Coil–globule transition in the denatured state of a small protein

Eilon Sherman and Gilad Haran*

Department of Chemical Physics, Weizmann Institute of Science, P.O.B. 26, Rehovot 76100, Israel

Edited by George H. Lorimer, University of Maryland, College Park, MD, and approved June 13, 2006 (received for review February 18, 2006)

Upon transfer from strongly denaturing to native conditions, proteins undergo a collapse that either precedes folding or occurs simultaneously with it. This collapse is similar to the well known coil–globule transition of polymers. Here we employ single-molecule fluorescence methods to fully characterize the equilibrium coil–globule transition in the denatured state of the IgG-binding domain of protein L. By using FRET measurements on freely diffusing individual molecules, we determine the radius of gyration of the protein, which shows a gradual expansion as the concentration of the denaturant, guanidinium hydrochloride, is increased all the way up to 7 M. This expansion is observed also in fluorescence correlation spectroscopy measurements of the hydrodynamic radius of the protein. We analyze the radius of gyration measurements using the theory of the coil–globule transition of Sanchez [Sanchez, I. C. (1979) Macromolecules 12, 980–988], which balances the excluded volume entropy of the chain with the average interresidue interaction energy. In particular, we calculate the solvation energy of the denatured protein, a property that is not readily accessible in other experiments. The dependence of this energy on denaturant concentration is nonlinear, contrasting with the common linear extrapolation method used to describe denaturation energy. Interestingly, a fit to the binding model of chemical denaturation suggests a single denaturant binding site per protein residue. The size of the denatured protein under native conditions can be extrapolated from the data as well, showing that the fully collapsed state of protein is only ~10% larger than the folded state.

Results and Discussion

Radius of Gyration. Protein L has been extensively studied in the past using a variety of spectroscopic techniques, which indicated that this protein belongs to the group of two-state folders, lacking any intermediate states on its folding pathway (15, 17). A SAXS experiment on this protein did not detect an expansion of its denatured state as the denaturant concentration was increased up to 5 M (18). However, the small number of points and the relatively large errors inherent in SAXS measurements at high GuHCl concentrations (where the contrast diminishes) might have precluded observation of the expansion reported below. We prepared a mutant of protein L with cysteine residues inserted near the N and C termini and labeled it with donor and acceptor dyes for FRET, using a sequential process. FRET histograms were constructed from fluorescence bursts created by individual molecules freely diffusing in and out of a focused laser beam (13, 19, 20). Sample histograms are shown in Fig. 1. The histogram at 0 M GuHCl shows a peak at $E = 0.9$ ($E$ is the energy transfer efficiency), which can be assigned to the native protein. As the concentration of GuHCl is increased, a second peak appears, first as a shoulder on the native-state peak and then as a separate peak, gradually shifting to lower and lower $E$ values. This peak can be assigned to the denatured state. A third peak, appearing at $E = 0$, is a well known artifact, arising either from molecules labeled with the donor only or from molecules in which the acceptor was prematurely photobleached (13, 20) and therefore is not shown. We fitted the native-state distribution by using a beta function, which takes into account the photon noise effect on the FRET efficiency (21). This distribution has three parameters: the overall amplitude, the average FRET efficiency, and the average number of photons per burst. From the 0 M GuHCl we could fix the FRET efficiency at $E = 0.9$. We also fixed the number of photons at $n = 25$, so that only the amplitude was fitted. The denatured-state distribution of each histogram was fitted by using a Gaussian function.

From the relative areas of the denatured and native distribu-

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CG, coil–globule; FCS, fluorescence correlation spectroscopy; GuHCl, guanidinium hydrochloride; smFRET, single-molecule FRET; SAXS, small-angle x-ray scattering.

*To whom correspondence should be addressed. E-mail: gilad.haran@weizmann.ac.il.

© 2006 by The National Academy of Sciences of the USA
tions, we could construct a denaturation curve, which was compared with the bulk denaturation curve based on FRET efficiency. The transition midpoint, $C_F$, obtained from the histograms was $1.67 \pm 0.04$ M, which is very similar to $C_F$ extracted from the bulk FRET denaturation curve, $1.64 \pm 0.02$ M (the two curves are shown in Fig. 5, which is published as supporting information on the PNAS web site). This result confirms the validity of our procedure for separating the two distributions in the histograms. Importantly, a gradual shift of the peak position of denatured-state distribution is seen in Fig. 1, from $E = 0.75$ at 1 M GuHCl to $E = 0.45$ at 7 M GuHCl. This shift to a lower FRET efficiency value indicates an expansion of the chain, driven by increased concentrations of the denaturant. The continuous shift of the denatured-state distribution implies a similarly continuous expansion of the denatured protein. This behavior is different from that associated with the folding transition of the protein, which involves coexistence of two populations, folded and denatured. This observation strongly suggests that the expansion/collapse transition is second-order in nature, as is usually observed for the CG transition of flexible polymers (2). The smFRET experiment thus is nicely capable of deciphering the order of a thermodynamic intramolecular transition in a protein.

The denatured-state FRET efficiency distribution does not directly provide the distribution of the end-to-end distance. Recent experiments (22, 23) suggest that the end-to-end distance dynamics for a denatured protein of the size of protein L are on the microsecond time scale, whereas the smFRET data are averaged over millisecond fluorescence bursts. We therefore took the mean $E$ of the denatured-state distribution as an average over the end-to-end distance distribution, $P(R_{ee})$, of the protein

$$\langle E \rangle = \frac{1}{N} \int P(R_{ee}) 1/(1 + R_{ee}^6/R_0^6)dR_{ee}. \quad [1]$$

Here $R_0$ is the Förster radius for energy transfer, taken as $54 \pm 3$ Å (see details on determination of $R_0$ in Supporting Text, which is published as supporting information on the PNAS web site). Adopting a particular form for $P(R_{ee})$ allows us to calculate the possible $\langle E \rangle$ values, then match the experimental values to a particular set of parameters, from which the average mean-squared end-to-end distance, $\langle R_{ee}^2 \rangle$, and the radius of gyration, $R_g$, can be obtained. Here, we used the $P(R_{ee})$ of a Gaussian chain, and in Fig. 2A we plot the calculated values of $R_g = \sqrt{\langle R_{ee}^2 \rangle}/6$ as a function of the GuHCl concentration. We also used the $P(R_{ee})$ of the worm-like chain model (24) for calculating $R_g$ values from the FRET efficiency distributions (data not shown). The $R_g$ values found using the worm-like chain model compare well with values obtained from the Gaussian chain model. The data in Fig. 2A show a strong expansion of $R_g$ from $\sim 18$ Å at 1 M GuHCl to $\sim 27$ Å at 7 M GuHCl. Although work on other proteins already detected chain expansion (12–14), we are not aware of any other example where the expansion continues all the way to 7 M GuHCl. It is clear from the shape of the curve that the full collapse/expansion of the denatured chain is captured in this experiment, from the lowest value in the globule state in 1–2 M GuHCl to essentially the maximal value in the coil state in 6–7 M GuHCl.

**Hydrodynamic Radius.** To further confirm the expansion of the denatured state of protein L, we performed FCS measurements on single-labeled molecules of the protein to extract the hydrodynamic radius as a function of GuHCl concentration. Several fluorescence correlation functions are shown in Fig. 3A. To obtain the hydrodynamic radius of the denatured protein from the correlation functions, we fitted them to the following two-component equation (25), which takes into account the coexistence of folded and denatured molecules:

$$G(\tau) = \frac{1}{N} \left[ \frac{f_f}{1 + \frac{\tau}{\tau_f}} + \frac{f_d}{1 + \frac{\tau}{\tau_d}} \right] \left[ \frac{1 + \frac{\tau}{\tau_f}}{1 + \frac{\tau}{\tau_d}} \right] \left[ \frac{1 + \frac{\tau}{\tau_f}}{1 + \frac{\tau}{\tau_d}} \right]. \quad [2]$$

$G(\tau)$ is the time-correlation function of the fluorescence. $f_f$ and $f_d$ are the fractions of the folded and denatured species, respectively, taken from the denaturation curve of the protein. For the 0 M GuHCl correlation function, $f_d$ was set to 0, whereas for correlation functions at GuHCl concentrations of $>3$ M, $f_d$ was set to 0. $\tau_f$ and $\tau_d$ are the diffusion times of the folded and unfolded (denatured) species, respectively. The former is fixed by the measurement at 0 M denaturant, with a correction for the viscosity change in GuHCl solutions. $N$ is the total average number of molecules in the sampling volume, and $r_{fr}$ and $z_0$ are the dimensions of the sampling volume in the $x$–$y$ plane and the $z$ direction, respectively. The hydrodynamic radius of the folded or denatured protein, $R_{H}^f$ or $R_{H}^d$, is calculated through the diffusion
This term resists chain contraction. The second term is the average interresidue interaction, which is in effect the balance of internal attraction between amino acid residues and residue-solvent attractive or repulsive energy. Contrary to other models of the CG transition, this model does not truncate the excluded volume interaction at the secondary or ternary level but takes into account all higher orders, allowing for a better description of the compact globule state. Minimizing the free energy, Sanchez (16) derived the following transcendental equation for the expansion factor (order parameter) $\alpha^c = R_g^2 / R_{g,0}^2$, where $R_{g,0}$ is the radius of gyration of a reference state

$$\frac{7}{3n\phi_0} (\alpha^3 - \alpha^5) - \frac{\alpha^3}{\phi_0} - \frac{\alpha^6}{\phi_0} \log \left( 1 - \frac{\phi_0}{\alpha^3} \right) = \frac{1}{2T}. \quad [3]$$

In this equation, $n$ is the number of amino acid residues in the chain, and $\phi_0 = R_{g,N}^2 / R_{g,0}^2$ is the volume fraction of the reference state. $R_{g,N}$ is the radius of gyration of the maximally compact state, for which the volume fraction is 1 by definition. In the current context, this value is the radius of gyration of the native state, 16.2 Å (18). $\theta/T$ is the interresidue interaction energy, in $k_BT$ units, which is responsible for the collapse of the chain in a bad solvent. For an infinitely long polymer, the collapse occurs at the $\theta$ temperature, i.e., when $\theta/T = 1$, but this rule is not necessarily true for a finite-length polymer.

**Solvation Energy of the Denatured State.** In the more common case of temperature-induced CG transition, the relevant variable is $T$. However, in the case of chemical denaturation, $T$ is constant and $\theta$ becomes a function of GuHCl concentration. The interaction energy $\theta(C)/T$ is maximal in water, opposing the excluded-volume entropy and contracting the chain to a globular state. As the concentration of GuHCl is increased, the interaction energy decreases, leading to expansion of the chain. By inverting numerically Eq. 3, we can obtain this energy, but to perform this operation we need to know the radius of gyration of the reference state, $R_g,0$. It is reasonable to associate the reference state with the CG transition point, $C_w$ which we define as the point of fastest change in the order parameter $\alpha^2$. This point can be simply calculated by using the relation $d^2(\alpha^2)/dC^2|_{C_w} = 0$. We first make a guess for $R_{g,0}$ and use it together with Eq. 3 to calculate $\theta(C)/T$. We then recalibrate the transition point and iterate the whole calculation several times until we obtain convergence. It is important to note that this calculation leads to an estimate of the energy that is free of any model for the way GuHCl interacts with proteins. The resulting $\theta(C)/T$ is shown in Fig. 2B. It decreases by >1 kT in going from 0 to 7 M GuHCl. This change is solely because of an increase in solvation energy of the protein residues in GuHCl solutions. The overall solvation energy for the denatured protein at a particular denaturant concentration can be obtained from $n(\theta(0) - \theta(C))/T$. It is important to remember that this property is one that cannot be readily measured in folding experiments. In principle, the difference between the solvation energies of the denatured and folded states can be equated with the change in folding energy between aqueous and denaturant solutions (32), which is obtained from the denaturation curve of a protein. However, the solvation energy of the denatured state itself is not accessible in standard experiments.

**Binding Model.** The interaction energy shown in Fig. 2B, or the denatured-state solvation energy that can be calculated from it, are nonlinear in the denaturant concentration. In the common linear extrapolation method, the unfolding energy, which is the difference between the denatured- and folded-state energies, is modeled as a linear function of the denaturant concentration (32). Our finding stands in contrast to this method, or at least puts stringent constraints on the type of models that might lead...
Unfolded Protein at Native Conditions. It is of great interest to describe the unfolded state of a protein under native conditions. It is presumably the state of the protein just after its synthesis on the ribosome (if no cotranslational folding occurs), and it is also the state from which many kinetic folding experiments start, after fast dilution of a denaturant. A short extrapolation of the measurement of the radius of gyration (Fig. 2A) to 0 M GuHCl allows us to find that under native conditions, \( R_g = 17.5 \) Å. Comparing this number with the radius of gyration of the native states as measured by SAXS, 16.2 Å (18), we find that the unfolded state of protein L under native conditions is highly compact, only \( \leq 10\% \) larger than the folded state. A calculation of \( R_g \) of the folded state from the Protein Data Bank coordinates of the protein (ID code 2PTL without the unstructured N-terminal tail) gives a smaller value of 13 Å. If this number is used, the unfolded globular state of protein L is \( \approx 30\% \) larger than the folded state, still rather compact. The \( R_g \) of the unfolded state of drkN-SH3 under nondenaturating conditions also was estimated to be \( 30\% \) larger than that of the folded state (40).

Conclusion. In this work, we used two different and complementary single-molecule fluorescence methods to show that the denatured state of protein L undergoes an expansion in denaturant solutions, which is akin to the well studied CG transition in polymers. Clearly, the CG transition of protein L is thermodynamically distinct from the folding transitions, with a transition point that lies at a much higher concentration of denaturant than the folding point. It is not possible to infer from the thermodynamic measurement whether the collapse and folding transitions should also be kinetically distinct. In fact, Plaxco et al. (18) provided evidence that in protein L these two transitions are simultaneous, as might be the case in two other proteins, studied by Sosnick and coworkers (41). In any case, it is clear from the present results that the denatured state of protein L continues to change with denaturant concentration well beyond the folding point. This kind of behavior, where an apparent two-state folder shows changes in one of its states, was termed “variable two-state denaturation” by Ferreon and Bolen (42).

We used ideas from polymer theory to obtain the solvation energy of the denatured protein as a function of GuHCl concentration. This quantity is particularly interesting because it cannot be measured in standard folding experiments. Interestingly, the binding model for the effect of chemical denaturants could nicely account for the calculated solvation energy, with a binding constant that matches literature values and, most satisfactorily, a single interaction site for GuHCl per residue, a very sensible result that suggests that the denaturant interacts mostly with the protein backbone. Interaction of the backbone with a large number of GuHCl molecules in a concentrated denaturant solution should lead to a more locally rigid structure than in a low-concentration solution. This assertion is supported by our calculations of the radius of gyration using the worm-like chain model (data not shown). These calculations also generated values for the persistence length of the protein at each denaturant concentration, showing an increase of this length by almost a factor of 3 between aqueous solution and concentrated denaturant solution. Because the persistence length is a measure of local chain rigidity, this finding is another indication for the role of denaturant–protein backbone interaction. Changes in persistence length could perhaps be directly measured by force spectroscopy of denatured proteins or by fitting distance–distribution functions obtained from SAXS experiments (43).

An interesting question is whether the collapse transition of proteins is driven just by average changes in solvent conditions, as is the case in homopolymers, or whether specific interactions within the protein drive and guide this transition. As suggested by Hagen and Eaton (44), as well as Roder and coworkers (45), the observation of exponential kinetics even for the earliest...
folding phases constitutes evidence for specific, barrier-crossing transitions [although it has been shown that this conclusion might not be universal (46)]. Support for a specific collapse also was provided. Rev. Let. the time-dependent SAXS experiment of Kimura et al. (10), 86, 4062–4065 a collapsed state of monoclin whose overall structure is oblate, similar to that of the native state of the protein. Our equilibrium experiment does not directly bear on this point. As noted above, the smFRET results imply that the CG transition is second order. However, even a second-order transition, although not involving a barrier-crossing event, still needs not be completely random and might be driven by formation of specific structure elements or contacts in the protein. These can be included in the future in a more refined modeling of the interaction energy.

Materials and Methods

Protein Samples and Labeling. A plasmid containing a histidine-tagged version of protein L was a kind gift of David Baker (University of Washington, Seattle, WA). Donor and acceptor fluorescent probes (Alexa Fluor 488 C3 maleimide and Alexa Fluor 594 C3 maleimide, respectively; Molecular Probes) were attached to cysteine residues inserted close to the N and C termini of the protein [at positions 1 and 64, numbers following (15)], by using a sequential process (47). Details on labeling procedures are provided in Supporting Text.

smFRET Experiments. smFRET measurements were carried out on a home-made confocal microscope (48). Signal detection was performed in a time-stamping mode, i.e., by registering the time lag between consecutive photons arriving at the detectors. Data analysis was performed by using methods similar to those reported by Seidel and coworkers (49). Fluorescence bursts due to passage of individual molecules through the laser beam were identified, and the FRET efficiencies $E$ for each burst was calculated, after correction for the donor leakage, by $E = 1/(1 + \gamma I_D/I_A)$, where $I_D$ and $I_A$ are the donor and acceptor intensities, respectively, and $\gamma$ is a correction factor ($\gamma = 0.7 \pm 0.07$). Further details on the instrument and data analysis are provided in Supporting Text.

Measured samples of freely diffusing protein contained a protein concentration of 150 PM, guaranteeing a probability of $<0.01$ for the simultaneous passage of two protein molecules in the $\sim 1$ fl sample volume. All smFRET measurements were conducted at room temperature. All samples contained 50 mM phosphate buffer (pH 7.4) with varying concentration of GuHCl.

FCS. FCS measurements were performed on the same confocal microscope used for smFRET, but using a 60×, 1.2-NA water-immersion objective (Olympus). A 50-μm pinhole was introduced at the image plane to reject out-of-focus photons. Measurements were conducted on the MIC single mutant labeled with Alexa Fluor 488, as well as on the double mutant, labeled with a single Alexa Fluor 488 dye. The single and double mutants were measured at concentrations of 10 and 20 nM, respectively. The latter sample contained $\sim 1$ mM Tris-(2-carboxyethyl)phosphine (TCEP) to avoid dimerization of the protein. Cross-correlation signals were obtained by splitting the fluorescence into two detectors using a nonpolarizing beam splitter. Data collection and generation of FCS curves were performed by using a hardware correlator (Flex02-12D; Correlator.com).

We thank Rita August for excellent technical help and Wayne Bolen, Shirley Daube, Amnon Horovitz, George Makhatadze, Simon Moehrie, Joan-Emma Shea, Dave Thirumalai, and Guy Ziv for useful comments and suggestions. This work was supported in part by the Israel Science Foundation (administered by the Israel Academy of Science) and by the U.S.–Israel Binational Science Foundation.