J|A|C|S
COMMUNICATIONS

Published on Web 12/02/2003

Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange

Michele Vendruscolo,[†] Emanuele Paci,[‡] Christopher M. Dobson,[†] and Martin Karplus*,[||]

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge Massachusetts 02138

Received June 5, 2003; E-mail: marci@tammy.harvard.edu

Equilibrium hydrogen-exchange measurements are a powerful tool for investigating the structures, stabilities, and dynamics of native and nonnative states of proteins.[1−4] Previous analyses of hydrogen-exchange data have provided considerable insight into such properties for a variety of proteins, but the results have generally been qualitative in nature,[1,3] although more quantitative thermodynamic analyses have been proposed.[4−7] Here we present a model for the exchange of amide hydrogens with solvent water and determine the structures of species populated during the rare fluctuations of the native state that are required for the exchange. The method is based on an exploration of conformational space using Monte Carlo sampling biased by experimental data.[8] The data involved are the experimental protection (*P*) factors, which represent the ratios of the exchange rates of amide hydrogens in unstructured polypeptides to those of the same residues in the native protein.[2]

We assume that protection from hydrogen exchange is due to hydrogen bonding in secondary structure elements and to the burial of the amide groups in the interior of the protein structure.[2,9−11] Since the detailed chemical mechanism of the hydrogen-exchange process is not fully understood, we introduce these two factors in a phenomenological expression; i.e., the calculated protection factor, $P_i(C)^{\mathrm{sim}}$, of the amide hydrogen of residue i in conformation C is given by

$$\ln P_i^{\mathrm{sim}}(C) = \beta_c N_i^c + \beta_h N_i^h \qquad (1)$$

where N_i^c and N_i^h are, respectively, the number of contacts of residue i with other residues and the number of hydrogen bonds formed by the amide hydrogen of residue i.[12] The experimental free energy difference between a fully closed state (no hydrogen exchange) and a fully open state (where hydrogen exchange takes place at the same rate as that of the equivalent peptide unit in solution) for a particular residue i, in the case of the local unfolding (EX2) mechanism,[2] is $\Delta G_i^{\mathrm{HX}} = RT \ln P_i^{\mathrm{exp}}$. Consequently, the parameter $RT\beta_c$ in eq 1 represents the contribution to the free energy of the formation of an interresidue contact, and the parameter $RT\beta_h$ represents the contribution to the free energy of the formation of a hydrogen bond. The parameters β_c and β_h were estimated by fitting experimental hydrogen-exchange data for a series of native proteins, as described below.

Equilibrium protection factors are a property of a Boltzmann ensemble of conformations, and no single conformation is required, or indeed expected, to satisfy all the observed protection factors simultaneously. The calculated protection factors of eq 1 were therefore taken as averages over M replicas of the molecule; i.e. $\overline{\ln P_i^{\mathrm{sim}}} = \frac{1}{M}\sum_k \ln P_i^{\mathrm{sim}}(C_k)$ where the replicas have conformations C_k ($k = 1, ..., M$). In most of the calculations described in this paper we used $M = 10$, a value found to be sufficient to obtain convergence for the structural properties that were considered here (see below). We defined the pseudo-energy

$$E_{\mathrm{HX}} = \sum_i (\overline{\ln P_i^{\mathrm{sim}}} - \ln P_i^{\mathrm{exp}})^2 \qquad (2)$$

The goal of the conformational sampling with the experimental restraints was to obtain $\langle \overline{\ln P_i^{\mathrm{sim}}} \rangle = \ln P_i^{\mathrm{exp}}$ where the average $\langle \overline{\ln P_i^{\mathrm{sim}}} \rangle$ is taken over the entire ensemble of conformations.

The sampling was carried out with a Monte Carlo procedure, using eq 2 as the energy function. The protein was represented by the main-chain with six atoms per residue, C, C_α, O, N, H, and C_β (or H_α for glycine). Bond lengths and angles were fixed at the canonical experimental values for the native structures, and a hard-sphere nonbonded interaction was included based on the van der Waals radii of the atoms.[13] At each Monte Carlo step a torsion move was attempted for each of the M replicas. The autocorrelation time for the root-mean-square distance (rmsd) from the native state was computed (10^4 Monte Carlo steps for $M = 10$) to generate 10 000 independent conformations for each of the M replicas by repeated annealing cycles down to a zero Monte Carlo temperature.

To determine the values for the two parameters β_c and β_h in eq 1, we considered six proteins. The proteins were horse heart cytochrome *c*,[15] staphylococcal nuclease,[16] ribonuclease H,[17] chymotrypsin inhibitor 2,[18] equine lysozyme,[19] and the basic pancreatic trypsin inhibitor.[20] Using the native structures of these proteins we optimized the correlation of $\langle \overline{\ln P_i^{\mathrm{sim}}} \rangle$ with P_i^{exp} for all i, and obtained $\beta_c = 1$, $\beta_h = 5$ with a correlation coefficient of 0.91; the correlation coefficients for the individual proteins vary between 0.89 and 0.99. Given these values for the β_c and β_h parameters, the loss of a nonbonded contact corresponds to 0.6 kcal/mol and the loss of a hydrogen bond to 3 kcal/mol, in reasonable agreement with other estimates.[2] Using nonbonded contacts alone in eq 1 (i.e., setting $\beta_h = 0$), the largest correlation coefficient (with $\beta_c = 1$) is 0.5; using hydrogen bonding alone in eq 1 (i.e. setting $\beta_c = 0$), the largest correlation coefficient (with $\beta_h = 12$) is 0.4. These results indicate that the nonbonded contacts and hydrogen-bonding terms are both required for a meaningful analysis of the protection factors.

Protection factors measured for the native state of human α-lactalbumin at pH 6.3 and 15 °C in 3 mM Ca^{2+} were used;[21] under these conditions the exchange behavior has been shown to be well described by an EX2 mechanism.[21] Figure 1 shows the comparison between experimental and calculated protection factors in the native-state ensemble of α-lactalbumin. The correlation coefficient is 0.80; it increases to 0.89 if we do not include data for residues Thr86 and Asp87 which have an additional contribution to the protection factors from the Ca^{2+}; the latter was not included as a nonbonded contact in the analysis. For most residues shown with $\ln P_i^{\mathrm{exp}} = 0$ (the experiments were only able to place upper

[†] University of Cambridge.
[‡] University of Zürich.
[||] Harvard University.

10.1021/ja036523z CCC: $25.00

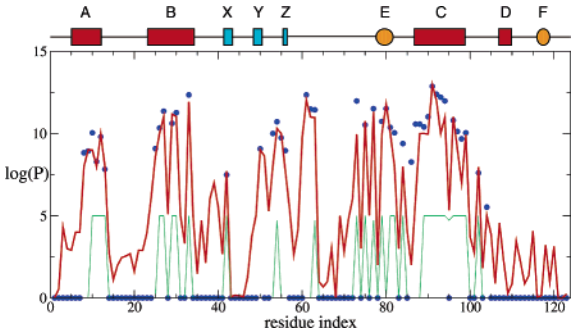

Figure 1. Values of $\ln(P_i^{exp})$ (blue points) compared with values of $\langle \ln P_i^{sim} \rangle$ calculated as described in the text for the native state ensemble (red solid line) of α-lactalbumin. The coefficient of correlation is 0.89. The contribution to $\langle \ln P_i^{sim} \rangle$ due solely to hydrogen bonds is shown by the thin green line. In the calculation of the coefficient of correlation only those amide hydrogens whose protection factors were measured are included. Secondary structure elements in the native state are shown at the top of the figure.

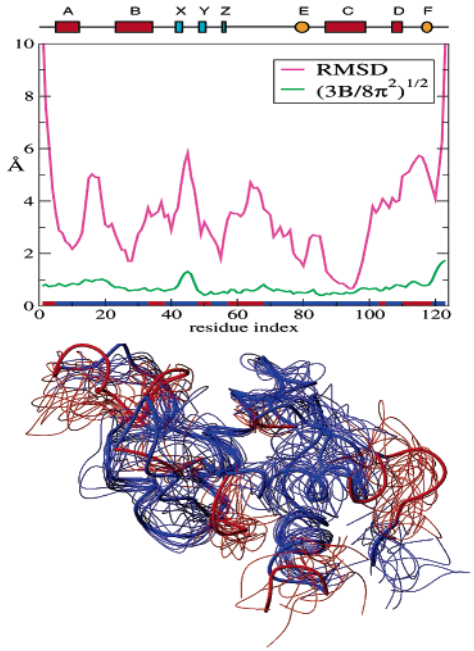

Figure 2. (a) Structural fluctuations (rmsd) per residue in the ensemble calculated in this work (red line) compared with the X-ray *B* factors (green line).[14] Interestingly, the coefficient of correlation between rmsd (averaged over a window of three amino acids) and *B* factors is 0.63, although the magnitudes are very different. (b) Comparison of the crystal structure of α-lactalbumin[14] (bold line) with 15 representative structures in the ensemble (thin lines). The latter represent large but rare fluctuations in the native-state structure that permit hydrogen exchange to take place much more rapidly than in the highly populated species. Residues whose rmsd is larger by a factor of 5 or more, relative to the *B* factor, are shown in red.

limits on the protection factors of the rapidly exchanging hydrogens) we calculate $\ln P_i^{sim} \leq 5$. For two of the 79 residues in this category the predicted behavior is outside these limits; residues Tyr36 and Asn71 have $\langle \ln P_i^{sim} \rangle$ equal to 6.5 and 6.1, respectively.

Due to thermal fluctuations, the native state is an ensemble of very similar structures whose average rmsd is less than 2 Å, as estimated from X-ray *B* factors[14] (see Figure 2a). Under equilibrium conditions many more unfolded conformations are sampled, but only very rarely. HX techniques are unique in providing information concerning such species, as they are required for exchange of amide hydrogens that are otherwise inaccessible to solvent. The ensemble of such rarely populated structures that we have determined in the present work is shown in Figure 2b. For these structures, the average rmsd from the native state is 4 ± 1 Å, although the fluctuations of individual residues are as large as 6 Å (see Figure 2a). Protection factors are large in α-helices A, B, and C, as well in the small 3_{10} helix E, in the β domain; in these regions the average fluctuations are relatively small (<2 Å). A previously puzzling finding from the experimental studies[21] is that there is very little protection in the C-terminal part of the molecule, formed by α-helix D and the 3_{10} helix F, despite the fact that this part of the structure has a well-defined architecture. Our results reveal that the absence of protection is due to fluctuations of the loop between between helices D and F rather than to large fluctuations of the helices themselves; the rmsd from the crystallographic structure of the calculated conformations of this locally cooperative region is 6 Å.

We have shown that it is possible to obtain a detailed model for the structures populated during the rare equilibrium fluctuations of the native state that are sampled by hydrogen exchange. The approach can be applied also to explore the structural characteristics of nonnative states populated under equilibrium conditions. Given the wide applicability of hydrogen-exchange methods, the ability to define and analyze detailed 3D structural ensembles using procedures such as that described here will make possible a significant advance in our understanding of the structure and dynamics of molecules.

Acknowledgment. We thank Jane Clarke and Alan Fersht for helpful discussions. The work was supported by the Royal Society (M.V.), in part by a Program Grant from the Wellcome Trust (C.M.D.), and by the National Institute of Health (M.K.).

References

(1) Clarke, J.; Itzhaki, L. S. *Curr. Opin. Struct. Biol.* **1998**, *8*, 112−118.
(2) Fersht, A. R. *Structure and Mechanism in Protein Science: W. H. Freeman & Co.: New York, 1999.*
(3) Li, R.; Woodward, C. *Protein Sci.* **1999**, *8*, 1571−1590.
(4) Hoang, L.; Bedard, S.; Krishna, M. M.; Lin, Y.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12173−12178.
(5) Hilser, V. J.; Freire, E. *J. Mol. Biol.* **1996**, *262*, 756−772.
(6) Bahar, I.; Wallqvist, A.; Covell, D. G.; Jernigan, R. L. *Biochemistry* **1998**, *37*, 1067−1075.
(7) Viguera, A. R.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5730−5735.
(8) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641−645.
(9) Bentley, G. A.; Delepierre, M.; Dobson, C. M.; Wedin, R. E.; Mason, S. A.; Poulsen, F. M. *J. Mol. Biol.* **1983**, *170*, 243−247.
(10) Miller, D. W.; Dill, K. A. *Protein Sci.* **1995**, *4*, 1860−1873.
(11) Clarke, J.; Itzhaki, L. S.; Fersht, A. R. *Trends Biochem. Sci.* **1997**, *22*, 284−287.
(12) It is of interest in this regard that the number of native contacts (i.e., the packing density) is directly related to the crystallographic temperature (B) factors (see Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1274−1279 2002.
(13) The van der Waals radii used are: 1.15 Å (H), 1.70 Å (C), 1.52 Å (N), 1.50 Å (O), and 1.80 Å (S). Hydrogen bonds were defined to be present if the angle between the NH vector and the OH vector is below 0.7 rad and the OH distance is below 2.4 Å. Contacts between two residues were defined to be present when any pair of their atoms are closer than 8.5 Å.
(14) Acharya, K. R.; Ren, J. S.; Stuart, D. I.; Phillips, D. C.; Fenna, R. E. *J. Mol. Biol.* **1991**, *221*, 571−581.
(15) Milne, J. S.; Mayne, L.; Roder, H.; Wand, A. J.; Englander, S. W. *Protein Sci.* **1998**, *7*, 739−745.
(16) Loh, S. N.; Prehoda, K. E.; Wang, J.; Markley, J. L. *Biochemistry* **1993**, *32*, 11022−11028.
(17) Chamberlain, A. K.; Handel, T. M.; Marqusee, S. *Nat. Struct. Biol.* **1996**, *3*, 782−787.
(18) Itzhaki, L. S.; Neira, J. L.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 89−98.
(19) Morozova, L. A.; Haynie, D. T.; Arico-Muendel, C.; Van Dael, H. V.; Dobson, C. M. *Nat. Struct. Biol.* **1995**, *2*, 871−875.
(20) Kim, K. S.; Fuchs, J. A.; Woodward, C. K. *Biochemistry* **1993**, *32*, 9600−9608.
(21) Schulman, B. A.; Redfield, C.; Peng, Z.-y.; Dobson, C. M.; Kim, P. S. *J. Mol. Biol.* **1995**, *253*, 651−657.

JA036523Z