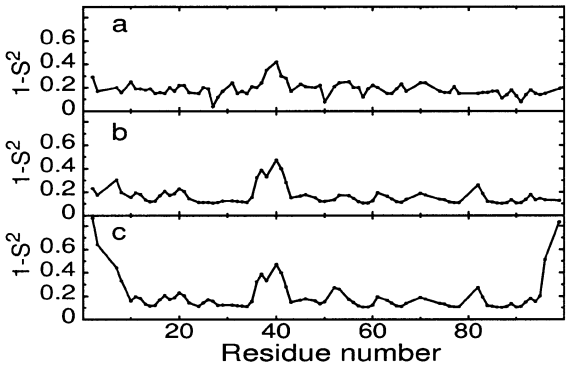

Figure 3. Backbone $N-H^N$ $1 - S^2$ values of HIV-1 protease complexed with dmp323. (a) Experimental values.[20] (b) Values predicted from the dimeric X-ray structure.[12] (c) Values predicted from a monomer taken from the dimer structure.[12]

1 is consistent with this well-known difference between the X-ray and the NMR structures.[9,11]

Application of the method to a protein dimer is illustrated for the HIV-1 protease−dmp323 complex. Figure 3 shows the results for the X-ray structure of the complex.[12] The results for the NMR structures[13] are similar (Table 1). The heavy atoms for the ligand dmp323 are treated in the calculation the same way as the heavy atoms of the polypeptide chain. The only region with increased mobility is the surface loop around amino acid 40, which is reproduced by the prediction. Experimental[20] and predicted $1 - S^2$ values are low at the N- and C-termini of the polypeptide chain. When eq 1 is applied to a single peptide chain only, significantly increased $1 - S^2$ values are predicted at both ends (Figure 3c), which is in disagreement with the experiment and reflected in a drop of the correlation coefficients to $r = 0.39$ and $r_s = 0.43$. Thus, the S^2 prediction method is consistent with the existence of the dimeric state in solution.[13]

In summary, eq 1 allows the easy and rapid estimation of the magnitude of fast time-scale backbone dynamics provided that a high-resolution X-ray structure or a representative ensemble of NMR structures is available. Comparison of experimental and predicted S^2 values allows one to cross-validate local contact properties of the 3D structure and to assess the state of the protein (free vs bound, monomer vs dimer, crystal vs solution).

A possible limitation of eq 1 is that long-range motional effects are not included: a $N-H^N$ vector that belongs to a protein region with high mobility can experience the effects of cumulative motion along the backbone. For the globular proteins discussed here such effects however appear to be small.

S^2 order parameters can be estimated from molecular dynamics (MD) computer simulations.[21] While the agreement with experiment is typically not better than the one obtained using eq 1, MD simulations provide additional insights, such as information about correlated dynamics between different protein parts. For globular proteins, it was found that the largest reorientational amplitude motions generally involve a small number of atoms.[22,23] This explains why low-order parameters can be well-predicted by local contacts entering eq 1.

Related observations were recently made by Jacobs et al.,[24] who qualitatively predicted protein flexibility in terms of local constraints and by Halle,[25] who explained crystallographic *B*-factors of proteins in terms of the local packing density.

The results presented here support the notion that local protein flexibility is to a significant extent directly encoded in the average 3D structure, illuminating the intimate relationship between molecular structure and dynamics. In particular, the contact strength of neighboring atoms to the peptide plane is a powerful indicator for the amount of ns and sub-ns time scale reorientational N−H motion in globular proteins.

Acknowledgment. This work was supported by NSF Grant MCB-0211512.

Supporting Information Available: Two figures with experimental and predicted order parameters for lysozyme and ubiquitin (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Palmer, A. G. *Curr. Opin. Struct. Biol.* **1997**, *7*, 732−737. (b) Kay, L. E. *Nat. Struct. Biol.* **1998**, *5*, 513−517. (c) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740−743.
(2) (a) Stone, M. *Acc. Chem. Res.* **2001**, *34*, 379−388. (b) Spyracopoulos, L.; Sykes, B. D. *Curr. Opin. Struct. Biol.* **2001**, *11*, 555−559. (c) Wand, A. J. *Nat. Struct. Biol.* **2001**, *8*, 926−931. (d) Prompers, J. J.; Brüschweiler, R. *J. Phys. Chem. B* **2000**, *104*, 11416−11424.
(3) (a) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559. (b) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.
(4) Goodman, J. L.; Pagel, M. D.; Stone, M. J. *J. Mol. Biol.* **2000**, *295*, 963−978.
(5) Redfield, C.; Boyd, J.; Smith, L. J.; Smith, R. A.; Dobson, C. M. *Biochemistry* **1992**, *31*, 10431−10437.
(6) Müller, T.; Oehlenschlager, F.; Buehner, M. *J. Mol. Biol.* **1995**, *247*, 360−372.
(7) Müller, T.; Dieckmann, T.; Sebald, W.; Oschkinat, H. *J. Mol. Biol.* **1994**, *237*, 423−436.
(8) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: New York, 1992.
(9) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269−5278.
(10) Taylor, D. A.; Sack, J. S.; Maune, J. F.; Beckingham, K.; Quiocho, F. A. *J. Biol. Chem.* **1991**, *266*, 21375−21380.
(11) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632−638.
(12) Ala, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y.; Hodge, C. N.; Chang, C. H. *Biochemistry* **1997**, *36*, 1573−1580.
(13) Yamazaki, T.; Hinck, A. P.; Wang, Y. X.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Chang, C. H.; Domaille, P. J.; Lam, P. Y. *Protein Sci.* **1996**, *5*, 495−506.
(14) Walsh, M. A.; Schneider, T. R.; Sieker, L. C.; Dauter, Z.; Lamzin, V. S.; Wilson, K. S. *Acta Crystallogr., Sect. D* **1998**, *54*, 522−546.
(15) Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C.; Smith, L. J. *Protein Sci.* **2001**, *10*, 677−688.
(16) Buck, M.; Boyd, J.; Redfield, C.; MacKenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M. *Biochemistry* **1995**, *34*, 4041−4055.
(17) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(18) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(19) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12565−12566.
(20) Tjandra, N.; Wingfield, P.; Stahl, S.; Bax, A. *J. Biomol. NMR* **1996**, *8*, 273−284.
(21) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325−331.
(22) Lienin, S. F.; Brüschweiler, R. *Phys. Rev. Lett.* **2000**, *84*, 5439−5442.
(23) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.
(24) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins* **2001**, *44*, 150−165.
(25) Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1274−1279.

JA027847A