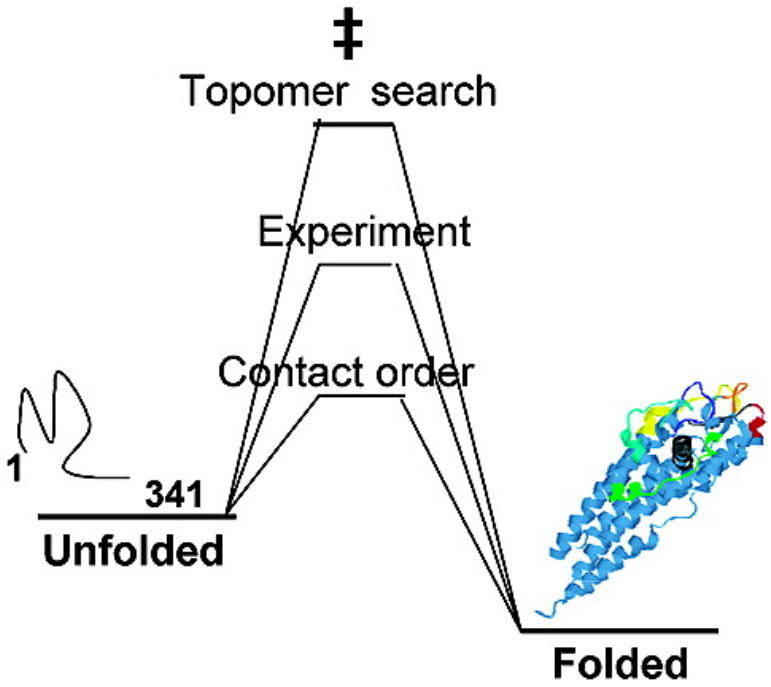
Communication

The Largest Protein Observed To Fold by Two-State Kinetic Mechanism Does Not Obey Contact-Order Correlation

Kathryn Jones, and Pernilla Wittung-Stafshede

J. Am. Chem. Soc., **2003**, 125 (32), 9606-9607• DOI: 10.1021/ja0358807 • Publication Date (Web): 19 July 2003



J|A|C|S
COMMUNICATIONS

Published on Web 07/19/2003

The Largest Protein Observed To Fold by Two-State Kinetic Mechanism Does Not Obey Contact-Order Correlation

Kathryn Jones and Pernilla Wittung-Stafshede*

Molecular and Cellular Biology Graduate Program and Chemistry Department, Tulane University, New Orleans, 70118 Louisiana

Received April 30, 2003; E-mail: pernilla@tulane.edu

Small (<110 residues), single-domain proteins often fold in a two-state kinetic process indicative of a smooth energy landscape lacking discrete traps and heterogeneous roughness. To probe the energy landscape for a very large single-domain protein, we used *Borrelia burgdorferi* VlsE as a model. Flexible loops surround the predominantly α-helical core of this 341-residue, single-domain protein. Equilibrium unfolding of VlsE is two-state and corresponds to a low thermodynamic stability. Remarkably, time-resolved folding and unfolding processes also follow two-state behavior. VlsE is by far the largest protein characterized that folds by a two-state *kinetic* mechanism. In contrast with smaller helical proteins, the folding speed for VlsE is slow (5 ± 2 s^{-1}, pH 7, 20 °C) and does not agree (by 4 orders of magnitude!) with the speed predicted on the basis of its native-state topology.

The *B. burgdorferi* spirochete expresses an abundant surface lipoprotein, VlsE (Variable major protein-Like Sequence, Expressed) that is thought to play a major role in the immune response to Lyme disease.[1] In vivo, an N-terminal cysteine residue links the protein to a lipid, which allows for membrane attachment: without the cysteine, VlsE is highly soluble. In accord with the recent crystal structure,[2] a preliminary analysis of pure recombinant VlsE (without N-terminal cysteine; 341 residues) showed that the protein adopts a single-domain, globular structure dominated by α-helices.[3] Six flexible loops, constituting variable regions (25% of total primary sequence), link to the invariable core including 11 α-helices and 4 small β-strands.[2] In solution of low ionic strength and in the crystal structure, VlsE forms a weak homo-dimer. In accord, thermal unfolding (pH 7) probed by calorimetry involves two transitions: dimer-to-monomer conversion (around 40 °C) followed by protein unfolding (~55 °C).[3] Upon addition of chemical denaturants (20 °C, pH 7), however, dissociation into monomers occurs readily and the unfolding transition as monitored by far-UV circular dichroism takes place in a single transition.[3]

More than 30 single-domain proteins have been reported to fold via two-state kinetic mechanisms with speeds defined by the folded proteins' gross structural properties (as measured by relative contact order).[4] This results in helical proteins, with many short-range interactions and low contact order, folding faster than β-sheet proteins, with more long-range contacts and high contact order, together spanning a 6 orders of magnitude range of folding speeds.[4] Length is usually a determinant of two-state versus non-two-state kinetic folding behaviors: proteins less than 110 residues usually exhibit two-state kinetics whereas larger proteins often populate intermediates.[4,5] However, experimental folding studies on single-domain proteins larger than 150 residues are sparse and it is not clear if there is a size limit for being a two-state folder. We here use VlsE to reveal if a single-domain protein that is dramatically larger in size than those previously studied adopts to the same rules as found for smaller proteins.

Equilibrium unfolding of VlsE was induced by chemical denaturants GuHCl and urea and monitored by tyrosine emission (VlsE has two tyrosines, Tyr18 and Tyr87, but no tryptophan[3]) and far-UV CD (pH 7, 20 °C).[6] With both denaturants, unfolding is 100% reversible. The negative far-UV CD signal, reporting on the large content of helical structure, decreases in magnitude at higher denaturants and the tyrosine emission decreases upon solvent exposure in the unfolded state. For both chemical denaturants, single transitions are observed when the spectroscopic signals are plotted as a function of denaturant (Figure 1). CD and fluorescence-monitored transitions overlap, and there is no protein-concentration dependence in either transition. The transition midpoints occur at 0.4 ± 0.05 M and 1.5 ± 0.3 M GuHCl and urea, respectively. The equilibrium-unfolding data were analyzed in terms of a two-state mechanism[7] resulting in a free-energy of unfolding in water, $\Delta G_U(H_2O)$, of 19 ± 3 kJ/mol (independent of denaturant; see Supporting Information). VlsE is one of a few large proteins exhibiting reversible two-state equilibrium-unfolding mechanisms. *Escherichia coli* maltose binding protein (MBP), 370 residues, and *Alteromonas haloplanctis* α-amylase, 453 residues, are the only other demonstrated examples.[8,9]

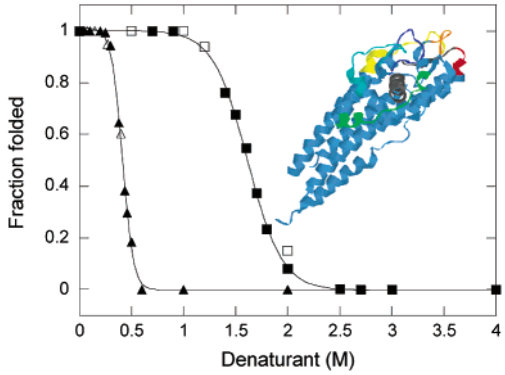


Figure 1. Equilibrium-unfolding transitions (fraction folded versus denaturant) for VlsE using urea (filled squares) and GuHCl (filled triangles) [pH 7, 20 °C]. Open symbols are refolding data; solid lines are two-state fits. Inset: VlsE structure (pdb.1L8W), invariable regions are gray.

The thermodynamic stability of VlsE is very low for a protein of this size. In general, single-domain proteins (100−150 residues long) have stability around 20−50 kJ/mol;[7] the 370-residue MBP has a stability of ~40 kJ/mol.[8] We speculate that since the in vivo environment for VlsE is to be packed among other proteins on the spirochete's surface, and (one of) VlsE's function(s) is to present variable regions to the host's immune system,[10] low stability is advantageous. Low protein stability will allow for rapid turnover and contribute to high antigenic variation and, therefore, to spirochete survival and growth.

Time-resolved folding and unfolding experiments were monitored by changes in tyrosine emission at 320 nm and by far-UV CD at 220 nm.[11] At each denaturant (GuHCl and urea) concentration, both detection probes gave identical kinetic traces and there is no missing CD or emission amplitude within the instrument's dead time (2−4

10.1021/ja0358807 CCC: $25.00

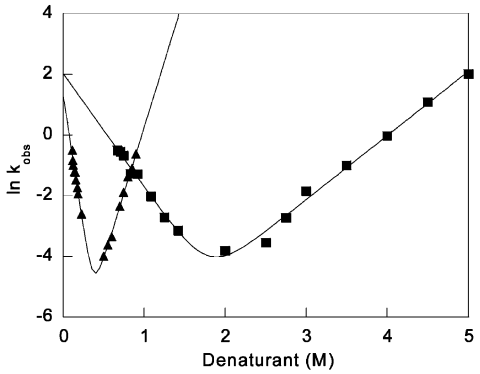


Figure 2. Semilogarithmic plot of ln *k* versus denaturant concentration for VlsE folding and unfolding using urea (squares) and GuHCl (triangles) [pH 7, 20 °C]. Solid curves are two-state fits to the data.

ms). Unfolding and refolding processes are both single-exponential processes and there was no protein-concentration dependence in the first-order rate constants. In addition, the semilogarithmic plot (Chevron plot) of folding and unfolding rate constants versus denaturant concentration exhibits the characteristic V-shape,[6] indicative of two-state kinetic behavior (Figure 2). The folding speed for VlsE in the absence of denaturant (pH 7, 20 °C) is 5 ± 2 s^{-1}. The $\Delta G_U(H_2O)$ value that can be derived from the unfolding and folding rate constants extrapolated to water is 23 ± 4 kJ/mol (independent of denaturant; see Supporting Information), which is in excellent agreement with the equilibrium experiments. Taken together, it is clear that the two-state equilibrium-unfolding mechanism for VlsE holds true also for the kinetic reaction. To date, the largest protein found to fold by two-state kinetics is flavodoxin, an α/β protein of 146 residues.[12] VlsE with its 341 residues is thus by far the largest protein characterized, folding by a simple two-state *kinetic* mechanism.

The quantitative correlation between native-state topology (measured as relative contact order) and folding speed for proteins folding with a two-state mechanism can be used to predict the folding speed of VlsE. The relative contact order (based on the crystal structure) of VlsE is 8.1%,[13] corresponding to a predicted folding rate constant of 150 000 s^{-1} ($1/k_f = 7$ μs).[4] Thus, the experimental folding speed in water for VlsE is 4 orders of magnitude slower than the topology-based prediction, making it the first two-state folder to be significantly off the prediction. The relative contact order for VlsE is similar to that of cytochrome b_{562}, another highly helical protein (but only 106 residues long) that was shown to fold in $\sim$5 μs in water, in excellent accord with the contact-order-based prediction.[14] The effect of polypeptide length has not been directly tested since the proteins included in the original contact-order correlation are all within the range of 50 to 150 residues.[4] Flavodoxin with its 146 residues is so far the largest protein observed to fold in a two-state kinetic process with a rate that matches the contact-order prediction.[12]

The contact-order correlation (an empirical measure) was recently explained in terms of the topomer-search model, a theory that stipulates that the search for those unfolded conformations with a grossly correct topology is the rate-limiting step in folding.[15] Relative contact order and the number of sequence-distant contacts (Q_D, the metrics used in the topomer-search model) are highly correlated.[15] Thus, few systems can be used to distinguish which model predicts folding rates better. The extreme length of the VlsE polypeptide (and the accordingly large number of sequence-distant contacts) makes Q_D high (predicting slow folding), although the relative contact order is very low due to the helical topology (predicting fast folding). In fact, Q_D for VlsE predicts a folding speed of 0.00001 s^{-1} (Kevin Plaxco, personal communication), 5 orders of magnitudes slower than the experimental value. If length-dependence is taken into account, the topomer-search model predicts (counterintuitively) that longer proteins will fold more rapidly.[15] Using the correlation including length dependence,[15] a folding speed of 1700 s^{-1} is now predicted for VlsE. Thus, VlsE is a rare case for which the various models give dramatically different predictions, unfortunately none of which agrees with the experimental data.

We propose that the presence of long flexible loops in folded VlsE slows down the folding speed as compared to the contact-order and length-dependent topomer-search predictions. According to the crystal structure of VlsE, there are six loops 10−20 residues long and additional disorder in the N-terminal part.[2] Although not included as a variable in the current version of the topomer-search model,[15] the probability of achieving a given topomer in the unfolded state should be a function of the chain-length separating the residues that need to form pairs. In accord, recent theoretical work suggests that steric effects in the unfolded ensemble can affect protein folding.[5,16] Experimental evidence for the importance of loop lengths comes from a study of the Arc repressor: the two-state folding speed decreased 4 orders of magnitude when the linker length between two monomers was increased from 13 to 47 residues.[17]

In summary, we here show that a single-domain protein of 341 residues can have a smooth folding-energy landscape lacking heterogeneous roughness and discrete traps. Thus, the common rule of thumb, that proteins larger than 110 residues fold by complex, multistate kinetic mechanisms, must be used with caution. In addition, it is questionable if the barrier height for folding is defined by gross topology for large proteins folding by two-state kinetics.

Acknowledgment. We thank Kevin Plaxco (University of California Santa Barbara) for calculating Q_D. This work was supported by the National Institutes of Health (GM5966301A2).

Supporting Information Available: Examples of kinetic folding and unfolding traces. Table of thermodynamic data for each denaturant. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Zhang, J. R.; Hardham, J. M.; Barbour, A. G.; Norris, S. J. *Cell* **1997**, *89*, 275−285.
(2) Eicken, C.; Sharma, V.; Klabunde, T.; Lawrenz, M. B.; Hardham, J. M.; Norris, S. J.; Sacchettini, J. C. *J. Biol. Chem.* **2002**, *277*, 21691−6.
(3) Jones, K.; Guidry, J.; Wittung-Stafshede, P. *Biochem. Biophys. Res. Commun.* **2001**, *289*, 389−94.
(4) Plaxco, K. W.; Simons, K.; Ruczinski, I.; Baker, D. *Biochemistry* **2000**, *39*, 11177−83.
(5) Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (7), 3555−3556.
(6) Recombinant VlsE was prepared as in ref 3. Equilibrium unfolding (10−100 μM VlsE) was performed at 20 °C, 5 mM phosphate, pH 7. Far-UV CD was measured on an OLIS spectrometer, tyrosine fluorescence on a Varian Eclipse. Guanidine hydrochloride (GuHCl) and urea were of highest purity. Data were analyzed in terms of two-state mechanisms.
(7) Creighton, T. E., Ed. *Protein Folding*; Freeman & Co.: New York, 1992.
(8) Ganesh, C.; Shah, A.; Swaminathan, C.; Surolia, A.; Varadarajan, R. *Biochemistry* **1997**, *36*, 5020−5028.
(9) Feller, G.; d'Amico, D.; Gerday, C. *Biochemistry* **1999**, *38*, 4613−4619.
(10) Lawrenz, M. B.; Hardham, J. M.; Owens, R. T.; Nowakowski, J.; Steere, A. C.; Wormser, G. P.; Norris, S. J. *J. Clin. Microbiol.* **1999**, *37* (12), 3997−4004.
(11) Stopped-flow experiments (20 °C, pH 7) were performed on a PiStar (Applied PhotoPhysics), monitoring fluorescence at 320 nm (ex 280 nm) and CD at 220 nm (2 mm cell, 1:5 mixing, final VlsE concentration 10 and 100 μM). Kinetics were monitored by fluorescence and CD for 17 GuHCl and 16 urea concentrations. In all cases, traces were best fit to single-exponential equations and CD and fluorescence data gave identical results. In Figure 2, the average rate constant at each condition is shown.
(12) Apiyo, D.; Wittung-Stafshede, P. *Protein Sci.* **2002**, *11*, 1129−35.
(13) http://depts.washington.edu/bakerpg/contact_order.
(14) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6587−90.
(15) Markarov, D. E.; Plaxco, K. W. *Protein Sci.* **2003**, *12*, 17−26.
(16) Zagrovc, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 153−164.
(17) Robinson, C. R.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5929−5934.

JA0358807