Protein folding is slaved to solvent motions

H. Frauenfelder*, P. W. Fenimore, G. Chen, and B. H. McMahon

Theory Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Contributed by H. Frauenfelder, August 22, 2006

Proteins, the workhorses of living systems, are constructed from chains of amino acids, which are synthesized in the cell based on the instructions of the genetic code and then folded into working proteins. The time for folding varies from microseconds to hours. What controls the folding rate is hotly debated. We postulate here that folding has the same temperature dependence as the α-fluctuations in the bulk solvent but is much slower. We call this behavior slaving. Slaving has been observed in folded proteins: Large-scale protein motions follow the solvent fluctuations with rate coefficient k_α but can be slower by a large factor. Slowing occurs because large-scale motions proceed in many small steps, each determined by k_α. If conformational motions of folded proteins are slaved, so *a fortiori* must be the motions during folding. The unfolded protein makes a Brownian walk in the conformational space to the folded structure, with each step controlled by k_α. Because the number of conformational substates in the unfolded protein is extremely large, the folding rate coefficient, k_f, is much smaller than k_α. The slaving model implies that the activation enthalpy of folding is dominated by the solvent, whereas the number of steps $n_f = k_\alpha/k_f$ is controlled by the number of accessible substates in the unfolded protein and the solvent. Proteins, however, undergo not only α- but also β-fluctuations. These additional fluctuations are local protein motions that are essentially independent of the bulk solvent fluctuations and may be relevant at late stages of folding.

folding energy landscape | fractional viscosity dependence | internal viscosity | Maxwell relation | protein–solvent interaction

Proteins in cells fold and unfold continuously. Consequently, an understanding of folding rates is key. The distribution and strength of contacts in the native state is one ingredient that influences the rate of folding (1). A second ingredient is the effect of the solvent, because protein motions are intimately linked to the motions of the environment. Our slaving model quantifies the linkage. The model is based on three concepts: Proteins assume a large number of different conformations or substates (2), their organization is described by a hierarchic energy landscape (3), and large-scale protein fluctuations follow the α-relaxation (Debye or dielectric relaxation) in the bulk solvent (4–6). Here we propose that these concepts also are valid for folding and that they lead to the model pictured in Fig. 1. Fig. 1*a* is a cartoon of folding and a 1D cross-section through the high-dimensional energy landscape. Fig. 1*b* is a 2D cross-section. Each valley in this landscape represents a conformational substate of the unfolded protein ensemble (U), the transition state ensemble (TSE), and the native ensemble (N). An unfolded protein starts out in U and makes a random walk in U until it reaches the TSE. Each step in this walk can occur only if the solvent moves and, hence, its rate is proportional to the rate coefficient $k_\alpha(T)$ of the α-relaxation. Once in the TSE, the protein either returns to U or falls into the folding funnel that leads to N (7–11). Fig. 1*a* is, however, misleading because it suggests that there is only one pathway for folding. The 2D cross-section in Fig. 1*b* shows that there are many pathways and that the density of substates can differ in different parts of the landscape. A dense region in U can act as an intermediate state. Proteins with dissimilar structures have been found to have the same folding activation enthalpy (12, 13). The slaving model explains the similar activation enthalpies as being dominated by $k_\alpha(T)$. Although proteins may possess some internal viscosity (14, 15), the slaving model does not require internal protein dissipation to explain the data. Thus, internal viscosity must be a negligibly small effect in existing experiments.

Slaving in the Native Protein

Before treating the effect of the solvent on protein folding, we discuss some aspects of the dynamics of the protein in the native state. Consider the rate coefficient, $k_{exit}(T)$, for the exit of CO or O_2 from myoglobin (Mb) embedded in a glycerol/water solvent (4–6). Because Mb does not have static channels, the exit of ligands must involve conformational fluctuations, which are controlled by the α-fluctuations in the solvent: $k_{exit}(T)$ parallels $k_\alpha(T)$ over almost six orders of magnitude but is slower by nearly five orders of magnitude (5). Slowing occurs because large-scale protein motions are not due to a single large fluctuation but are the result of a large number, n_{exit}, of elementary steps. Each step can only take place if the solvent moves. The rate coefficient $k_{exit}(T)$ is therefore given by

$$k_{exit}(T) = ck_\alpha(T)/n_{exit}(T). \quad [1]$$

The coefficient c is included because it may take somewhat more or less than one solvent fluctuation to induce a step. We take $c = 1$ for simplicity. If $n_{exit}(T)$ is temperature-independent, the enthalpic barrier governing the ligand exit is entirely due to the solvent. If the conformational random walk occurs on a sloping energy landscape or if an enthalpic protein barrier H_P is present, $n_{exit}(T)$ depends on temperature and is given by

$$n_{exit}(T) = n_s e^{\frac{H_P}{RT}}, \quad [2]$$

where R is the gas constant and n_s is the solvent-dependent number of steps if $H_P = 0$. In the language of transition state theory, $\ln n_s = S^\ddagger/R$, where $S^\ddagger$ is the activation entropy. Measurements of $k_{exit}(T)$ and $k_\alpha(T)$ yield the parameters H_P and n_s. Although the concept of slaving was introduced based on experiments (5), the theoretical understanding of slaving is from Lubchenko and Wolynes (4) and is based on the physics of glass-forming liquids. Glass-forming liquids have a dynamic mosaic structure described by the Random First-Order Transition theory. Applied to slaving, the theory not only describes how the molecular motions of the solvent distort the boundaries of the protein but also gives the length and time scales of the solvent-imposed constraints.

Protein folding and large-scale motions of folded proteins are related; they correspond to random walks on the global energy landscape while the protein is embedded in a solvent. It is therefore reasonable to postulate that folding, like ligand escape, is slaved to the solvent and proceeds in a large number of steps,

Author contributions: H.F., P.W.F., G.C., and B.H.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: Mb, myoglobin; N, native ensemble; U, unfolded ensemble; TSE, transition state ensemble.

*To whom correspondence should be addressed: E-mail: frauenfelder@lanl.gov.

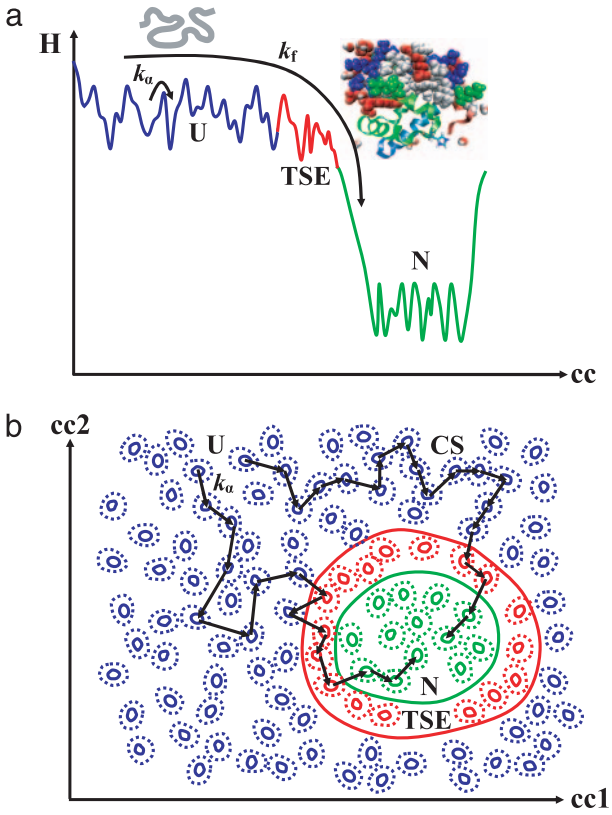

Fig. 1. A schematic description of protein folding. In real space, the unfolded polypeptide (U) folds into the working protein (N). In conformational space, the protein makes a random walk through the high-dimensional energy landscape. (*a*) A 1D cross-section through the energy landscape showing the U (blue), TSE (red), and N (green) conformational basins. The long arrow represents a folding path with an overall rate k_f, whereas the short arrow shows a single step, with a rate k_α, in the conformational diffusion during folding. (*b*) A 2D cross-section through the energy landscape illustrating two different paths for the folding motions of proteins. Starting from a U conformation, proteins make a Brownian walk in the conformational space until they finally fall into the ensemble of N substates.

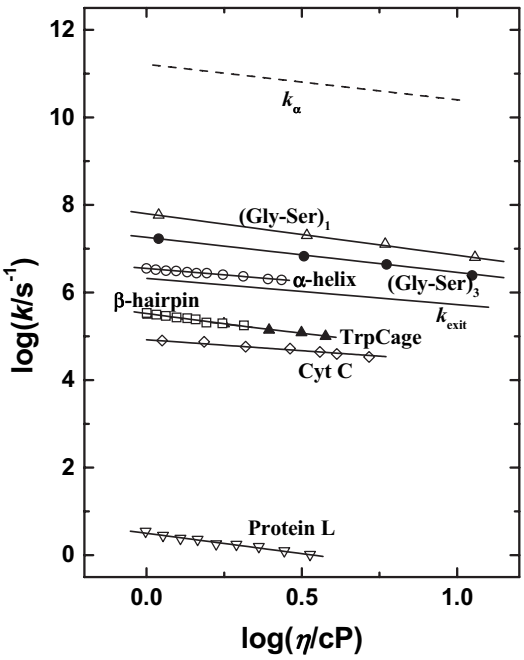

Fig. 2. The folding rates for various polypeptides and proteins versus the solvent viscosity: (Gly-Ser)$_n$ (n = 1 and 3) polypeptide chains (17) (from Eq. **4**, κ = 0.95 and 0.80, respectively), tryptophan cage (21) (κ = 0.84), cytochrome *c* (14, 15) (κ = 0.55), α-helix (κ = 0.53), β-hairpin (18) (κ = 0.93), and protein L (19) (κ = 0.93). The rate coefficients for the bulk α-fluctuations for glycerol–water mixtures (k_α) and for k_{exit} also are plotted for comparison.

n_f. In the slaving model, Eq. **2** also applies to folding. To test the model, $k_f(T)$ and $k_\alpha(T)$ should be measured in the same solvent. We have not found such data. However, many folding rates have been determined as a function of the solvent viscosity $\eta(T)$, which is related to $k_\alpha(T)$ by the Maxwell relation (16):

$$k_\alpha(T) = \frac{G_0(T)}{\eta(T)}, \qquad [3]$$

where G_0 is the shear modulus. At 300 K glycerol–water mixtures have $G_0 \approx 4 \times 10^{11}$ cP·s^{-1}. G_0 varies only weakly with solvent composition and temperature, providing usable values of $k_\alpha(T)$ from η(T). The Maxwell relation can therefore be used to approximately test Eq. **2** for folding when k_α is not known. Fig. 2 presents $k_f(\eta)$ for a few proteins together with $k_\alpha(\eta)$ for glycerol–water solvents. The data exhibit two outstanding features: (*i*) The coefficients k_f of all proteins are much smaller than k_α, consistent with the slaving model in which a large number of steps, n_f, slows protein motions relative to $1/k_\alpha$. For some proteins, k_f follows $1/\eta$, whereas others show a fractional viscosity dependence (e.g., refs. 14, 15, and 17–21):

$$k_f(\eta) \propto \left(\frac{\eta}{\eta_0}\right)^{-\kappa}, \quad \kappa \le 1, \quad \eta_0 = 1 \text{ cP}. \qquad [4]$$

Consider first the case in which $k_f(\eta) \propto 1/\eta$. Assuming G_0 to be constant, Eqs. **2** and **3** substituted into Eq. **4** gives $k_f(\eta) \propto k_\alpha$ and $H_P = 0$, implying that folding is slaved and that the control of folding is largely entropic and given by the solvent. The overdamped Kramers equation (22) leads to the same conclusion but does not determine n_f.

Fractional Viscosity Dependence of Protein Motions

The rate coefficients k_f in many folding experiments do not follow the overdamped Kramers law and deviate from the expected $1/\eta$ dependence. An example is shown in Fig. 3*a*, where k_f^{-1} is plotted as a function of viscosity at 293 K (14, 15). The data up to 5 cP can be fit equally well with a linear fit and a power law, Eq. **4** with κ = 0.55. The linear fit implies an internal protein viscosity, and the power-law fit calls for an explanation of the fractional viscosity dependence. Which model is correct? Flash photolysis experiments in Mb break the deadlock because dynamic data as a function of temperature (5) and viscosity (23) are available. The experiments start with CO bound to the heme iron in Mb. A laser flash breaks the Fe–CO bond, CO escapes into the solvent, and the escape rate coefficient k_{exit} is measured as functions of T in different solvents. Fig. 3*b* shows the rate coefficients for folding and for CO exit as function of logη. The isothermal data in Fig. 3*b* demonstrate that $k_{exit}(\eta)$ follows Eq. **4** with κ = 0.55. Also shown in Fig. 3*b* is k_f taken from Fig. 3*a*. The linear (Fig. 3*b*, solid line) and the power-law (Fig. 3*b*, dashed line) fits diverge above 10 cP. Because the power-law fits for k_f and k_{exit} give the same coefficient $\kappa \approx 0.55$ and because folding and ligand escape both involve large-scale protein motions, the data favor the slaving model. We consequently propose that the power-law fit to the isothermal data is the better description of the viscosity data and that the protein's internal viscosity is not relevant.

The data in Fig. 3 raise another question: What causes the fractional viscosity dependence of k_{exit} and k_f? The information

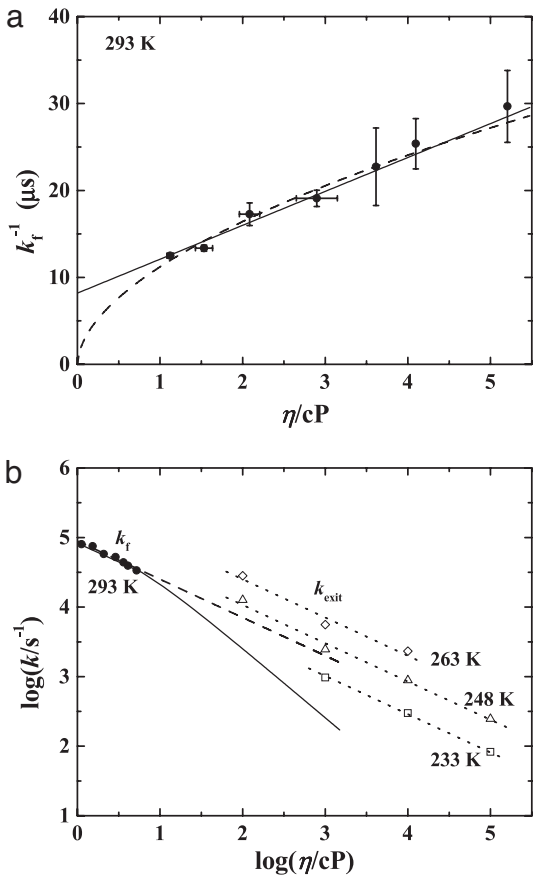

Fig. 3. The viscosity dependence of large-scale protein motions. (*a*) Viscosity dependence of the folding time at 293 K for cytochrome *c* (14, 15). The solid line is a linear fit predicting an internal friction within the protein (14, 15), and the dashed line is a fit to Eq. **4** for a power-law viscosity dependence of folding, with $\kappa = 0.55$. Error bars are the standard deviation. (*b*) Plot of the isothermal rate coefficients versus viscosity. The dotted lines are the rates for the exit of CO from Mb (k_{exit}) at the indicated temperatures; the solid and dashed lines are the folding rates (k_f) corresponding to the linear (solid line) and power-law (dashed line) fits in *a*, respectively.

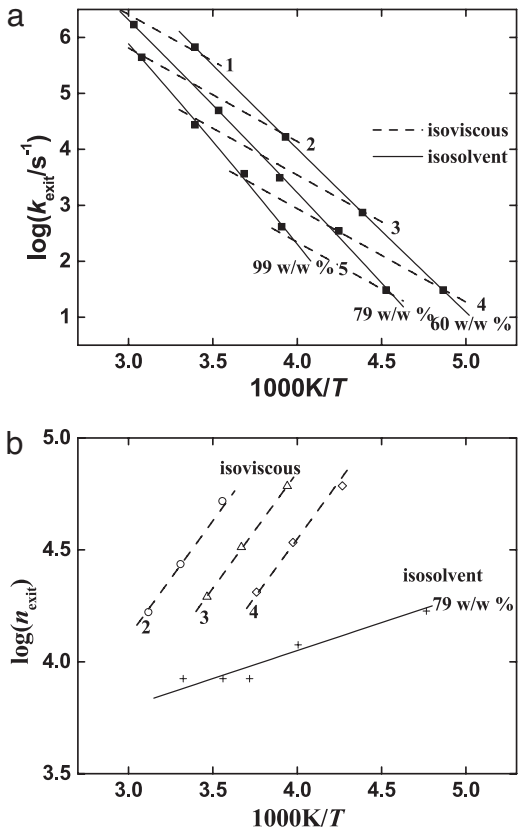

Fig. 4. The exit of CO from Mb (23). (*a*) Arrhenius plot of the CO exit rate. The symbols are the experimental data (23), the solid lines connect the data points measured in different glycerol–water mixtures [99% (wt/wt), 79% (wt/wt), and 60% (wt/wt)], and the dashed lines are the Arrhenius fits to the isoviscous data for five viscosities [log(η/cP) = 1, 2, 3, 4, and 5]. (*b*) Arrhenius plot of n_{exit}. The solid line shows the data for the 79% (wt/wt) mixture, and the dashed lines show the data for three of the viscosities.

from protein dynamics studies provides insight. There are three main ways to look at k_{exit} and k_f, isosolvent, isoviscous, and isothermal. The isosolvent and isoviscous data are in Fig. 4*a*, and the isothermal data are in Fig. 3*b*. In the isosolvent approach (Fig. 4*a*, solid lines) $k_{exit}(T)$ is measured in a given solvent. Evaluating the data with Eq. **2** gives $\log n_s \approx 3$ and $H_P \approx 5$ kJ/mol. Thus, CO needs $\approx 10^3$ steps and has to overcome a small enthalpic barrier to escape. Fig. 4*b* shows that n_s is temperature-independent if measured in the same solvent. A surprise comes from looking at the isoviscous data in Fig. 4 *a* and *b*: $k_{exit}(T)$ follows an Arrhenius law with $H_{exit} \approx 33$ kJ/mol. This value creates a puzzle. Eq. **3**, with G_0 constant, states that k_α and $1/\eta$ must have the same temperature dependence. The activation enthalpy in isoviscous solvents should therefore be given by H_P, apart from a correction for the solvent-dependence of G_0. The slopes of log n_{exit} versus $1/T$ should therefore be the same for isoviscous and isosolvent data: H_P for the isoviscous data should not be as large as 33 kJ/mol. Fig. 4*b* shows, however, that the isoviscous curves have a much larger slope than the isosolvent one. The large activation enthalpy of k_{exit} and the steep increase of n_{exit} and, therefore, of n_s, with $1/T$ in isoviscous solvents must be due to changing entropy. As pointed out earlier, $\ln n_s$ is essentially the activation entropy. Isosolvent data are strikingly different, with n_s constant, apart from the small effect of H_P.

What causes the different behavior of isosolvent and isoviscous experiments? The answer lies in an experimental requirement.

To measure $k_{exit}(T)$ at fixed viscosity requires a different solvent at each temperature. Different solvents can interact differently with the protein and its hydration shell and can change the dynamic and static properties and, in particular, the stability (19, 20, 24–29). Fig. 4*b* shows that the number of steps required for CO to escape at constant viscosity increases with decreasing temperature. The data also demonstrate that n_{exit} at constant temperature increases with decreasing viscosity. In either case, the increase in n_{exit} is caused by a decrease in the viscogen concentration. Thus, n_{exit} is largest and k_{exit} is smallest in the fully hydrated protein. The conclusion is consistent with experiments by Yedgar and coworkers (26). The temperature dependence of folding from the collapsed state of cytochrome *c* at constant viscosity also supports the arguments given here. Measured from 290 to 303 K, $k_f(T)$ has an apparent activation enthalpy of ≈30 kJ/mol at constant viscosity (15, 21). Most or all of the apparent activation enthalpy is due to the effect of the solvent on the protein. Solvent–protein interaction also is responsible for the power-law dependence in Eq. **4**. Present experiments in the literature have not been thorough enough to determine under what conditions κ might be constant between different proteins or protein processes.

The Slaving Model and Folding

Folding involves large-scale motions and takes place in solvents of different compositions and viscosities. The arguments dis-

cussed for ligand escape therefore also apply to folding, in particular to "two-state" folders (30). The shape of the energy landscape in Fig. 1 is important. In one extreme, the main part of the energy landscape is flat and large, and the protein stays in U for a long time before it finds the TSE and falls into one of the many substates of N. In the other extreme, the folding funnel extends to the entire region U, and the protein folds fast. It is difficult to get information about the shape of the folding landscape with experiments that monitor many proteins simultaneously. Consider as an example a landscape with many substates in U and few in the transition state ensemble. Fluctuations in the length-of-stay in U may then fake a broad TSE. Single-molecule experiments provide unique insight into the energy landscape (31). One impressive example is a study by Haran and collaborators (32): Based on previous work by Eaton and collaborators (33), they studied the cold shock protein CspTm in a solution adjusted so that the folded and unfolded state were equally populated. Protein and solvent were encapsulated in a surface-tethered vesicle so that the same protein could be observed many times. Folding and unfolding were then measured by using the Förster resonance energy transfer technique. The experiment shows that the protein stays approximately the same time ($\approx$1 s) in U and N but that the transitions between U and N take $<10^{-4}$ s. This result provides a peek into the energy landscape: U and N, surprisingly, contain about the same number of substates, but the protein spends little time in the TSE. The observation that U and N are nearly equally populated provides more evidence that folded proteins are not in a unique state, as is often stated, but can assume a very large number of conformational substates. Hagen *et al*. (15) have made the observation that there seems to be a "speed limit" near 10^6 s^{-1}. This also implies some minimum number of steps n_f for protein folding. Such a lower limit for n_f, if it exists, may be a measure of hydration shell entropy.

We have so far concentrated the discussion on the effect of the solvent on the folding process. This effect is dominated by the α-fluctuations. Proteins and solvents, however, also undergo β-fluctuations (6), which involve local motions, but we have so far not found direct evidence for their influence on folding. It is likely, however, that β-fluctuations also are important for many features of folding, for instance, the expulsion of water molecules and reorientation of side chains in the final stages of folding. More experiments are needed to explore the role of the β-fluctuations.

The slaving model solves some old problems but produces new ones. We mention just a few: How do solvents change n_f, the number of steps in the conformational space? How do solvents affect the hydration shell? Why is there a systematic connection between viscosity and n as the solvent is varied? How are folding and protein dynamics affected by crowding in cells (34)? Because protein function involves folding and unfolding, the answers to these questions should lead to a better understanding of how proteins work in the biological environment.

We thank Joel Berendzen, Leslie Chavez, Brian Dyer, Martin Gruebele, Peter Hänggi, Martin Karplus, Thomas Kiefhaber, Albert Migliori, John Portman, Peter Wolynes, and Robert Young for discussions, suggestions, and positive criticism. This work was supported by U.S. Department of Energy Contract W-7405-ENG-36 and the Laboratory Directed Research and Development program at Los Alamos National Laboratory.

1. Munoz V, Eaton WA (1999) *Proc Natl Acad Sci USA* 96:11311–11316.
2. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975) *Biochemistry* 14:5355–5373.
3. Frauenfelder H, Sligar SC, Wolynes PG (1991) *Science* 254:1598–1603.
4. Lubchenko V, Wolynes PG, Frauenfelder H (2005) *J Phys Chem B* 109:7488–7499.
5. Fenimore PW, Frauenfelder H, McMahon BH, Parak FG (2002) *Proc Natl Acad Sci USA* 99:16047–16051.
6. Fenimore PW, Frauenfelder H, McMahon BH, Young RD (2004) *Proc Natl Acad Sci USA* 101:14408–14413.
7. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) *Annu Rev Phys Chem* 48:545–600.
8. Portman JJ, Takada S, Wolynes PG (2001) *J Chem Phys* 114:5069–5081.
9. Wolynes PG (2005) *Philos Trans R Soc London A* 363:453–467.
10. Gruebele M (2005) *C R Soc Biol* 328:701–712.
11. Roder H, Maki K, Cheng H (2006) *Chem Rev* 106:1836–1861.
12. Tan YJ, Oliveberg M, Fersht AR (1996) *J Mol Biol* 264:377–389.
13. Scalley ML, Baker D (1997) *Proc Natl Acad Sci USA* 94:10636–10640.
14. Pabit SA, Roder H, Hagen SJ (2004) *Biochemistry* 43:12532–12538.
15. Hagen SJ, Qiu LL, Pabit SA (2005) *J Phys Condens Matter* 17:S1503–S1514.
16. Frenkel J (1955) *Kinetic Theory of Liquids* (Dover, New York).
17. Bieri O, Wirz J, Hellrung B, Schutkowski M, Drewell M, Kiefhaber T (1999) *Proc Natl Acad Sci USA* 96:9597–9601.
18. Jas GS, Eaton WA, Hofrichter J (2001) *J Phys Chem B* 105:261–272.
19. Plaxco KW, Baker D (1998) *Proc Natl Acad Sci USA* 95:13591–13596.
20. Möglich A, Krieger F, Kiefhaber T (2005) *J Mol Biol* 345:153–162.
21. Qiu LL, Hagen SJ (2004) *Chem Phys* 307:243–249.
22. Hänggi P, Talkner P, Borkovec M (1990) *Rev Mod Phys* 62:251–341.
23. Beece D, Eisenstein L, Frauenfelder H, Good D, Marden MC, Reinisch L, Reynolds AH, Sorensen LB, Yue KT (1980) *Biochemistry* 23:5147–5157.
24. Careri G (1998) *Prog Biophys Mol Biol* 70:223–249.
25. Herberhold H, Royer CA, Winter R (2004) *Biochemistry* 43:3336–3345.
26. Yedgar S, Tetreau C, Gavish B, Lavalette D (1995) *Biophys J* 68:665–670.
27. Shimizu S, Smith DJ (2004) *J Chem Phys* 121:1148–1154.
28. Rösgen J, Pettitt MB, Bolen DW (2005) *Biophys J* 89:2988–2997.
29. Mitra L, Smolin N, Ravindra R, Royer C, Winter R (2006) *Phys Chem Chem Phys* 8:1249–1265.
30. Akmal A, Munoz V (2004) *Proteins* 57:142–152.
31. Haran G (2003) *J Phys Condens Matter* 15:R1291–R1317.
32. Rhoades E, Cohen M, Schuler B, Haran G (2004) *J Am Chem Soc* 126:14686–14687.
33. Schuler B, Lipman E, Eaton WA (2002) *Nature* 419:1233–1235.
34. Ellis RJ, Minton AP (2003) *Nature* 425:27–28.