ϕ-Value Analysis of Apo-Azurin Folding: Comparison between Experiment and Theory†

Chenghang Zong,[§,||,‡] Corey J. Wilson,[⊥,○,‡] Tongye Shen,[§,||] Peter G. Wolynes,*[,§,||] and Pernilla Wittung-Stafshede*[,⊥,○,#]

Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0371, and Department of Biochemistry and Cell Biology, Keck Center for Structural Computational Biology, and Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas 77251

Received January 5, 2006; Revised Manuscript Received March 13, 2006

ABSTRACT: *Pseudomonas aeruginosa* azurin is a 128-residue β-sandwich metalloprotein; in vitro kinetic experiments have shown that it folds in a two-state reaction. Here, we used a variational free energy functional to calculate the characteristics of the transition state ensemble (TSE) for folding of the apo-form of *P. aeruginosa* azurin and investigate how it responds to thermal and mutational changes. The variational method directly yields predicted chevron plots for wild-type and mutant apo-forms of azurin. In parallel, we performed in vitro kinetic-folding experiments on the same set of azurin variants using chemical perturbation. Like the wild-type protein, all apo-variants fold in apparent two-state reactions both in calculations and in stopped-flow mixing experiments. Comparisons of phi (ϕ) values determined from the experimental and theoretical chevron parameters reveal an excellent agreement for most positions, indicating a polarized, highly structured TSE for folding of *P. aeruginosa* apo-azurin. We also demonstrate that careful analysis of side-chain interactions is necessary for appropriate theoretical description of core mutants.

Many aspects of protein folding reactions have been illuminated by energy landscape theory (*1*−*3*). With the minimal frustration principle as a constraint required for biological function, proteins usually are well-described by a simple funneled-shaped energy landscape (*4*). Variations of the energy landscape can change the effective free-energy profile along the folding reaction coordinate. It is clear that a simple funneled energy landscape can yield many different free-energy profiles depending on the topological details of the protein's native structure. Therefore, proteins exploit a variety of kinetic-folding mechanisms. While traditional transition-state theory can be adapted to study the kinetics of protein folding, the large number of degrees of freedom and the often broad distribution of transient states make detailed characterization of the complete free-energy landscape for protein folding necessary. Sampling of the landscape by simulations provides one route to obtain the free-energy profile for folding. Yet, for studies of many variants of a protein, and their folding at different concentrations of denaturant or temperature, performing simulations case by case becomes cumbersome. Furthermore, many point-mutated protein variants only make delicate changes in the folding-free energy landscape, which can be hard to discriminate from the statistical errors occurring during sampling.

Portman, Takada, and Wolynes (PTW)[1] have developed a scheme based on a variational free energy functional for characterization of the folding-free energy landscapes of minimally frustrated proteins (*5*−*7*). Numerical calculations based on the variational principle were found to be both efficient and accurate when compared to most simulation schemes. Calibration of the variational calculation avoids the statistical disadvantages confronted in simulations and, therefore, allows more precise studies of small changes in folding-free energy landscapes caused by denaturants and site mutations. With the variational method, multiple folding pathways for the same protein can be clearly discerned. The variational approach also enables the discrimination of multiple transient states within one folding pathway as well as the detection of shifts of transition states. The latter behavior has been proposed in several experimental protein-folding studies (*8*−*12*). It was recently demonstrated that the PTW variational method predicts folding-kinetic measurables that match in vitro results reasonably well for the small protein U1A (*13*).

† Support for this project was provided by grants from NIH (PGW GM44557 and PWS GM059663) and the Robert A. Welch Foundation (PWS C-1588). C.J.W. is supported by the Houston Area Molecular Biophysics Program (GM08280).

* To whom correspondence should be addressed. P.G.W.: phone, (858) 822-4825; fax, (858) 822-4560; e-mail, pwolynes@ucsd.edu. P.W.-S.: phone, 713-348-4076; fax, 713-348-5154; e-mail, pernilla@rice.edu.

§ Department of Chemistry and Biochemistry, University of California.

|| Center for Theoretical Biological Physics, University of California.

‡ These authors made equal contributions to this work.

⊥ Department of Biochemistry and Cell Biology, Rice University.

○ Keck Center for Structural Computational Biology, Rice University.

Department of Chemistry, Rice University.

[1] Abbreviations: CD, circular dichroism; GuHCl, guanidine hydrochloride; PTW, Portman, Takada, and Wolynes; TSE, transition state ensemble.

10.1021/bi060025w CCC: $33.50
Published on Web 05/02/2006

Experimental studies of protein folding have revealed that small, single-domain proteins often fold by apparent two-state kinetic mechanisms (i.e., involving only the denatured, transition, and native states) (*14*−*16*). Experimentally, phi (ϕ)-value analysis is the most important strategy to explore the nature of the transition state ensemble (TSE) for folding (*17*, *18*). In such experiments, the TSE is probed by measuring the kinetic and thermodynamic effects of hydrophobic-to-alanine mutations in different regions of the protein. The ϕ-values represent the change in stability of the TSE accompanying the mutation of a residue relative to the effect the same mutation has on the stability of the native state, assuming there is no effect on the unfolded state by the mutation. Thus, a ϕ equal to 1 suggests that the corresponding residue makes interactions that contribute equally to the stability of the TSE and the native states (i.e., nativelike interactions in the transition state). In contrast, a ϕ equal to 0 indicates that the corresponding residue forms very few, if any, interactions with other residues in TSE. Fractional ϕ-values are often interpreted as possessing different degrees of structure in the folding nucleus (*19*, *20*). Several small proteins folding by two-state kinetic mechanisms have been subjects of mutagenesis to obtain a picture of the TSE for folding with residue-specific resolution. Such in vitro studies have led to the distinction of two classes of TSE for folding (*21*, *22*), diffuse transition states, in which most side chains have similar relatively low ϕ-values (*22*−*25*), and polarized transition states that display distinct substructures with very high ϕ-values (*26*−*30*).

Pseudomonas aeruginosa azurin is an excellent model system to investigate the applicability of the PTW method to ϕ-value analysis (*5*, *6*). Azurin is a small (128 residues), single-domain metalloprotein with a β-barrel structure composed of eight β-strands, which belongs to the sandwich-like protein family (*31*). In vivo, a redox-active copper is coordinated to the protein allowing for electron-transfer activity (*31*). The copper in azurin can be eliminated, creating apo-azurin, or substituted with zinc without change of the overall structure (*32*, *33*). Equilibrium and kinetic folding processes for apo- and zinc-forms of azurin are two-state reactions (*34*−*39*). The significantly curved chevron plot for zinc-substituted azurin has been attributed to movement of a highly diffuse folding-transition state and was recently used to investigate the growth of the folding-transition state with residue-specific resolution (*12*). In contrast, apo-azurin folds via a fixed TSE (*40*): of the six highly conserved core residues investigated, three positions had ϕ-values near one, whereas the others had near-zero ϕ-values.

In this paper, we probe the free-energy landscape for folding of wild-type and 16 point-mutated forms of *P. aeruginosa* apo-azurin by theory, using the PTW variational method, and by in vitro experiments, using stopped-flow mixing and purified proteins. Theoretical relaxation curves are first constructed from the predicted folding pathways; next, the rate coefficients for folding/unfolding are used to construct theoretical chevron plots for all variants. The resulting theoretical ϕ-values describing apo-azurin's TSE are compared to the corresponding values obtained from the in vitro kinetic measurements. We find excellent agreement between theoretical and experimental ϕ-values; moreover, the estimated native-state stabilities of the various apo-azurin variants also agree between theory and experiment.

THEORETICAL BASIS

The PTW Method. The PTW variational scheme starts with a simple model Hamiltonian of the protein system (*5*−*7*). Both a chain model of the backbone and contact interactions are included in the Hamiltonian which is expressed as $H = H_{\text{chain}} + H_{\text{int}}$. The first term ($H_{\text{chain}}$) models the collapsed stiff chain (i.e., a linear chain of "monomers", each representing a residue in the protein's primary sequence) as follows (*5*):

$$H_{\text{chain}} = \frac{3}{2a^2}\sum_{ij} r_i \Gamma_{ij} r_j + \frac{3}{2a^2} B \sum_i r_i^2$$

where a is a microscopic length scale taken to be the mean square distance between adjacent monomers in the chain, B is the conjugate to the radius of gyration of the chain, and r_i is the position of monomer i in the polymer chain (*5*). The correlations between monomer positions are given by Γ^{-1}. The second term in the Hamiltonian (H_{int}) represents the contact interactions, which are modeled by a pairwise potential $u(r_{ij})$ with an interaction strength coefficient ϵ_{ij}; $H_{\{\text{int}\}} = \sum_{\{ij\}} \epsilon_{ij} u(r_{ij})$. We parametrized the ϵ_{ij} coefficients with Miyazawa−Jernigan contact energies. The interaction potential $u(r_{ij})$ is modeled as a sum of three Gaussian potentials representing short (s)-, intermediate (i)-, and long (l)-range parts

$$u(r) = \sum_{k=(\text{s,i,l})} \gamma_k \exp\left[-\frac{3}{2a^2}\alpha_k r^2\right]$$

where $\alpha_l < \alpha_i < \alpha_s$ are the long-, intermediate-, and short-range widths, respectively. The long-range term is attractive; the intermediate- and short-range terms are repulsive (i.e., $\gamma_l < 0$; $\gamma_i > 0$; $\gamma_s > 0$, respectively) (*5*). The long and intermediate components are chosen to present an energy potential well that has the native distances between residues at the minimum; the short-range term approximates hard-core repulsion between residues (*5*). To construct the free-energy surface, a reference Hamiltonian is first chosen as $H_0 = H_{\text{chain}} + (3/2a^2)\sum_i C_i (r_i - r_i^N)^2$, where N indicates the native state. The constraint parameter, C_i, describes the fluctuations of the ith residue about its native location. With a reasonable choice of values in the reference Hamiltonian, the variational free energy is calculated as

$$F[C] = -k_B T \log Z_o + \langle H - H_0 \rangle_0$$

where, Z_0 is the partition function of the reference Hamiltonian and $\langle H - H_0 \rangle_0$ denotes the average with respect to H_0. When this relation is used, calculations of the folding landscape's energy and entropy are straightforward (*13*).

From Theory to ϕ. The ϕ-values can be defined as $\Delta\Delta G^{\ddagger}/\Delta\Delta G$ (where $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{\text{mut}} - \Delta G^{\ddagger}_{\text{wt}}$ and $\Delta\Delta G = \Delta G_{\text{mut}} - \Delta G_{\text{wt}}$; mut, mutant; wt, wild-type form of protein) if folding is a two-state process with a sharp transition state. In in vitro experiments of protein folding, the directly measured data are presented as relaxation curves corresponding to the rate of equilibration of folding and unfolding events at the particular condition (in terms of temperature, chemical-denaturant concentration and/or pH): $k_{\text{obs}} = k_u + k_f$. Theoretical relaxation curves that correspond to experimental k_{obs} are extracted from the folding pathways predicted by

the PTW method, as we describe in detail below. The free-energy changes corresponding to the folding barrier, $\Delta G^{\ddagger}$, and the protein stability, ΔG, are then estimated via linear extrapolations of the logarithms of the observed relaxation rates to a common condition (such as zero denaturant concentration or a specific temperature). In our theoretical calculations, we altered the balance of folding and unfolding by changing the temperature. The Arrhenius rate coefficients k_u and k_f are expressed as follows, with $\Delta G^{\ddagger}_{u,f} = -(G^{\ddagger} - G_{F,U})$ (A, pre-factor):

$$k_{u,f} = A \exp\left(-\frac{\Delta G^{\ddagger}_{u,f}}{k_B T}\right)$$

EXPERIMENTAL PROCEDURES

Construction of Variants. Construction of the following azurin variants has been reported previously (*36*, *40*, *41*): Val31Ala, Leu33Ala, His46Gly, Trp48Ala, Leu50Ala, Val95Ala, Phe97Ala, Tyr108Ala, Phe110Ala, and His117Gly. An identical scheme was used to prepare the remaining variants used in this work: Ile7Ala, Ile20Ala, Val22Ala, Val60Gly, Ile81Ala, and Leu125Ala azurins. All variants were expressed in *Escherichia coli* BL21pLysS cells, and purification was performed as in ref *36*. The apo-form of each variant was prepared by extensive cyanide dialysis to remove both copper and zinc; completion of the reaction was verified by copper-titration monitoring absorption at 630 nm (*36*). In our in vitro measurements, the balance of folding and unfolding was altered by changing the chemical denaturant guanidine hydrochloride, GuHCl, concentration.

Equilibrium Unfolding. GuHCl-induced equilibrium unfolding was performed in 100 mM Tris-HCl, pH 7.0, at 25 °C using fluorescence (excitation at 285 nm; emission monitored at 308 nm) and far-UV circular dichroism (CD) detection (200−300 nm). Samples were incubated for 2 h before measurements. The equilibrium-unfolding reactions were reversible without any display of protein-concentration dependence (in 5−50 μM protein range). The equilibrium-unfolding curves were analyzed using a two-state model (*42*, *43*) to determine $\Delta G_U(H_2O)$ values. For each azurin variant, far-UV CD- and fluorescence-detected transitions overlapped.

Folding and Unfolding Dynamics. Time-resolved folding and unfolding was probed by fluorescence (excitation at 285 nm; emission monitored at 308 nm) and far-UV CD (at 220 nm) using an Applied Photophysics Pi-Star stopped-flow mixer. Both detection modes gave identical kinetic traces in all cases. Apo-azurin variants were mixed in a 1:10 ratio with appropriate GuHCl/buffer solutions. At each condition, six kinetic traces were averaged before fitting. There were no missing amplitudes (within the 2−3 ms dead time) in either fluorescence or far-UV CD kinetic traces, and no protein-concentration dependence was found (5−50 μM protein range). Both unfolding and refolding reactions were fit to monoexponential decay equations. The observed rate constants, k_{obs}, at different GuHCl concentrations were then fit assuming standard linear dependence of $\ln k_f$ and $\ln k_u$ on GuHCl concentration (*44*). The experimental ϕ-values were calculated as $\phi = \Delta\Delta G^{\ddagger}(H_2O)/\Delta\Delta G_U(H_2O)$, where $\Delta\Delta G_U(H_2O)$ is ΔG_U(mutant, H_2O) − ΔG_U(wild-type, H_2O) and $\Delta\Delta G^{\ddagger}(H_2O)$ corresponds to $RT \ln[k_f$(mutant, H_2O)/k_f(wild-type, H_2O)].

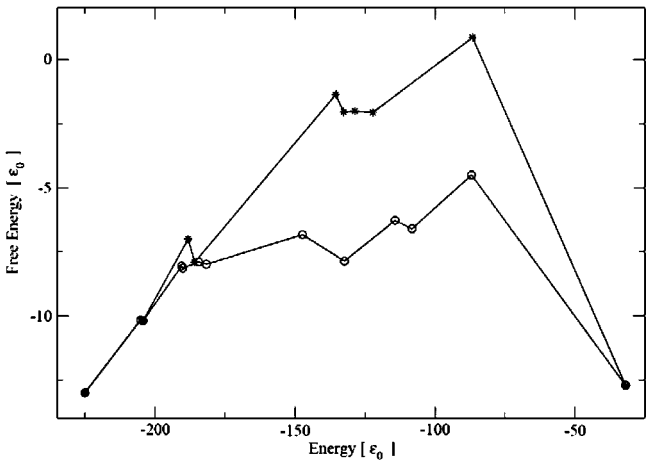

FIGURE 1: Variational free energy versus energy of the two folding pathways [(*) high-energy barrier and (○) low-energy barrier] for apo-azurin at $T = 1.91$ (i.e., at the folding temperature). The point with the highest free energy is denoted as the folding-transition state in each path.

RESULTS AND DISCUSSION

A contact map (input in the PTW variational calculations) for wild-type *P. aeruginosa* apo-azurin was constructed based on the distances between all heavy atoms in the side chains (pdb file 1E65). The contacts are classified into side-chain and backbone contacts according to the distances between side-chain atoms and their angular orientations with respect to each other. As found in the original PTW work (*5*, *6*), the energy unit ϵ_0 is best fit by a factor of 0.6 to scale the Miyazawa−Jernigan contact energies (*45*). Using this scaling, we calculated a relative folding temperature of 1.91 for apo-azurin. To match the experimentally determined thermodynamic stability of *P. aeruginosa* apo-azurin (*40*), the stability (ΔG) at 298 K was set to $\sim 10 k_B T$ in our model. With this, the relative folding temperature of 1.91 corresponds to a value of 320 K. Notably, this is only somewhat lower than the melting temperature reported in in vitro unfolding experiments of *P. aeruginosa* apo-azurin (*46*).

The free-energy profiles predicted from the PTW variational method for folding of wild-type apo-azurin appear two-state-like with one broad barrier; the profiles at the folding temperature (i.e., at $T = 1.91$) are shown in Figure 1. The finer details of the profiles correspond to irregular compensations of entropy loss by free energy gain. We identify two folding paths for apo-azurin at the folding temperature; the path with the lower barrier is the pathway we focus on in this paper (since it is most probable). Although many saddle points are found in both pathways, there are no distinct or highly populated intermediate states found on either pathway. In the present treatment, therefore, we denote the point with the highest free energy as the folding-transition state and the folding rates are calculated as the relaxation times to cross this barrier. A structural interpretation of the folding-transition state, that is, the TSE for folding, can be made by examining the mean-square deviation (MSD) of each residue as predicted by the PTW variational algorithm. Here, the MSD of residue i is defined as how much residue i fluctuates around its mean position in structures belonging to the TSE. The MSD as a function of residue number is shown in Figure 2A for the globular (unfolded), native, and transition states on the two folding routes for wild-type apo-azurin. Similar

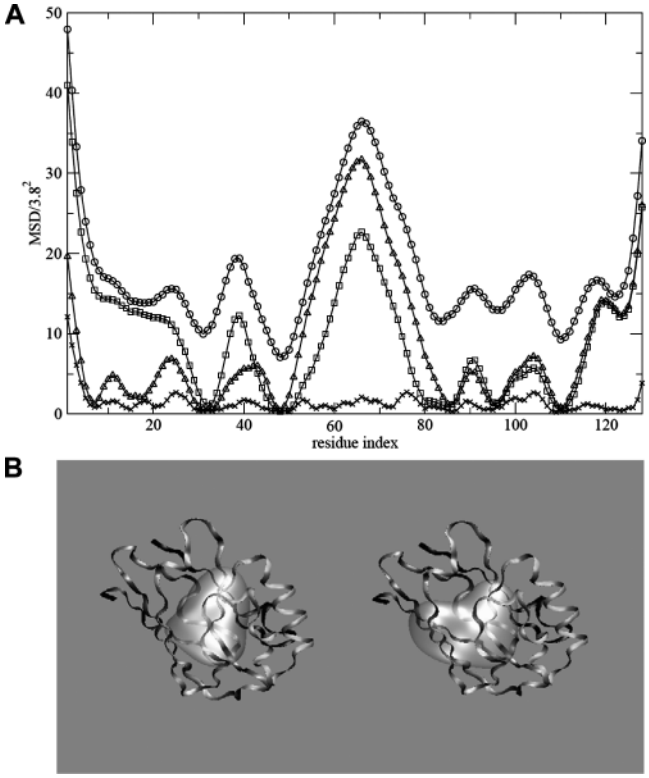

FIGURE 2: (A) MSD as predicted by the variational algorithm for different states on wild-type apo-azurin's folding pathways at $T = 1.91$. The MSD as a function of residue number is shown for the globular/unfolded state (○), the native state (×), as well as the high-barrier TSE (△), and the low-barrier TSE (□) of the two folding pathways identified in Figure 1. (B) The isodensity surface ($\rho = 0.005$) for the two TSE for apo-azurin folding at $T = 1.91$. The left model corresponds to the low-energy barrier TSE (i.e., the focus of this study), and the right model corresponds to the high-energy barrier TSE.

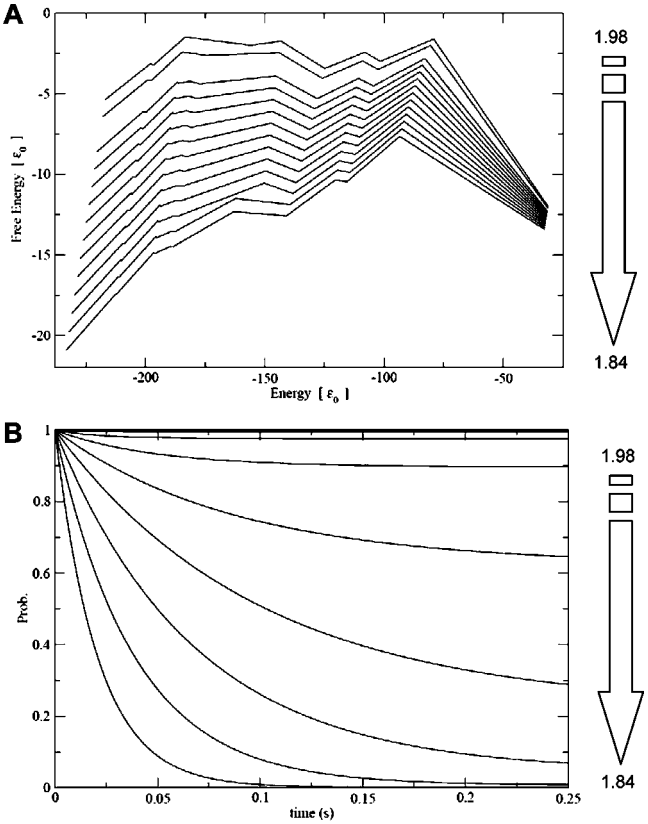

FIGURE 3: (A) Folding paths (variational free energy versus energy) at different temperatures for wild-type apo-azurin. Only the low-energy-barrier pathway is shown. The lines correspond to temperatures from 1.98 to 1.84 with 0.01 increments from top to bottom. (B) Relaxation curves for wild-type apo-azurin of the probability to remain in the globular/unfolded state at various temperatures (with a zero probability value meaning full conversion to the folded state). The curves correspond to relaxation data at temperatures from 1.98 to 1.84 with 0.02 increments from the top to the bottom.

information is visualized in isodensity plots for the two TSEs (see Figure 2B). The construction of this type of plot has been described previously (*13*). In both transition states, a nativelike structure appears in the core around residues 30, 50, 85, 95, and 110 with some additional nativelike structure proximal to the N-terminus in the pathway with the higher barrier.

In Figure 3A, we show the predicted folding routes for wild-type apo-azurin at a set of temperatures. The data shows that, at the higher temperatures (top traces), the unfolded state is energetically favored, whereas at the lower temperatures (bottom traces), the folded state is favored. Following the pathway at each temperature, a simple relaxation curve is calculated using the Arrhenius equation. In Figure 3B, we plot such relaxation curves (reported as probability to remain unfolded versus time) at different temperatures. It is obvious that the lower the temperature, the faster and more complete is the conversion to the folded state. The relaxation curves are fitted to exponential decay equations, which provide rate constants that directly correspond to k_{obs} values, as are normally the parameters determined in in vitro folding measurements. We set the pre-factor of all rate constants at 10^4 to make the theoretical relaxation-time scales similar to those observed in the corresponding in vitro folding experiments on apo-azurin. We note that this pre-factor is approximately 10^2-folds smaller than the speed-limit estimate for protein folding (*47*). This difference may be explained by pre-TSE collapse (*48*) due to azurin's complex native topology which may account for additional friction. The frictional effect on the pre-factor may be even larger than this estimate, since recent experiments (*49*) as well as theoretical work (*6*) have demonstrated that the pre-factor can exceed the speed limit of 1 μs. Alternatively, the current energy model may underestimate the barrier height due to the absence of nonadditive effects.

The folding/unfolding barrier depends linearly on the temperature change to a first-order approximation

$$\Delta G^{\ddagger}_{u,f}(T) = \Delta G^{\ddagger}_{u,f}(T_0) + m_{u,f} \times (T - T_0)$$

Then $k_{u,f}(T)$ is given as follows:

$$k_{u,f}(T) = A \exp\left[\frac{-\Delta G^{\ddagger}_{u,f}(T)}{T}\right] = A \exp\left[\frac{-\Delta G^{\ddagger}_{u,f}(T_0) + m_{u,f} \times T_0}{T} - m_{u,f}\right]$$

Now the logarithms of the observed rate coefficients, $\ln k_{obs}$, can be expressed as

$$\ln k_{obs}(T) = \ln[\exp(a + bT^{-1}) + \exp(c + dT^{-1})] + \ln A$$

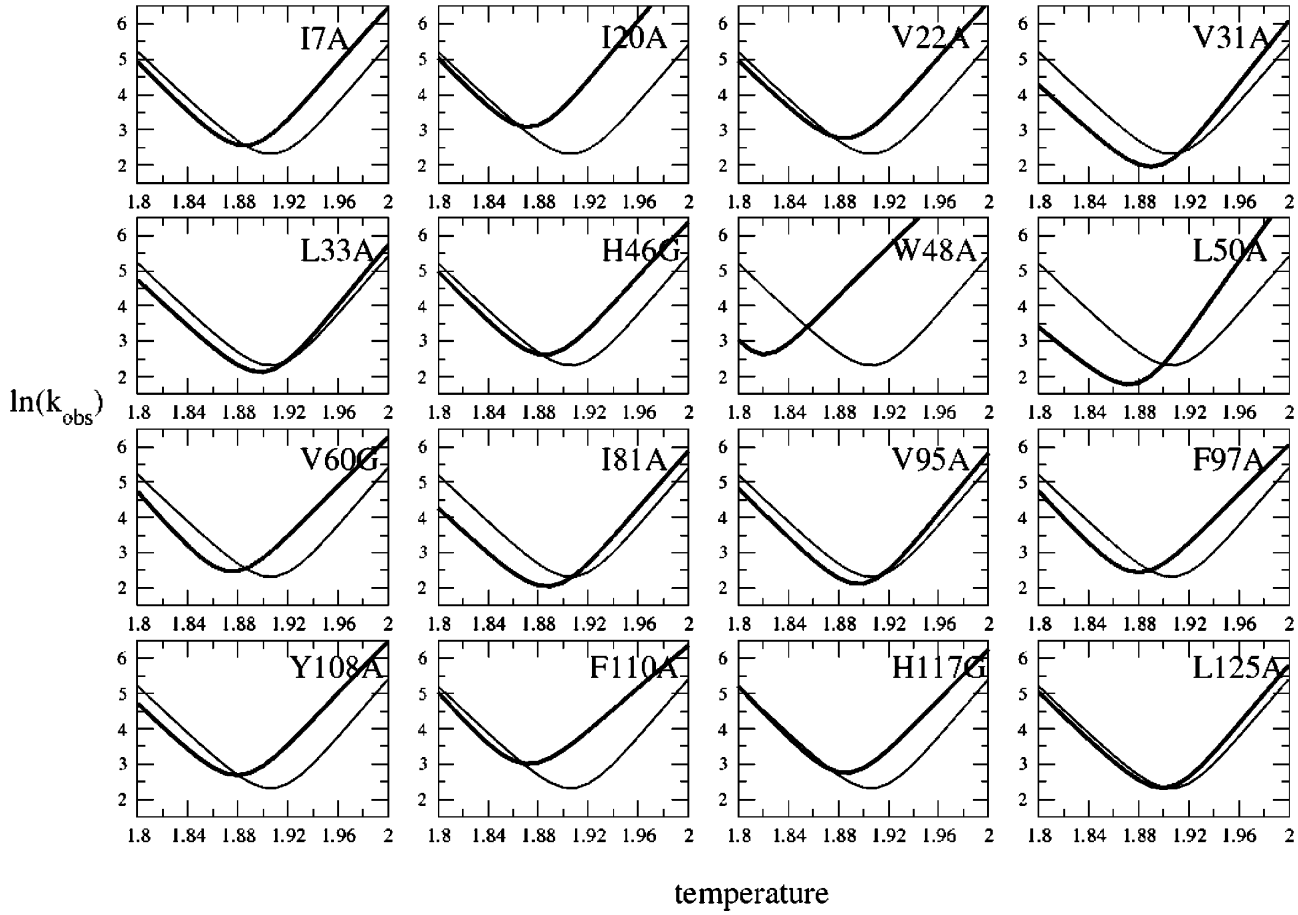

FIGURE 4: Theoretical chevron plots (ln k_{obs} versus temperature) for the 16 apo-azurin variants studied herein. In each graph, the mutant data (thick curve) is overlaid with the wild-type data (thin curve). The arrangement of panels is the same in Figures 4 and 5.

with $a = -m_u$, $b = -\Delta G^{\ddagger}_u(T_0) + m_u T_0$, $c = -m_f$, and $d = -\Delta G^{\ddagger}_f(T_0) + m_f T_0$. The relaxation data for wild-type apo-azurin was used to prepare a theoretical chevron plot (i.e., a plot of the ln $k_{obs}(T)$ versus T data) for wild-type apo-azurin that was then fitted with the above expression to assign values to the a, b, c, and d parameters. In the fitting procedure, parameters a and c will give opposite signs, corresponding to unfolding and folding, respectively.

To derive ϕ-values for specific positions, 16 apo-azurin variants with point-mutations (Ile7Ala, Ile20Ala, Val22Ala, Val31Ala, Leu33Ala, His46Gly, Trp48Ala, Leu50Ala, Val60Gly, Ile81Ala, Val95Ala, Phe97Ala, Tyr108Ala, Phe110Ala, His117Gly, and Leu125Ala) were created based on the contact map of wild-type apo-azurin. In the theoretical description of the point-mutated variants, only the side-chain contacts of the residue in question are eliminated in each case. Backbone contacts, that is, the hydrogen bonds in secondary structures, are not altered in the mutated variants. The PTW variational calculation was performed on each azurin variant to ultimately obtain a set of chevron plots (Figure 4). These were then fitted with the above expression to define the parameters a, b, c, and d for each variant. From the obtained values, the folding barrier ($\Delta G^{\ddagger}$) and the protein stability, $\Delta G = -(\Delta G^{\ddagger}_f - \Delta G^{\ddagger}_u)$, of each system were extracted. Next, the $\Delta G^{\ddagger}$ and ΔG values for wild-type and variants of apo-azurin were combined to yield the ϕ-value for each mutated position. We summarize the theoretically obtained stability and ϕ-value results in Table 1.

To assess the accuracy of the theoretical results as compared to in vitro experiments, we also collected experimental equilibrium- and kinetic-folding data on all the apo-azurin variants. Of the 16 point-mutated variants of azurin analyzed by theory, 10 had been characterized earlier (*36*, *40*), and 6 were investigated in the current work (Ile7Ala, Ile20Ala, Val22Ala, Val60Gly, Ile81Ala, and Leu125Ala azurins). Equilibrium-unfolding curves as a function of the chemical denaturant GuHCl for wild-type and all mutants are single transitions, and unfolding measurements using two modes of detection (i.e., fluorescence and far-UV CD) are coincidental (Supporting Information). This suggests the reactions to be two-state processes; the stability values obtained from two-state fits for all variants in the absence of GuHCl, $\Delta G_u(H_2O)$, are listed in Table 1. In Figure 5, we show the resulting chevron plots for folding and unfolding kinetics of the 16 apo-azurin variants. In accord with two-state kinetics, the constituting arms of a given chevron are linear with neither protein concentration dependence nor missing amplitudes. The ϕ-values calculated from these experiments upon extrapolation to 0 M GuHCl are listed in Table 1. As a control, experimental and theoretical ϕ-values were also calculated at the transition midpoints in the chevron plots. The results using this approach are very similar to those reported in Table 1 (see Supporting Information for details).

The experimentally and theoretically derived ϕ-values are compared in Figure 6A; the correlation coefficient γ between the two sets of data is 0.80. The ϕ-value of Leu125Ala has

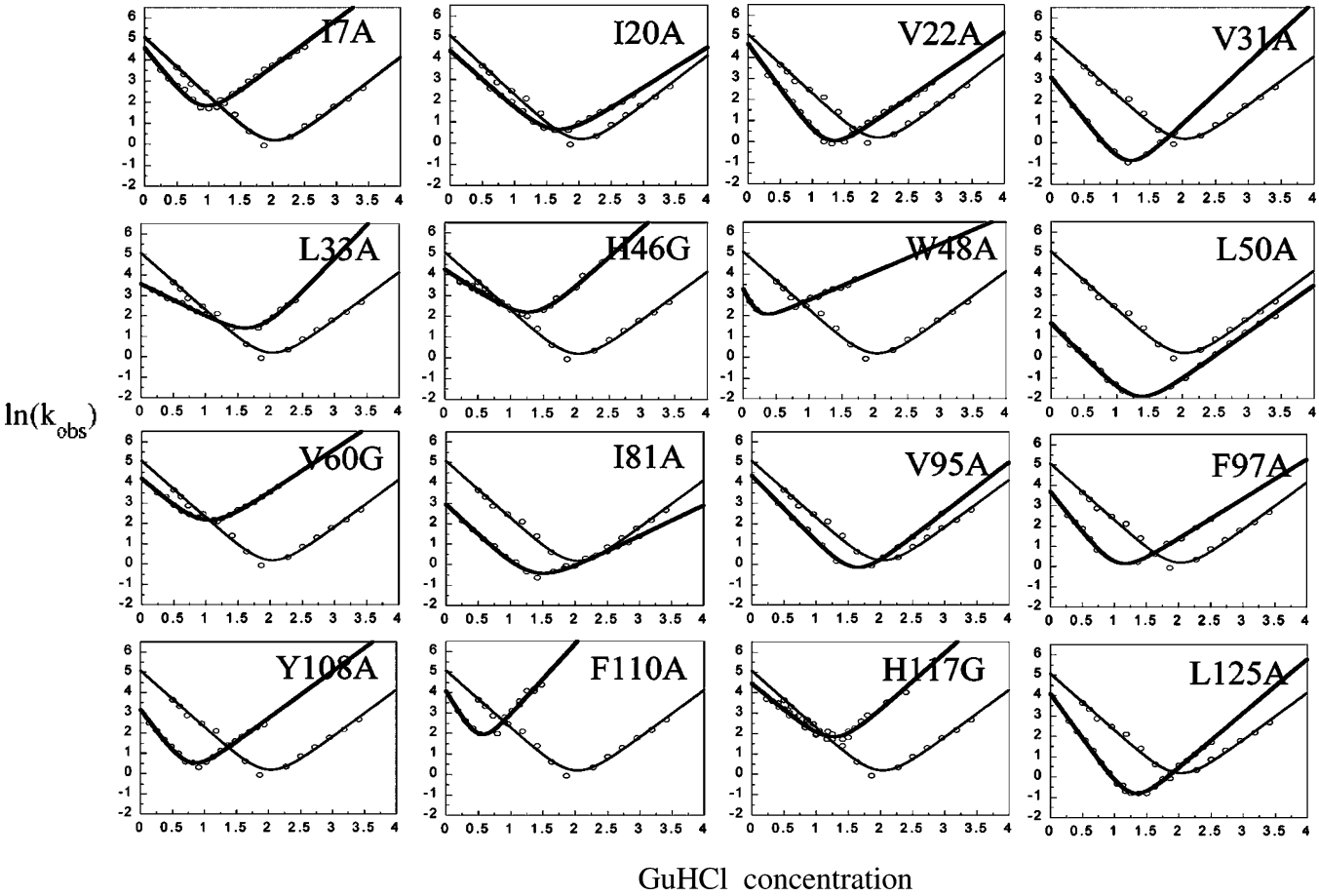

FIGURE 5: Experimental chevron plots (ln k_{obs} versus GuHCl concentration) for the sixteen apo-azurin variants. In each graph, the mutant data (thick curve) is overlaid with the wild-type data (thin curve). The arrangement of panels is the same in Figures 4 and 5.

Table 1: Protein Stability and Kinetic ϕ-Value Data for *P. aeruginosa* Apo-Azurin (Wild-Type and 16 Point-Mutated Variants) As Derived from Theoretical Calculations (the PTW Variational Method) and in Vitro Measurements (pH 7, 25 °C), Respectively[a]

apo-azurin variant	mutated β-strand	eliminated contacts	ΔG (k_BT)	ϕ (theory)	$\Delta G_U(H_2O)$ (kJ/mol)	ϕ^b (experiment)
wild-type	-	-	8.82	-	25.0 ± 0.7	-
Ile7Ala	1	15, 16, 17, 31, 33	7.15	0.18	12.0 ± 0.3	0.10 ± 0.05
Ile20Ala	2	29, 48	6.28	0.07	18.0 ± 0.5	0.27 ± 0.03
Val22Ala	2	29, 99, 125, 127	6.95	0.14	19 ± 0.2	0.18 ± 0.06
Val31Ala[c]	3	48	7.91	1.04	20.0 ± 0.4	0.93 ± 0.14
Leu33Ala[c]	3	84, 87	8.40	1.14	21.0± 0.3	0.91 ± 0.18
His46Gly[d]	4	9, 10, 35, 46	6.96	0.13	14 ± 0.6	0.1 ± 0.02
Trp48Ala[c]	4	20, 31, 84, 95, 110	1.56	0.33	4.0 ± 0.3	0.23 ± 0.03
Leu50Ala[c]	4	81, 97	7.09	0.67	18.0 ± 0.2	1.04 ± 0.11
Val60Gly	helix	111, 113, 118	5.99	0.17	12 ± 0.6	0.18 ± 0.03
Ile81Ala	5	97, 101, 108	7.19	0.58	15 ± 0.3	0.52 ± 0.08
Val95Ala[c]	6	48	8.34	0.81	22.5 ± 0.2	0.63± 0.2[e]
Phe97Ala[c]	6	29, 50	6.25	0.18	15.0 ± 0.7	0.35 ± 0.04
Tyr108Ala[c]	7	81, 102, 103, 125	6.26	0.19	12.5 ± 0.2	0.37 ± 0.03
Phe110Ala[c]	7	15, 17, 18	5.20	0.05	11.0 ± 0.9	0.19 ± 0.02
His117Gly[d]	loop	13, 42, 112	6.76	0.00	15 ± 0.4	0.1 ± 0.03
Leu125Ala	8	50	8.09	0.20	22 ± 0.6	0.81 ± 0.12[e]

[a] The theoretical ϕ-values are calculated as $\phi = \Delta\Delta G^{\ddagger}/\Delta\Delta G$, where $\Delta\Delta G$ is calculated as ΔG(mutant) − ΔG(wild-type) and $\Delta\Delta G^{\ddagger}$ as $\Delta G^{\ddagger}$(mutant) − $\Delta G^{\ddagger}$(wild-type). Similarly, the experimental ϕ-values are calculated as $\phi = \Delta\Delta G^{\ddagger}(H_2O)/\Delta\Delta G_U(H_2O)$, where $\Delta\Delta G_U(H_2O)$ is calculated as ΔG_U(mutant, H_2O) − ΔG_U(wild-type, H_2O) and $\Delta\Delta G^{\ddagger}(H_2O)$ as RT ln[k_f(mutant, H_2O)/k_f(wild-type, H_2O)]. For each mutant, the localization in the folded structure of the residue exchanged is indicated, as well as the side-chain contacts eliminated in its theoretical description. [b] Standard deviations (σ) for the ϕ-values were calculated as in ref 20. [c] Experimental data taken from ref *40*. [d] Experimental data taken from ref *36*. [e] $\Delta\Delta G_U(H_2O)$ may be too low (i.e., 3 kJ or less) for ϕ to be accurate (*20*, *50*).

the largest deviation between experiment and theory (Table 1). However, this mutation only shows a ~3 kJ change in the experimental $\Delta G_U(H_2O)$ value (Table 1), which may make our experimental ϕ-value calculation somewhat unreliable (*20*, *50*). If we exclude the point for Leu125Ala in Figure 6A, the correlation coefficient increases to 0.90. The correlation between theoretical, $\Delta G(k_BT)$, and experimental, $\Delta G_U(H_2O)$, stability values for wild-type and mutant apo-azurins is also excellent; γ is 0.88 (Figure 6B). Taken together, the comparisons show that using the PTW varia-

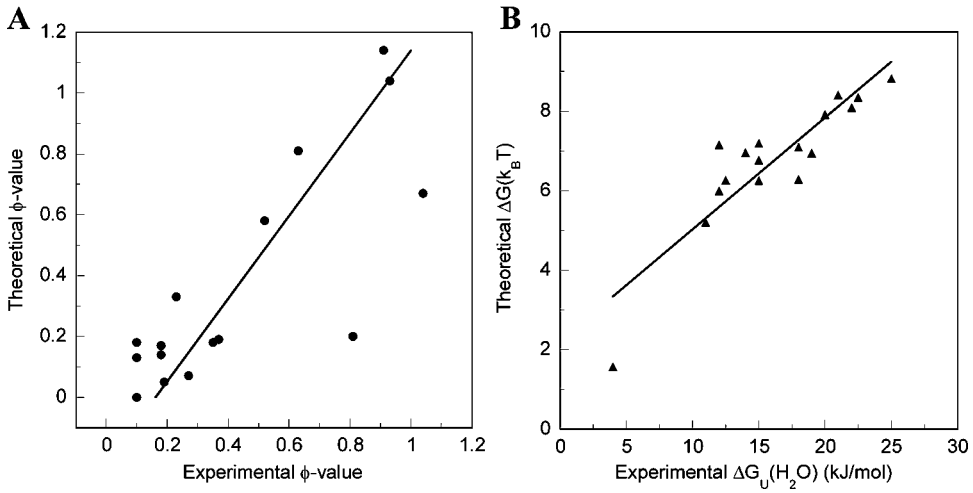

FIGURE 6: (A) Correlation between calculated and experimental ϕ-values (data from Table 1) at room temperature. (B) Correlation between calculated and experimental stability ($\Delta G(k_BT)$ and $\Delta G_U(H_2O)$; data from Table 1) values at room temperature.

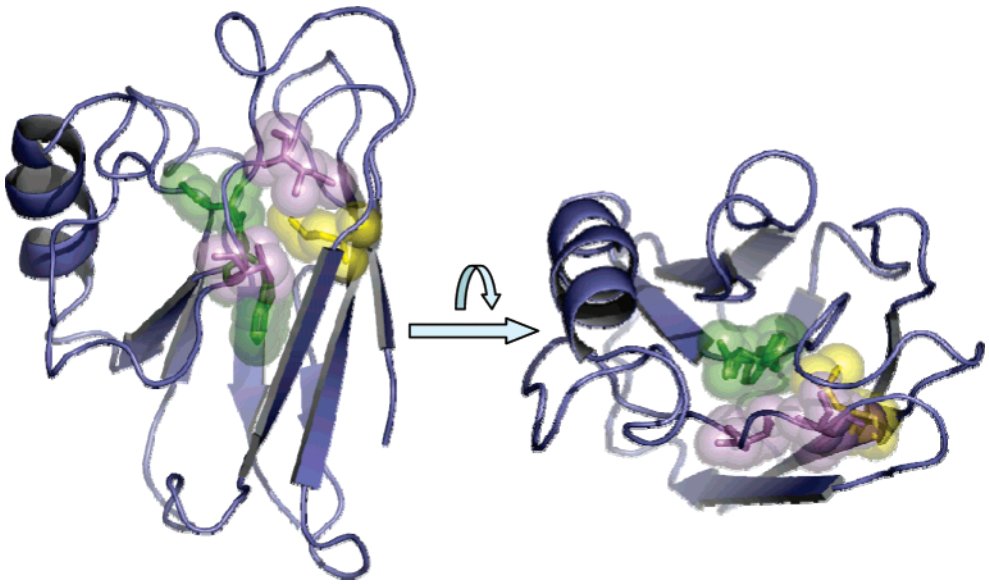

FIGURE 7: Illustration that shows the packing of Leu33 (yellow), Thr84 and Ile87 (purple), and Asn47 and Trp48 (green) in apo-azurin's core (prepared in PyMol using pdb file 1E65).

tional method to calculate folding dynamics provides chevron plots and, thus, ϕ-values that are in fine agreement with data from in vitro protein-folding experiments. Furthermore, from the ϕ-values for the 16 positions, it emerges that apo-azurin has a localized TSE with almost nativelike interactions around Val31, Val33, Leu50, Ile81, and Val95 bringing parts of β-strands 3−6 together. The remaining positions, which cover all other secondary-structure elements in azurin (i.e., β-strands 1, 2, 7, and 8, as well as the α-helix), are not structured in apo-azurin's TSE for folding. This description of the TSE for folding of apo-azurin is in good agreement with the isodensity plot for the low-energy barrier TSE (Figure 2B, left).

In the theoretical analysis of Leu33Ala apo-azurin, it became evident that the choice of contacts to be eliminated to accurately describe this variant must be carefully evaluated. In wild-type, folded apo-azurin, residues 47, 48, 84, and 87 form contacts with the side chain of residue 33. To define the Leu33Ala variant in the theoretical calculations, all four contacts should be removed based on the distance criterion. However, deletion of all four contacts altered the folding barrier significantly more than what was actually observed in the in vitro experiments (i.e., ϕ became unreasonably high). To address this discrepancy, we evaluated the packing quality between the involved residues. As shown in Figure 7, the side chain of leucine-33 is mainly packed with residues 84 and 87. Therefore, the contacts with residues 84 and 87 are considered to be side-chain contacts, but the contacts with 47 and 48 are instead classified as backbone contacts. When we only deleted the contacts between residue 33 and residues 84 and 87 to create the azurin variant Leu33Ala, both the stability and the ϕ-value agreed well with the experimental observations. For the variant Leu31Ala, based on the same packing argument as in the case of Leu33Ala, only the contact with residue 48 was deleted, since leucine-31 is mainly packed against the indole group of tryptophan-48. These examples suggest that side-chain contacts, especially those in the hydrophobic core, to be eliminated in theoretical descriptions of mutated variants cannot be determined by distance criteria alone. To match completely the experimental data, fine analysis of side-chain packing is important. The deleted contacts for each

theoretical description of the apo-azurin variants studied here are summarized in Table 1.

Here, we have compared ϕ-values derived from theoretical experiments with ϕ-values calculated from in vitro chemical-denaturant measurements. Although temperature-jump experiments may seem more appropriate in providing experimental data that correspond to the theoretical analysis, temperature changes can induce nontrivial dynamic solvent effects (*51*, *52*), which can complicate straightforward analysis based on equilibrium free-energy landscape theory. Importantly, unfolding of apo-azurin when induced by temperature- and chemical-denaturant perturbations results in similar unfolded states with respect to far-UV CD and fluorescence characteristics (data not shown); moreover, both methods of perturbation are reversible and correspond to two-state transitions (*35*, *36*, *39*). From a technical perspective, chemical-denaturant jumps are less complicated to execute as compared to temperature jumps when complete chevron plots are desired.

CONCLUSIONS

We used the PTW variational method to investigate the folding TSE for the β-sandwich protein *P. aeruginosa* azurin. We also prepared all azurin variants in the lab and tested their folding behavior in vitro via kinetic measurements. We found excellent agreement between theoretically and experimentally determined ϕ-values: apo-azurin's transition state is fixed with a set of nativelike interactions involving core residues from strands in both β-sheets. Detailed comparisons of theoretical and experimental data demonstrate that fine-tuning is needed in the theoretical description of point-mutated variants. Free-energy functional methods, such as the PTW scheme, allow one to readily calculate chevron plots/ϕ-values and, therefore, compare theoretical results directly with experimental measurements. For apo-azurin, this direct comparison shows that theory and experiment agree in a quantitative fashion. Combined theoretical/experimental studies to investigate how the presence of zinc induces a switch from a fixed to a moving TSE in azurin are underway.

SUPPORTING INFORMATION AVAILABLE

Correlation of ϕ-values calculated at the transition midpoints and equilibrium-unfolding curves detected by fluorescence and far-UV CD. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Wolynes, P. G. (2005) Energy landscapes and solved protein-folding problems, *Philos. Trans. R. Soc. London, Ser. A 363*, 453−464; discussion 464−467.
2. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: the energy landscape perspective, *Annu. Rev. Phys. Chem. 48*, 545−600.
3. Onuchic, J. N., and Wolynes, P. G. (2004) Theory of protein folding, *Curr. Opin. Struct. Biol. 14*, 70−75.
4. Bryngelson, J. D., and Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding, *Proc. Natl. Acad Sci. U.S.A. 84*, 7524−7528.
5. Portman, J. J., Takada, S., and Wolynes, P. G. (2001) Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach, *J. Chem. Phys. 114*, 5069.
6. Portman, J. J., Takada, S., and Wolynes, P. G. (2001) Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics, *J. Chem. Phys. 114*, 5082.
7. Portman, J. J., Takada, S., and Wolynes, P. G. (1998) Variational theory for site resolved protein folding free energy surfaces, *Phys. Rev. Lett. 81*, 5237.
8. Oliveberg, M. (2001) Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis, *Curr. Opin. Struct. Biol. 11*, 94−100.
9. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots, *Biochemistry 38*, 6499−6511.
10. Sanchez, I. E., and Kiefhaber, T. (2003) Hammond behavior versus ground-state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states, *J. Mol. Biol. 327*, 867−884.
11. Sanchez, I. E., and Kiefhaber, T. (2003) Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding, *Biophys. Chem. 100*, 397−407.
12. Wilson, C. J., and Wittung-Stafshede, P. (2005) Snapshots of a dynamic folding nucleus in zinc-substituted *Pseudomonas aeruginosa* azurin, *Biochemistry 44*, 10054−10062.
13. Shen, T., Hofmann, C. P., Oliveberg, M., and Wolynes, P. G. (2005) Scanning malleable transition state ensembles: comparing theory and experiment for folding protein U1A, *Biochemistry 44*, 6433−6439.
14. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding Des. 3*, R81−91.
15. Kamagata, K., Arai, M., and Kuwajima, K. (2004) Unification of the folding mechanisms of non-two-state and two-state proteins, *J. Mol. Biol. 339*, 951−965.
16. Lindorff-Larsen, K., Vendruscolo, M., Paci, E., and Dobson, C. M. (2004) Transition states for protein folding have native topologies despite high structural variability, *Nat. Struct. Mol. Biol. 11*, 443−449.
17. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) Transient folding intermediates characterized by protein engineering, *Nature 346*, 440−445.
18. Matouschek, A., and Fersht, A. R. (1991) Protein engineering in analysis of protein folding pathways and stability, *Methods Enzymol. 202*, 82−112.
19. Fersht, A. R., and Sato, S. (2004) Phi-Value analysis and the nature of protein-folding transition states, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7976−7981.
20. Sanchez, I. E., and Kiefhaber, T. (2003) Origin of unusual phi-values in protein folding: evidence against specific nucleation sites, *J. Mol. Biol. 334*, 1077−1085.
21. Goldenberg, D. P. (1999) Finding the right fold, *Nat. Struct. Biol. 6*, 987−990.
22. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding, *J. Mol. Biol. 254*, 260−288.
23. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding, *Biochemistry 34*, 13914−13919.
24. Lopez-Hernandez, E., and Serrano, L. (1996) Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2, *Folding Des. 1*, 43−55.
25. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions, *Nat. Struct. Biol. 4*, 305−310.
26. Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J., and Poulsen, F. M. (1999) The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP, *Nat. Struct. Biol. 6*, 594−601.
27. Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state, *Nat. Struct. Biol. 9*, 126−130.
28. Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002) Im7 folding mechanism: misfolding on a path to the native state, *Nat. Struct. Biol. 9*, 209−216.

29. Kim, D. E., Fisher, C., and Baker, D. (2000) A breakdown of symmetry in the folding transition state of protein L, *J. Mol. Biol. 298*, 971−984.
30. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding, *Nat. Struct. Biol. 6*, 1016−1024.
31. Adman, E. T. (1991) Copper protein structures, *Adv. Protein Chem. 42*, 145−197.
32. Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M., and Canters, G. W. (1992) Crystal structure of *Pseudomonas aeruginosa* apo-azurin at 1.85 Å resolution, *FEBS Lett. 306*, 119−124.
33. Nar, H., Huber, R., Messerschmidt, A., Filippou, A. C., Barth, M., Jaquinod, M., van de Kamp, M., and Canters, G. W. (1992) Characterization and crystal structure of zinc azurin, a by-product of heterologous expression in *Escherichia coli* of *Pseudomonas aeruginosa* copper azurin. *Eur. J. Biochem. 205*, 1123−1129.
34. Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. (2001) Copper stabilizes azurin by decreasing the unfolding rate, *Arch. Biochem. Biophys. 390*, 146−148.
35. Pozdnyakova, I., and Wittung-Stafshede, P. (2001) Copper binding before polypeptide folding speeds up formation of active (holo) *Pseudomonas aeruginosa* azurin, *Biochemistry 40*, 13728−13733.
36. Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. (2002) Studies of *Pseudomonas aeruginosa* azurin mutants: cavities in beta-barrel do not affect refolding speed, *Biophys. J. 82*, 2645−2651.
37. Pozdnyakova, I., and Wittung-Stafshede, P. (2003) Approaching the speed limit for Greek Key beta-barrel formation: transition-state movement tunes folding rate of zinc-substituted azurin, *Biochim. Biophys. Acta 1651*, 1−4.
38. Fuentes, L., Oyola, J., Fernandez, M., and Quinones, E. (2004) Conformational changes in azurin from *Pseudomona aeruginosa* induced through chemical and physical protocols, *Biophys. J. 87*, 1873−1880.
39. Wittung-Stafshede, P. (2004) Role of cofactors in folding of the blue-copper protein azurin, *Inorg. Chem. 43*, 7926−7933.
40. Wilson, C. J., and Wittung-Stafshede, P. (2005) Role of structural determinants in folding of the sandwich-like protein *Pseudomonas aeruginosa* azurin, *Proc. Natl. Acad. Sci. U.S.A. 102*, 3984−3987.
41. Marks, J., Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. (2004) Methionine-121 coordination determines metal specificity in unfolded *Pseudomonas aeruginosa* azurin, *J. Biol. Inorg. Chem. 9*, 281−288.
42. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman and Company, New York.
43. Pace, C. N., and Shaw, K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves, *Proteins* (Suppl.) *4*, 1−7.
44. Fersht, A. R. (1997) Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol. 7*, 3−9.
45. Miyazawa, S., and Jernigan, R. L. (1996) Residue−residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading, *J. Mol. Biol. 256*, 623−644.
46. Engeseth, H. R., and McMillin, D. R. (1986) Studies of thermally induced denaturation of azurin and azurin derivatives by differential scanning calorimetry: evidence for copper selectivity, *Biochemistry 25*, 2448−2455.
47. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
48. Silow, M., and Oliveberg, M. (2003) High concentrations of viscogens decrease the protein folding rate constant by prematurely collapsing the coil, *J. Mol. Biol. 326*, 263−271.
49. Krieger, F., Fierz, B., Axthelm, F., Joder, K., Meyer, D., and Kiefhaber, T. (2004) Intrachain diffusion in a protein loop fragment from carp parvalbumin, *Chem. Phys. 307*, 209.
50. de los Rios, M. A., Muralidhara, B. K., Wildes, D., Sosnick, T. R., Marqusee, S., Wittung-Stafshede, P., Plaxco, K. W., and Ruczinski, I. (2006) On the precision of experimentally determined protein folding rate and phi-values, *Protein Sci. 15*, 553−563.
51. Socci, N. D., Onuchic, J. N., and Wolynes, P. G. (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels, *J. Chem. Phys. 104*, 5860.
52. Lee, C.-L., Stell, G., and Wang, J. (2003) First-passage time distribution and non-Markovian diffusion dynamics of protein folding, *J. Chem. Phys. 118*, 959.

BI060025W