Folding Mechanism of a Multiple Independently-Folding Domain Protein: Double B Domain of Protein A[†]

Pooja Arora,[‡,§] Gordon G. Hammes,[||] and Terrence G. Oas*[,‡,||]

Departments of Chemistry and Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

Received May 10, 2006; Revised Manuscript Received July 28, 2006

ABSTRACT: The antibody binding properties of staphylococcal protein A (SpA) can be attributed to the presence of five highly homologous domains (E, D, A, B, and C). Although the folding of the B domain of protein A (BdpA) is well-characterized, the folding behavior of this domain in the context of full-length SpA in the cell remains unexplored. The sequence of the B domain is 89 and 91% identical to those of domains A and C, respectively. We have fused B domain sequences (BBdpA) as a close approximation of the A−B or B−C portion of SpA. Circular dichroism and fluorescence-detected denaturation curves of BBdpA are experimentally indistinguishable from those of BdpA. The rate constants for folding and unfolding from NMR line shape analysis for the single- and double-domain proteins are the same within experimental uncertainties (±20%). These results support the designation of SpA as a multiple independently-folding domain (MIFD) protein. We develop a mathematical model that describes the folding thermodynamics and kinetics of MIFD proteins. The model depicts MIFD protein folding and unfolding as a parallel network and explicitly calculates the flux through all parallel pathways. These fluxes are combined to give a complete description of the global thermodynamics and kinetics of the folding and unfolding of MIFD proteins. The global rates for complete folding and unfolding of a MIFD protein and those of the individual domains depend on the stability of the protein. We show that the global unfolding rate of a MIFD protein may be many orders of magnitude slower than that of the constituent domains.

Although folding pathways of numerous single-domain proteins have been mapped experimentally and theoretically (*1*), few systematic studies have focused at the same level on proteins with multiple domains (*2*−*6*). Protein domains are defined as substructures that are capable of folding into their native structure when excised from the parent protein (*7*−*9*). These structural elements are usually identified by using one or more of the following criteria: structural and/or functional analysis, deconvolution of thermal transitions, limited proteolysis, and sequence homology at the amino acid and/or DNA level (*9*). The simplicity associated with single-domain proteins makes them commonly used models for protein folding studies. The results from these studies have provided scientists with valuable insights into the processes associated with the interconversion of a polypeptide chain between its native and denatured states. Presumably, most of the general principles learned from these studies apply to the understanding of the folding of bigger, more complex systems with multiple domains (*1*). However, because most proteins are multidomain in nature [more than 75% in eukaryotes (*10*)], it is crucial to test these principles with multiple domain proteins. In this study, we present experimental results and a mathematical model to characterize the folding behavior of proteins with multiple domains that are thermodynamically and kinetically uncoupled: multiple independently-folding domain (MIFD)[1] proteins.

The extent of coupling between domains in a multidomain protein is determined by the magnitude of energetic interaction between them. The domains of a multidomain protein are defined as being coupled if the folding behaviors of a constituent domain in isolation versus that in the presence of neighboring domains are significantly different. If the magnitude of the coupling is sufficiently high, the multidomain protein unfolds as a single, cooperative unit (*5*, *11*, *12*). Elegant studies have been preformed previously using various natural proteins constructed from linear arrays of tandem repeats that show a high degree of cooperativity among the repeating units (*13*−*17*). Conversely, in the absence of significant coupling, the presence of neighboring domains in a multidomain protein does not perturb the thermodynamic and kinetic parameters associated with the unfolding of each domain (*3*, *4*, *19*−*21*). Therefore, to determine the extent of coupling present in a multidomain protein, experimental studies of constituent domains in both isolation and multidomain proteins are necessary.

Five tandemly repeated antibody binding domains comprise the N-terminal half of protein A from *Staphylococcus aureus* (Swiss-protein entry SPA2_STAAU, Primary Accession Number P38507). As depicted in Figure 1, these domains have a high degree of sequence identity and are

[†] Supported by National Institutes of Health Grant GM045322.

* To whom correspondence should be addressed: Box 3711, Duke University Medical Center, Durham, NC 27710. Telephone: (919) 684-4363. Fax: (919) 681-8862. E-mail: oas@duke.edu.

[‡] Department of Chemistry.

[§] Current address: KBI BioPharma Inc., 1101 Hamlin Rd., Durham, NC 27704.

[||] Department of Biochemistry.

[1] Abbreviations: MIFD, multiple independently-folding domain; CD, circular dichroism; NMR, nuclear magnetic resonance.

10.1021/bi060923s CCC: $33.50
Published on Web 09/19/2006

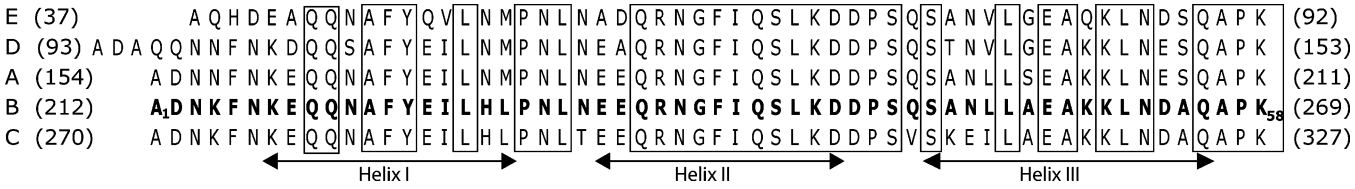


FIGURE 1: Sequence alignment of the five antibody binding domains (E, D, A, B, and C) of SpA shows 65−90% sequence identity between each domain. The start and end of each domain are numbered in parentheses in accordance with Uniprot entry SPA2_STAAU. The residues conserved among the five domains are highlighted in boxes. The BdpA sequence is highlighted in bold and is numbered from A_1 to K_{58} in this paper. Residues K_7-L_{17}, $E_{24}-D_{36}$, and $S_{41}-A_{54}$ constitute the three helices in BdpA (*18*). BBdpA has an additional A_1-K_{58} (with F13W and G29A substitutions) sequence immediately after K_{58} from the first domain.

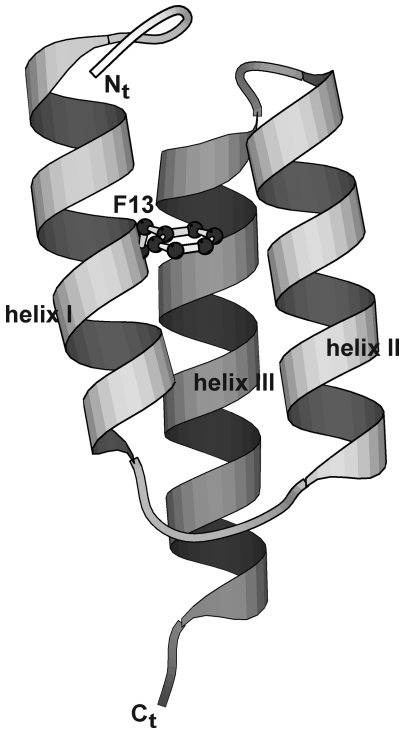


FIGURE 2: Ribbon diagram of BdpA showing the side chain of F13 that was substituted with Trp for the studies described here. This figure was generated using MOLSCRIPT (*22*), and the coordinates were obtained from PDB entry 2SPZ (model 2) (*18*).

designated, starting from the N-terminus, E, D, A, B, and C. The B domain has been structurally and functionally well characterized; however, important questions regarding the behavior of these domains in the presence of each other in full-length protein A (SpA) remain unanswered. In addition to these five domains, SpA consists of an N-terminal signal sequence (36 residues) and a C-terminal sorting sequence (180 residues) that tethers protein to the peptidoglycan moiety of the cell wall. The amino acid sequence alignment of the five domains of protein A is shown in Figure 1. The B domain of protein A (BdpA), the NMR-based structure of which is shown in Figure 2, has been well-studied. BdpA in conjunction with four other domains (E, D, A, and C) is responsible for the antibody binding properties of protein A. Previous studies have shown that this domain folds cooperatively in a two-state manner under all experimental conditions (*23*−*27*). Similar studies of BdpA when attached to its neighboring domains would provide insights into the folding of full-length protein A. These studies would answer the following questions: (1) What is the extent of thermodynamic and kinetic coupling between the folding of the repeated antibody binding domains? (2) How does the behavior of a domain compare in the presence and absence of other domains? (3) Does the extent of interaction between domains play a role in the efficiency of biological processes involving protein A such as secretion through the cell wall of *S. aureus*? As a step toward addressing these questions, we have constructed and characterized the folding of a protein that consists of two B domains fused together (BBdpA).

BBdpA is an excellent candidate for extending the folding studies of BdpA in an effort to understand the effects of neighboring domains (A and C) in full-length protein A. The primary sequence of BdpA is closest to the consensus sequence for the five antibody binding domains. Moreover, the degree of sequence similarity between BdpA and each of its adjacent domains is highest (71, 78, 89, and 91% with E, D, A, and C domains, respectively) among all the pairs. Therefore, the BBdpA construct is a very close approximation of the A−B or B−C portion of SpA.

In this work, we employed BBdpA as a model in an effort to understand the effect of neighboring domains of BdpA in protein A. We present experimental evidence that shows the thermodynamic coupling between the two domains in BBdpA, if any, is within the experimental uncertainty of ±0.7 kJ/mol. Also, the changes in kinetic parameters caused via addition of a second domain are experimentally insignificant (±20%). The absence of thermodynamic and kinetic coupling between the two domains in BBdpA suggests that the five domains of SpA fold and unfold completely independently of each other.

To extend further the understanding of the folding of MIFD proteins, we develop thermodynamic and kinetic models to describe the overall folding behavior of these systems. The results from the equilibrium model show that the thermodynamic parameters associated with the global unfolding of a MIFD protein can be determined by utilizing the parameters of each constituent domain. To derive the kinetic model, we utilize the rules of flux additivity to define the rate of flow of particles through *n*! pathways, where *n* is the number of domains (*28*). In the context of protein folding, flux is defined as the number of protein molecules folding or unfolding through a specific pathway per unit time. Our model demonstrates that the global folding and unfolding rates for a MIFD protein depend strongly on the values of the equilibrium constants.

MATERIALS AND METHODS

Cloning, Expression, and Purification of BBdpA

The start and end of each of the five domains in SpA are defined by the Universal protein knowledgebase (entry

SPA2_STAAU) and show the absence of a linker between any of the two adjacent domains (Figure 1). The molecular biology experiments in this study were designed to ensure the same in BBdpA. The 3′-end of the BdpA plasmid sequence with F13W and G29A substitutions (*27*) was examined carefully in designing primers for removal of the stop codon and creating a unique restriction site without changing the identity of the coded amino acids. A unique Bsu361 site was introduced by site-directed mutagenesis using the QuikChange procedure (Stratagene). This new restriction site is located near and on the 3′-side of a unique BamHI site near the 5′-end of the BdpA coding sequence. The resultant plasmid was amplified and purified using the QIAprep spin miniprep kit (Qiagen). The pure plasmid was digested with the two restriction enzymes to create a linear vector with Bsu361 and BamHI sticky ends at the 3′- and 5′-ends, respectively, for the second BdpA sequence to be inserted. In a parallel synthesis, the plasmid sequence encoding BdpA was PCR-amplified using primers that replaced the start codon at the 5′-end with the Bsu361 restriction site. The amplified sequence was doubly digested to produce Bsu361 and BamHI sticky ends at the 5′- and 3′-ends, respectively. The insert was purified from agarose gel and ligated to the linear vector produced in the parallel experiment to form the plasmid for the expression of BBdpA. The sequence of the resultant plasmid with F13W, G29A, F71W, and G87A substitutions was also confirmed.

BBdpA was expressed in *Escherichia coli* BL21(DE3) cells using the T7 expression system (*29*). The protein was purified as described previously (*25*), with an additional step of size exclusion chromatography on a Sephadex G25 column (Pharmacia) to remove a few higher-molecular mass impurities that coeluted with BBdpA in the first column-based purification step. Both SDS−PAGE and analytical reverse-phase HPLC confirmed the purity of the protein. The molecular mass of the purified protein was determined to be 13 284 Da by electrospray mass spectrometry (expected molecular mass of 13 282 Da).

CD- and Fluorescence-Based Equilibrium Experiments

For data collection, CD experiments were performed at 37 °C on an AVIV 62DS spectropolarimeter, equipped with a Hamilton syringe titration apparatus. The protein concentrations were determined in 6 M GuCl, using the absorbance measurements of Trp and Tyr residues at 276 and 280 nm (*30*). The native sample for the chemical denaturation experiments consisted of 10 μM protein, 100 mM NaCl, and 20 mM sodium acetate (pH 5). The denatured sample consisted of ~7 M GuCl in addition to the native sample components. Refractive index measurements were used to determine the exact concentration of denaturant (*31*). Different volumes of the native and denatured samples were mixed to attain intermediate GuCl concentrations. Each sample was stirred for 90 s before the CD signal was averaged at 222 nm for 30 s. Fluorescence-based chemical denaturation experiments were performed on an SLM model 8100 fluorimeter, with excitation and emission wavelengths of 280 and 340 nm, respectively.

For data analysis, chemical denaturation curves were fitted to a two-state model based on the linear dependence of the free energy of unfolding (ΔG_{D-N}) on denaturant concentration (*32*):

$$y = \frac{y_N + m_N[\mathrm{GuCl}] + (y_D + m_D[\mathrm{GuCl}]) \times \exp[m_{eq}([\mathrm{GuCl}] - [\mathrm{GuCl}]_{1/2})/RT]}{1 + \exp[m_{eq}([\mathrm{GuCl}] - [\mathrm{GuCl}]_{1/2})/RT]} \quad (1)$$

where y is the CD signal at 222 nm, y_N and y_D are intercepts for the pretransition (native) and post-transition (denatured) baselines, respectively, m_N and m_D are gradients of the native and denatured baselines, respectively, m_{eq} is the gradient of the linear dependence of ΔG_{D-N} on [GuCl], T is temperature (310 K), and R is the universal gas constant (8.314 J mol^{-1} K^{-1}). The protein fraction present in the denatured form (F_D) was determined at various GuCl concentrations as

$$F_D = \frac{y_N + m_N[\mathrm{GuCl}] - y}{y_N + m_N[\mathrm{GuCl}] - (y_D + m_D[\mathrm{GuCl}])} \quad (2)$$

The linear extrapolation method (*32*) was used to determine the protein stability (ΔG^o_{D-N}) in the absence of denaturant:

$$\Delta G_{D-N} = \Delta G^o_{D-N} - m_{eq}[\mathrm{GuCl}] \quad (3)$$

Equilibrium experiments with the two-domain protein were also analyzed using eqs 1−3. The parameters obtained from this analysis represent properties expressed on a per-mol-of-domain basis, rather than per-mole-of-protein.

Kinetic Measurements

NMR Experiments. ^{1}H NMR spectra were collected at pH 5 and 310 K at different GuCl concentrations as described previously (*25*). The Hϵ peaks of His18 and His76 were fitted using Mathematica 5 (Wolfram Research) to equations described in refs *33* and *34*. The spectra under native conditions (<2.5 M GuCl) were fitted for two sharp peaks corresponding to His18 and His76, resulting in two native baselines. The denatured baseline was obtained by fitting the single sharp peak observed for His18 and His76 Hϵ at >5 M GuCl. The intercepts and slopes of the best-fit native and denatured lines were used to fit the broad peaks in the transition region. The linear dependence of the folding and unfolding rate constants on [GuCl] was used to determine the rate constants in the absence of denaturant:

$$\ln(k_f) = \ln(k^o_f) + m_f[\mathrm{GuCl}] \quad (4)$$

$$\ln(k_u) = \ln(k^o_u) + m_u[\mathrm{GuCl}] \quad (5)$$

where $\ln(k^o_f)$ and $\ln(k^o_u)$ are the folding and unfolding constants, respectively, in the absence of denaturant. The slopes of the best-fit lines, m_f and m_u, were used to calculate the beta-Tanford parameter (β_T) that is proportional to the

solvent accessibility of the transition state compared to that of the native state (*35*):

$$\beta_T = m_f/(m_u - m_f) \quad (6)$$

Stopped-Flow Fluorescence Experiments. The kinetic measurements were conducted using the Applied Photophysics SX 18MV stopped-flow instrument (dead time of ~1.5 ms) at 20 °C, in the presence of 3−4 M GuCl. The fluorescence intensity of W13 (and W71 for BBdpA) was measured as a function of time for 100 ms. Twelve scans were averaged for each denaturant concentration. Resultant decay curves were fitted to a single-exponential equation to yield k_{obs}:

$$y = y_o \exp(-k_{obs}t) + c \quad (7)$$

k_{obs} ($=k_u + k_f$) was used in conjunction with K_{eq} ($=k_u/k_f$) from equilibrium experiments to calculate k_u and k_f.

RESULTS

Equilibrium Studies Using CD and Fluorescence. F13W and G29A were used as the background substitutions for the single- and double-domain constructs in this study. These substitutions in BdpA have been shown in our previous studies to result in insignificant structural perturbations in the native fold of the protein (*25*−*27*). Also, these substitutions result in a marginal (~0.8 kJ/mol) increase in protein stability, reflected primarily in a 3-fold increase in the folding rate constant (*25*−*27*).

The unperturbed near- and far-UV CD spectrum of BBdpA compared to that of BdpA shows that the three-helix bundle of BdpA is retained in each domain of BBdpA (data not shown). The CD signal at 222 nm for BBdpA was approximately twice as great for BdpA at the same protein concentration, resulting in similar mean residue ellipticities. These results show that the presence of the second domain in BBdpA does not result in significant structural perturbation of the constituent domains. This conclusion is also supported by a previous NMR-based study of a similar double-B domain construct, which showed single peaks for most resonances (*36*).

The chemical denaturation curves obtained using CD and fluorescence suggest that BBdpA unfolds cooperatively in a two-state manner. In these experiments, the far-UV CD signal at 222 nm or fluorescence intensity (from W13 and W71) was measured as a function of guanidinium chloride (GuCl) concentration. The fluorescence and far-UV CD signal report on changes in the protein tertiary structure around fluorophore(s) and the amount of helical content, respectively. BBdpA demonstrated a sigmoidal, cooperative transition that fits well to an equation describing the two-state protein unfolding (eq 1). The important thermodynamic parameters obtained from the two-state analysis of such experimental data are m_{eq}, C_{mid}, and ΔG^o_{D-N}. The m_{eq} value defines the linear dependence of the free energy of protein unfolding on denaturant concentration. Extrapolation of this linear dependence yields the free energy of unfolding in the absence of denaturant, ΔG^o_{D-N}. C_{mid} is the concentration of denaturant at which fractional populations of native and denatured states of protein are equal, or ΔG_{D-N} equals zero. The two-state analysis of chemical transition curves obtained using CD and fluorescence for BBdpA results in the fraction denatured versus denaturant concentration curves shown in Figure 3. The differences between the best-fit thermodynamic parameters obtained from the two probes were smaller than the experimental uncertainty of ±5% (Table 1).

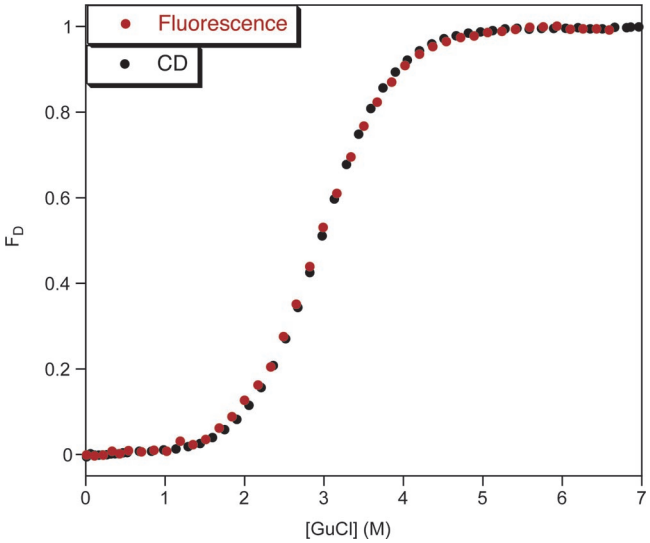


FIGURE 3: Chemical denaturation experiments of BBdpA using CD and fluorescence. The far-UV CD signal at 222 nm and fluorescence intensity of W13 and W71 were measured as a function of GuCl concentration. Two-state analysis of the resultant data (eq 1) yields fraction denatured (F_D) as a function of GuCl concentration. Excellent agreement between the two probes suggests the two-state unfolding of BBdpA, with no significant populations of partially folded intermediates.

Table 1: Thermodynamic Parameters Associated with the Unfolding of BBdpA and BdpA[a]

	m_{eq} (kJ mol^{-1} M^{-1})	$C_{1/2}$ (M)	ΔG^o_{D-N} (kJ/mol)
BBdpA, fluorescence	5.43 ± 0.21	2.92 ± 0.02	15.9 ± 0.6
BBdpA, CD	5.85 ± 0.21	2.94 ± 0.04	17.2 ± 0.7
BdpA, CD	5.96 ± 0.21	2.96 ± 0.02	19.2 ± 0.7

[a] The parameters are the best-fit values obtained using the two-state model (eq 1) to fit the chemical denaturation curves. For BBdpA, m_{eq} and ΔG are expressed in terms of moles of domains, rather than moles of protein (see text).

These results show that the secondary and tertiary structures of BBdpA change in a concerted manner as a function of denaturant concentration. The agreement between two independent probes is a very common test for verifying the absence of significant populations of partially folded intermediates under any experimental condition. Therefore, our experimental results suggest that only two thermodynamic states of protein are populated significantly under all experimental conditions. However, we show in the following sections that partially folded intermediates with only one of the domains folded are populated under certain experimental conditions. These species are not detected experimentally since the folding or unfolding of one domain does not perturb the spectroscopic properties of the adjacent domain, as discussed later.

Comparison of the equilibrium results obtained for BBdpA with those of BdpA shows that the two domains in BBdpA are not thermodynamically coupled within the experimental uncertainties. The fraction denatured versus denaturant concentration curves obtained using CD and two-state

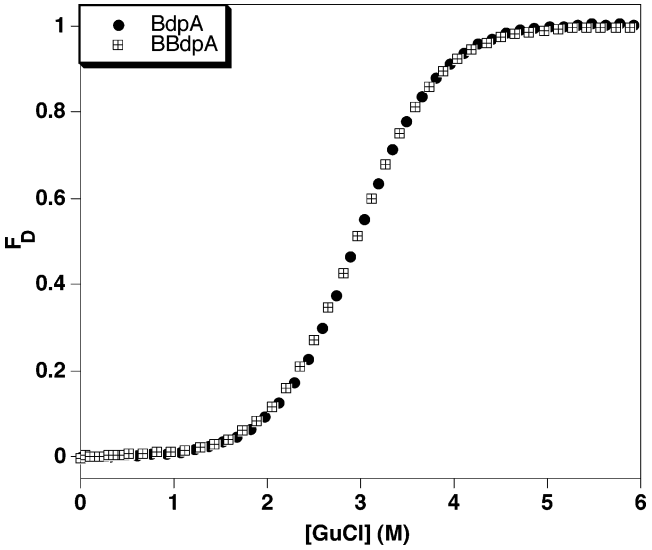


FIGURE 4: Fraction denatured vs GuCl concentration for BdpA and BBdpA from the two-state analysis of CD-based chemical denaturation experiments. The overlap between the two curves illustrates that the two domains in BBdpA are thermodynamically uncoupled.

analysis for BdpA and BBdpA are shown in Figure 4. The best-fit thermodynamic parameters obtained from these two curves for the single- and double-domain proteins are within the experimental error of those of BdpA (Table 1). The only difference observed between the experimental behavior of BBdpA and BdpA is the magnitude of the CD or fluorescence signal. The presence of two domains in BBdpA results in an ~100% increase in the magnitude of the observed signal at the same protein concentration. Therefore, when its concentration is expressed in terms of moles of domains per liter, the experimentally observed behavior of BBdpA is identical to that of BdpA under the same conditions, but with twice the protein concentration. To show that the excellent agreement between the equilibrium results of BBdpA and BdpA was not the consequence of a dimeric nature of BdpA, we performed sedimentation equilibrium measurements by analytical ultracentrifugation and measured a molecular mass corresponding to a single molecule (data not shown). We conclude that the best-fit parameters obtained by the two-sate analysis of BBdpA describe the unfolding of each individual domain. The absence of significant changes in these parameters due to the presence of the second domain in BBdpA shows that the magnitude of thermodynamic coupling, if any, between the two domains in BBdpA is smaller than the experimental error of 0.7 kJ/mol.

In a previous study, however, Karimi et al. concluded that there is significant coupling between the two domains in a wild-type (WT) BBdpA construct. This conclusion was based on the higher cooperativity they observed in the thermal denaturation experiments of WT BBdpA compared to that of WT BdpA. We repeated the thermal denaturation experiments with our single- and double-domain protein constructs. The denaturation curves obtained from these experiments were superimposable (data not shown) and yielded thermodynamic parameters within experimental error of each other ($T_{m,BdpA} = 76.9 \pm 0.3$ °C, and $T_{m,BBdpA} = 77.4 \pm 0.3$ °C). The two-domain constructs used in this study and by Karimi et al. differ significantly at the protein termini. The C-terminus of the double-domain protein studied by Karimi et al. consisted of six extra amino acid residues that could be responsible for the minor experimental differences observed between single- and double-domain proteins. On the basis of the results from thermal and chemical denaturation experiments, we conclude that the thermodynamic coupling between the two domains in BBdpA is insignificant.

Kinetic Studies Using NMR Line Shape Analysis of His18 (and His76). In our previous studies, we utilized the position and shape of the Hϵ peak of solvent-exposed His18 (*25*, *27*) to determine the (un)folding kinetics of BdpA. This peak shifts ~300 Hz (~0.6 ppm) upon denaturation and shows significant changes in its line width due to the exchange of protein between the native and denatured states (*25*, *27*). In the study presented here, we performed similar analyses using the Hϵ peak of His18 and His76 at various denaturant concentrations to determine the unfolding and folding kinetics of BBdpA.

The chemical shifts of the His18 and His76 peaks were slightly different ($\Delta\delta \approx 0.03$ ppm) at low GuCl concentrations, probably due to small changes in the environment of these two equivalent residues in the two domains. Doubled resonances were also observed for some of the residues in a similar double-B domain protein (*36*). Karimi et al. attributed these observed changes to either the asymmetry at the domain interface or the presence of interactions between the two domains. However, our experimental results show that there are no significant (>0.7 kJ/mol) energetic interactions between the two domains (see the next section). Hence, we conclude that the observed discrepancies in chemical shifts are a consequence of its high sensitivity to slight changes in the environment of equivalent residues in the two domains. The line-shape fitting equations for BBdpA were modified to account for the differences in the chemical shifts of His18 and His76 (see Materials and Methods).

Figure 5 shows the folding and unfolding rate constants obtained from line shape analysis of His peak(s) of BdpA and BBdpA plotted versus denaturant concentration. The data obtained for each protein have significant scatter around the best-fit lines, which is reflected in the large range of the 95% uncertainty lines. The best-fit values obtained for the extrapolated rate constants in the absence of denaturant and gradients defining the linear dependence for the single- and double-domain protein using experimental data in Figure 5are very similar (Table 2). These results suggest that the magnitude of kinetic coupling between the two domains in BBdpA is smaller than the experimental uncertainties of these kinetic measurements.

Because of significant (20%) uncertainties associated with the parameters estimated from NMR line shape analysis of BBdpA, we repeated the kinetic measurements using stopped-flow fluorescence. These experiments were conducted at a lower temperature (20 °C instead of 37 °C) and in the presence of 3−4 M GuCl, which yielded measurable k_{obs} values. The single-exponential fits to the fluorescence kinetic traces yielded very similar k_{obs} values for BdpA and BBdpA (Figure S1 of the Supporting Information). In conjunction with the experimentally indistinguishable K_{eq} ($=k_u/k_f$) values of BdpA and BBdpA under similar experimental conditions, the similar k_{obs} ($=k_f+k_u$) yielded very similar folding and unfolding rate constants for the single- and double-domain proteins. Therefore, we conclude that the unfolding and folding kinetics of each domain in BBdpA

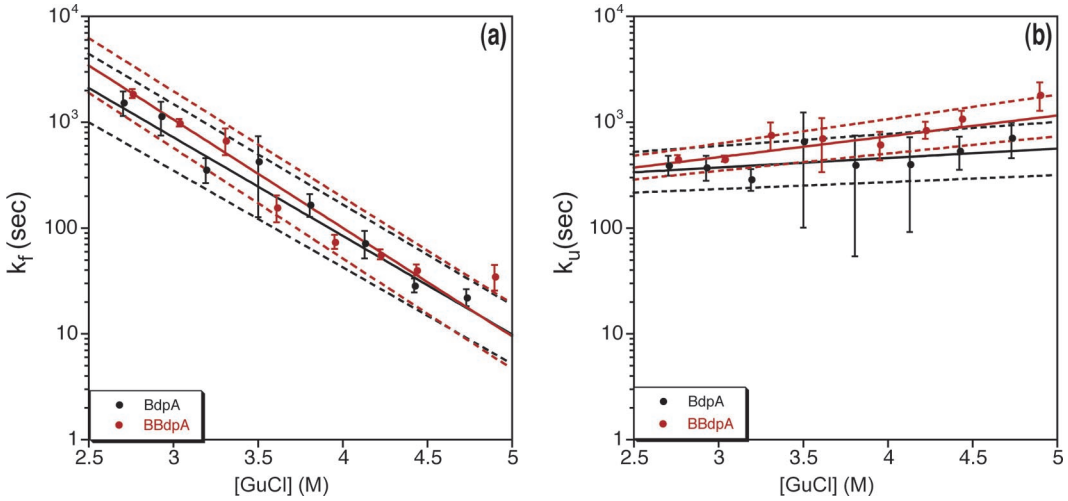


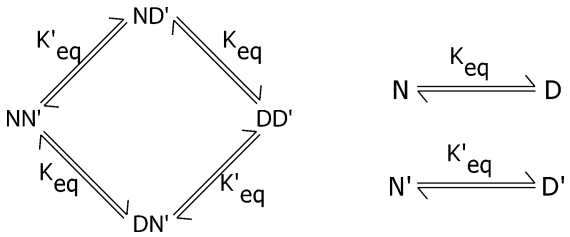
FIGURE 5: GuCl concentration dependence of folding (a, eq 4) and unfolding (b, eq 5) rate constants measured from the NMR line shape analysis of BdpA (black) and BBdpA (red). The solid and dashed lines represent the best-fit and 95% confidence interval lines for BdpA and BBdpA, respectively.

Table 2: Kinetic Parameters Associated with the Unfolding of BBdpA and BdpA Using NMR Line Shape Analysis[a]

	$\ln(k_f^o)$	m_f (kJ mol^{-1} M^{-1})	$\ln(k_u^o)$	m_u (kJ mol^{-1} M^{-1})	m_u (kJ mol^{-1} M^{-1})	β_T[b]
BBdpA	14 ± 0.5	-1.48 ± 0.18	4.8 ± 0.3	0.31 ± 0.21	2.9 ± 1.1	0.8 ± 0.1
BdpA	13 ± 0.5	-1.3 ± 0.18	5.3 ± 0.4	0.12 ± 0.21	2.9 ± 1.4	0.9 ± 0.1

[a] Best-fit values obtained from the linear dependence of folding (eq 4) and unfolding (eq 5) rate constants on GuCl concentration. [b] Calculated using eq 6.

Scheme 1



are unperturbed compared to those of BdpA, and hence, there is no significant kinetic coupling between the two domains in BBdpA.

Mathematical Model for Understanding the Equilibrium Unfolding of Proteins with Thermodynamically Uncoupled Domains. In this section, we derive the equilibrium parameters for the unfolding of a MIFD protein in terms of the parameters for the constituent domains. For a protein with n domains, there are 2^n possible thermodynamic states, including the fully native and fully denatured species (*38*). For a protein with thermodynamically uncoupled domains, the equilibrium constant associated with the unfolding of each domain is equal to that for the same domain in isolation.For example, for a protein with two uncoupled domains (Scheme 1)

$$\frac{[DN']}{[NN']} = \frac{[DD']}{[ND']} = \frac{[D]}{[N]} = K_{eq} \quad (8)$$

$$\frac{[ND']}{[NN']} = \frac{[DD']}{[DN']} = \frac{[D']}{[N']} = K'_{eq} \quad (9)$$

K_{eq} and K'_{eq} are the equilibrium constants that define the unfolding of the N and N′ domains, respectively. The prime symbol is used here to differentiate nonidentical domains of a MIFD protein. The global equilibrium constant associated with the complete unfolding of this two-domain protein (NN′ to DD′) can be defined using the K_{eq} values for the constituent domains as follows:

$$K_{eq,glob} = \frac{[DD']}{[NN']} = \frac{[DD']}{[ND']}\frac{[ND']}{[NN']} = K_{eq}K'_{eq} \quad (10)$$

In general, the global equilibrium constant associated with the complete unfolding of a MIFD protein with n domains is the product of equilibrium constants for all the constituent domains as shown in the following equation:

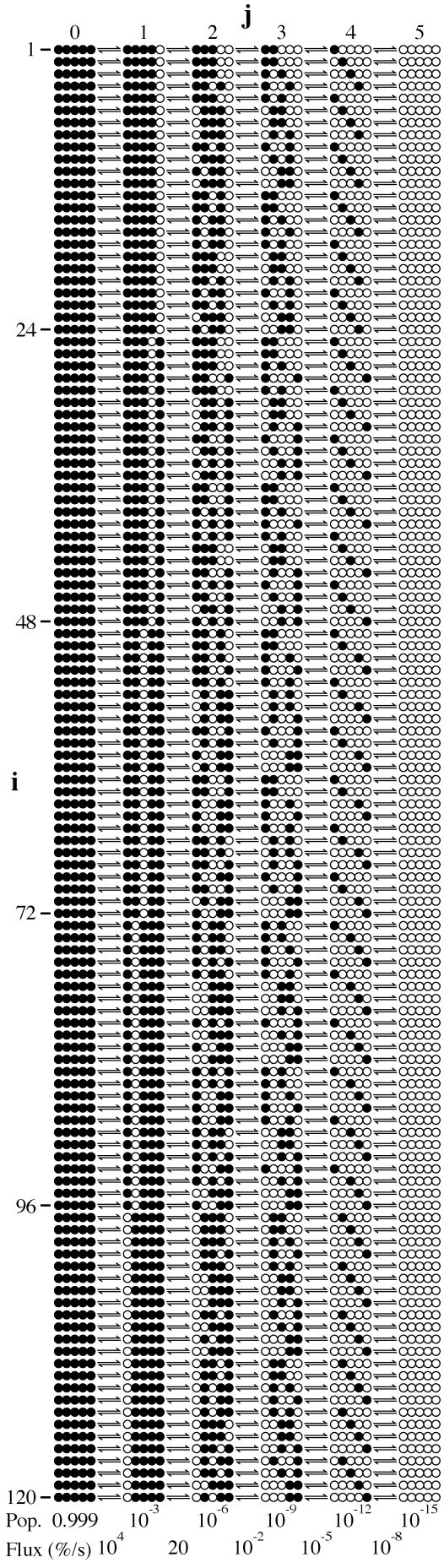
$$K_{eq,glob} = \prod_{i=1}^{n} K_{eq,i} \quad (11)$$

In the following derivation, we use i and j as indices to differentiate between various pathways and intermediates, respectively (Scheme 2). The index i denotes the pathway and varies from 1 to $n!$. The index j defines the identity of various states and varies from 0 to n. An intermediate denoted as $I_{i,j}$ is the thermodynamic species with j unfolded domains in the ith pathway.

To determine the concentration of each of the 2^n thermodynamic states for an n-domain system under any experimental condition, we use the fully native state ($j = 0$, N_n) as the reference species. The concentration of species with one or more domains unfolded ($j = 1$ to n) is equal to the product of the concentration of N_n and the equilibrium constants for all the j domains that are unfolded in that species as shown below:

$$[I_{i,j}] = [N_n]\prod_{j=1}^{j} K_{eq,i,j} \quad (12)$$

Scheme 2



where $K_{eq,i,j}$ is the equilibrium constant for the unfolding of a particular domain to form the jth intermediate on the ith pathway. Since the total protein concentration is distributed among the 2^n states, the sum of all concentrations obtained from eq 12 should sum to $[P]_{total}$. However, when the $n!$ pathways for the unfolding of an n domain protein are written explicitly as in Scheme 2, each state with j unfolded domains is repeated $(n - j)!j!$ times among these pathways. As a consequence, a simple summation of all the concentrations obtained using eq 12 would result in overcounting of various species. Therefore, $(n - j)!j!$ is the statistical factor that needs to be included while summing all the concentrations for the conservation of mass equation (eq A1 in the Appendix). Equation A1 can also be used to calculate the fractional population of the fully native species ($j = 0$):

$$f_{N_n} = \left[1 + \sum_{i=1}^{n!} \sum_{j=1}^{n} \frac{\prod_{k=1}^{j} K_{eq,i,k}}{(n-j)!j!}\right]^{-1} \tag{13}$$

The fractional population of all other thermodynamic species ($j > 0$) can be determined by dividing the concentration of that species by $[P]_{total}$ as shown below:

$$f_{I_{i,j}} = \frac{[I_{i,j}]}{[P]_{total}} = f_{N_n} \prod_{k=1}^{j} K_{eq,i,k} \tag{14}$$

If the K_{eq} values of all the constituent domains of a MIFD protein are known, eq 14 can be used to determine the fractional population of the various thermodynamic species. Equations 11, 12 and 14 can be further simplified to the following equations if the constituent domains are identical:

$$K_{eq,glob} = K_{eq,N_n \rightarrow D_n} = (K_{eq})^n \tag{15}$$

$$[I_{i,j}] = [N_n](K_{eq})^j \tag{16}$$

$$f_{I_{i,j}} = \frac{(K_{eq})^j}{(1 + K_{eq})^n} \tag{17}$$

Figure 6 shows the effects of addition of a second identical domain on the fractional populations of various thermodynamic species of a single-domain protein. For these simulation results, we used $\ln(K_{eq})$ as the independent variable to reflect changes in stability caused by, for example, a chemical denaturant. This variable transformation is straightforward if the single-domain m_{eq} and $C_{1/2}$ values are known: $\ln(K_{eq}) = m_{eq}/[RT([den] - C_{1/2})]$. The concentrations of four species (NN, ND, DN, and DD) for the double-domain protein become equal with experimental conditions under which the concentrations of two species (N and D) of the single-domain protein also become equal ($K_{eq} = 1$).

Kinetic Model for MIFD Proteins Derived from Rules of Flux Additivity. Here we derive a theoretical model in an effort to understand the changes in the unfolding and folding kinetics of a protein with two or more kinetically uncoupled domains. A protein with n domains has $n!$ independent pathways to completely unfold or fold (*38*). The number of protein molecules traversing each of these pathways can be defined using the concept of flux. The flux through a reaction

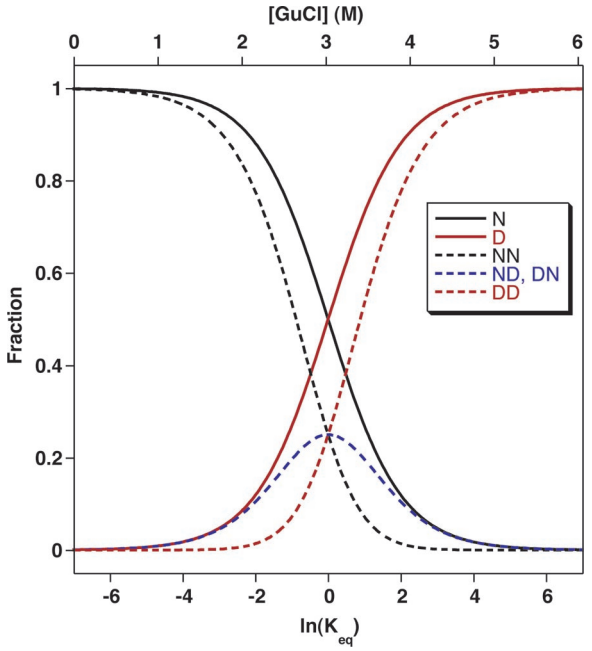


FIGURE 6: Simulations for the equilibrium model of a protein with two identical uncoupled domains. The fractional populations of various species for single-domain (solid lines) and double-domain (dashed lines) proteins are represented as a function of the change in the experimental conditions [$\ln(K_{eq})$, eq 17]. The fractional populations of ND and DN are identical under all experimental conditions. The top *x*-axis is calculated from [GuHCl] = $C_{1/2}$ + $(RT/m_{eq})\ln(K_{eq})$ using the m_{eq} and $C_{1/2}$ values for single-domain BdpA listed in Table 1.

is defined as the product of the forward rate constant and reactant concentration. At equilibrium, the forward and reverse fluxes are identical. The following two rules are used to calculate the total flux through multiple reactions in a system: (rule 1) $F_{total} = \sum_{i=1}^{n} F_i$ for the total flux through *n* parallel reactions, and (rule 2) $F_{total} = (\sum_{i=1}^{n} 1/F_i)^{-1}$ for the total flux through *n* sequential reactions.

In accordance with rule 1, the total flux for the global unfolding and folding of a multidomain protein ($D_n \rightarrow N_n$ or $N_n \rightarrow D_n$) through *n*! parallel pathways, such as those depicted in Scheme 2, is the arithmetic sum of the individual fluxes through each pathway. However, these pathways share multiple intermediates among each other, and therefore, a simple sum results in overcounting of fluxes. Hence, it is crucial to outline each pathway explicitly, with no cross equilibria between any of the pathways (Scheme 2). The sum of fluxes through these *n*! pathways defines the total flux as follows:

$$F_{total} = \frac{\sum_{i=1}^{n!} F_i}{n!} \tag{18}$$

The value of *n*! in the denominator of eq 18 is the statistical factor that accounts for the sharing of multiple states between the various pathways. Each of these *n*! pathways has *n* steps representing the (un)folding of the constituent domains in a different order. Therefore, rule 2 is used to calculate the total flux for each pathway via *n* sequential reactions as follows:

$$F_i = \left(\sum_{j=1}^{n} \frac{1}{F_{i,j}}\right)^{-1} \tag{19}$$

where $F_{i,j}$ is the flux associated with the formation of the *j*th intermediate in the *i*th unfolding pathway ($I_{i,j}$). $F_{i,j}$ through each of the *n* sequential steps is calculated using the definition of flux, by multiplying the forward (or unfolding in Scheme 2) rate constant by the concentration of the starting species as follows:

$$F_{i,j} = k_{u,i,j}[I_{i,j-1}] \tag{20}$$

where $I_{i,j-1}$ is the reacting intermediate with $j - 1$ unfolded domains in the *i*th pathway. The concentration of this partially folded species can be calculated using eq 12 from the previous section. Finally, the global rate constants associated with the complete folding and unfolding of a multidomain are calculated using the following equations:

$$k_{u,glob} = \frac{F_{total}}{[N_n]}; \; k_{f,glob} = \frac{F_{total}}{[D_n]} \tag{21}$$

As explained in the Appendix (eqs A2−A4), eqs 18−21 can be used in conjunction with eq 16 to calculate the total flux through *n*! pathways for the unfolding of a MIFD protein with *n* identical domains:

$$F_{total} = \frac{k_u[N_n]K_{eq}^{n-1}}{\sum_{j=1}^{n} K_{eq}^{j-1}} \tag{22}$$

The global unfolding rate constant can be determined by dividing F_{total} in eq 22 by $[N_n]$ (eq 21):

$$k_{u,glob} = k_u \frac{K_{eq}^{n-1}}{\sum_{j=1}^{n} K_{eq}^{j-1}} \tag{23}$$

Similar equations can be used to derive the rate constant for global folding of a MIFD protein with *n* identical domains:

$$k_{f,glob} = k_f \frac{1}{\sum_{j=1}^{n} K_{eq}^{j-1}} \tag{24}$$

Identical results are obtained if the rate laws for a multidomain protein are derived using the steady state approximations for all the partially folded intermediates, verifying the validity of our flux-based derivations.

DISCUSSION

The Two Domains in BBdpA Are Uncoupled. The experimentally indistinguishable thermodynamic behavior of the B domain in BdpA and BBdpA provides evidence for the absence of significant thermodynamic coupling between the two domains in BBdpA. The excellent agreement between thermodynamic results obtained from CD and fluorescence

suggests that the unfolding of BBdpA is two-state, with only the fully folded native state and the fully unfolded denatured state being populated significantly under all experimental conditions. However, our equilibrium model in the previous section suggests that the populations of partially folded species with only one domain folded become significant under specific experimental conditions. This discrepancy can be resolved by concluding that the CD or fluorescence signals from the native B domain are identical in NN, ND, and DN. Similarly, the signals of the denatured B domain in ND, DN, and DD must also be identical. Thus, the denaturation curve of BBdpA is identical to that of the single-domain protein at twice the concentration. Analysis of BBdpA equilibrium data yields thermodynamic parameters within experimental error of those obtained for BdpA, expressed per mole of domains. When expressed per mole of protein, these values double as expected for a protein containing two domains per protein molecule.

The folding and unfolding kinetics of the two domains in BBdpA also show no coupling. Experimental measurements using NMR line shape analysis (Figure 5) and stopped-flow fluorescence yielded the rate constants associated with the unfolding and folding of each domain of BBdpA. These rate constants are not significantly perturbed (±10%) relative to those of BdpA. Likewise, the slopes of $\ln(k_f)$ and $\ln(k_u)$ versus GuCl concentration (Figure 5, eq 6) are experimentally indistinguishable, indicating that the transition state is not perturbed in the double-domain protein (*35*). These results support the conclusion that the folding mechanism of the B domain is the same in the single-domain protein, the double-domain protein, and full-length protein A. This conclusion justifies the extensive experimental and theoretical work that has focused on the isolated B domain as being relevant to understanding the behavior of that domain in full-length protein A in the cell.

Mechanism of MIFD Protein Folding. The natural logarithms of rate constants for protein folding and unfolding usually exhibit a linear dependence on denaturant concentration, to which $\ln(K_{eq})$ is also linearly related. The natural logarithms of eqs 23 and 24 yield the following functions of K_{eq}:

$$\ln(k_{u,glob}) = \ln(k_u) + (n-1)\ln(K_{eq}) - \ln(\sum_{j=1}^{n} {K_{eq}}^{j-1}) \quad (25)$$

$$\ln(k_{f,glob}) = \ln(k_f) - \ln(\sum_{j=1}^{n} {K_{eq}}^{j-1}) \quad (26)$$

Evaluating eqs 25 and 26 using the values of k_u, k_f, and K_{eq} for BdpA shows that addition of each domain results in a more than 3 order of magnitude decrease in the global unfolding rate constant in the absence of denaturant (Figure 7). Under these conditions [$K_{eq} \ll 1$, $\sum_{j=1}^{n} {K_{eq}}^{j-1} \approx 1$, $\ln(\sum_{j=1}^{n} {K_{eq}}^{j-1}) \approx 0$], eqs 25 and 26 can be simplified to eqs 27 and 28, respectively:

$$\ln(k_{u,glob}) \approx \ln(k_u) + (n-1)\ln(K_{eq}) \quad (27)$$

$$\ln(k_{f,glob}) \approx \ln(k_f) \quad (28)$$

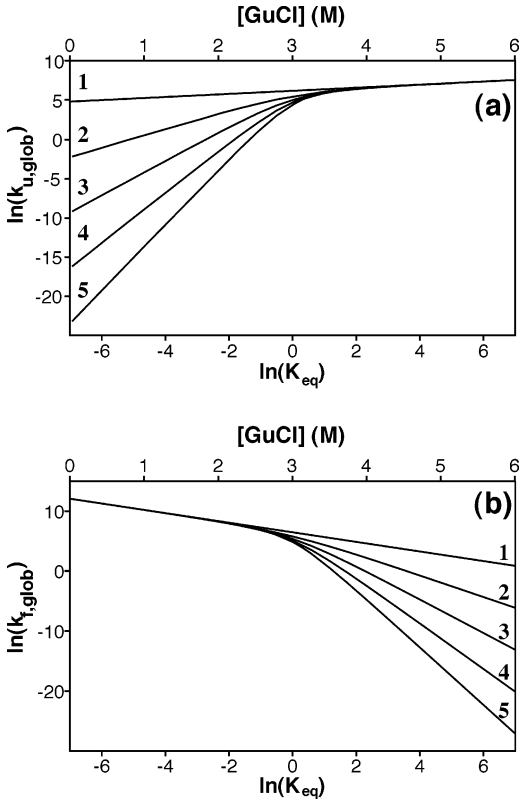


FIGURE 7: Simulations for the kinetic model of proteins with one to five identical, uncoupled domains. Perturbations in the global folding (eq 26) and unfolding (eq 25) rate constants are most pronounced under highly denaturing and native conditions, respectively. $\ln(K_{eq})$ was used as the independent variable for these simulations using the following equations: $\ln(k_f) = \ln(k_f^o) + [\ln(K_{eq}) - \ln(k_u^o/k_f^o)]\beta_T$ and $\ln(k_u) = \ln(k_u^o) + [\ln(K_{eq}) - \ln(k_u^o/k_f^o)]\beta_T$. The following parameters were used for the simulations: $\beta_T = 0.8$, $\ln(k_u^o) = 5$, and $\ln(k_f^o) = 12.3$. The top *x*-axis is calculated as described in the legend of Figure 6.

Similarly, the global folding kinetics differ significantly from those of a single domain only under denaturing conditions ($K_{eq} \gg 1$.) Under these conditions, addition of each domain results in a more than 4 order of magnitude decrease in the global folding rate constant. When $K_{eq} \gg 1$, the summation in eq 26 is approximately equal to the highest-order term ($\sum_{j=1}^{n} {K_{eq}}^{j-1} \approx {K_{eq}}^{n-1}$), so eqs 25 and 26 can be simplified to the following equations:

$$\ln(k_{u,glob}) \approx \ln(k_u) \quad (29)$$

$$\ln(k_{f,glob}) \approx \ln(k_f) - (n-1)\ln({K_{eq}}^{n-1}) \quad (30)$$

In the vicinity of the midpoint of the denaturation curve, where $K_{eq} \approx 1$, eqs 27−30 do not apply and eqs 25 and 26 must be used to calculate $k_{u,glob}$ and $k_{f,glob}$.

As depicted in Figure 7, the effect of adding domains on the global (un)folding kinetics of a MIFD protein depends on the stability of the protein. Under conditions that favor native individual domains, addition of domains has a significant effect on only the global unfolding rate constant and not the global folding rate constant. In contrast, under denaturing conditions, only the global folding rate constant is affected by additional domains. The biological implications of these mechanistic properties of MIFD protein folding are discussed below.

The magnitude of the difference between global (un)-folding kinetics of a MIFD protein and those of the individual domains depends on whether the experimental conditions are native or denaturing. Since the probability of multiple domains being unfolded simultaneously is the product of individual unfolded probabilities of each domain (all <1), we would intuitively expect the addition of domains to decrease the global unfolding rate constant. This decrease is indeed apparent in our simulation results (Figure 7) for experimental conditions that favor the fully native species ($K_{eq} < 1$). However, under denaturing conditions ($K_{eq} > 1$), addition of domains has no significant effect on the global unfolding rate constant because the term $K_{eq}^{n-1}/\sum_{j=1}^{n} K_{eq}^{j-1}$ in eq 23 approaches 1.

This counterintuitive result can be best understood by considering the flux through the reactions depicted in Scheme 2. The expected relative populations of the various forms of a five-B domain repeat protein under physiological conditions are listed under each column of intermediates in Scheme 2, based on the stability of BdpA in the absence of denaturant (Table 1). The flux for each of the equilibria depicted in Scheme 2 can be calculated from these populations and the estimated rate constant for unfolding of BdpA under the same conditions (Table 2) using eq 20. The total fluxes for each column of reactions in Scheme 2, calculated from the individual fluxes according to eq 18, are listed in Scheme 2. It is clear from these values that under physiological conditions any single domain of the fully folded protein unfolds and refolds very frequently (20% every millisecond). However, intermediates with two unfolded domains form 1000 times less frequently. The flux to form three, four, or five unfolded domains is extremely small because the populations of the precursor intermediates are very small. Since, according to eq 19, the step with the smallest flux in a sequential pathway dominates the total flux through the pathway, the flux from any intermediate with a single folded domain to a fully unfolded five-domain protein is extremely small and, thus, so is the global unfolding rate constant (eq 23). Although their fractional population is small, there would nevertheless be 12 fully unfolded molecules of a five-B domain protein in a 2 mL CD sample (10 μM). These molecules would fold with a rate constant equal to that of the single B domain because each successive folding event creates an intermediate which is much more likely to fold another domain than unfold. Although there are five domains to be folded in these molecules, there are 120 different pathways by which they can do so, which makes the global rate constant for folding the same as that for a single domain. This scenario is embodied in eq 21, which gives the global folding rate constant as the total flux divided by $[D_n]$, whose numerical value is much smaller than the flux for the five-domain protein under physiological conditions. A similar scenario can be envisioned to explain the global rate constants for folding and unfolding under highly denaturing conditions.

Scheme 2 highlights the intrinsic multiple-pathway nature of MIFD protein folding. Because each domain folds independently, we can be assured that folding of multiple domains can proceed in any order, biased by the relative stabilities and folding kinetics of individual domains. A multiple-pathway mechanism is also envisioned by the diffusion−collision theory for folding of microdomains within a single-domain protein (*39*). Although a diffusion−collision mechanism is likely to apply to many single-domain protein folding reactions (*25*, *40*, *41*), it is difficult to experimentally verify because microdomains are difficult to identify and study in isolation. However, for MIFD proteins whose individual domains can be studied individually, the predictions of the diffusion−collision model can be easily tested. Because of their independence, interdomain collisions do not play an important mechanistic role for MIFD proteins. However, for non-independently folding multidomain proteins, diffusion−collision theory may be a very useful tool in describing the overall folding mechanism, particularly when individual domains or pairs of domains can be studied in isolation.

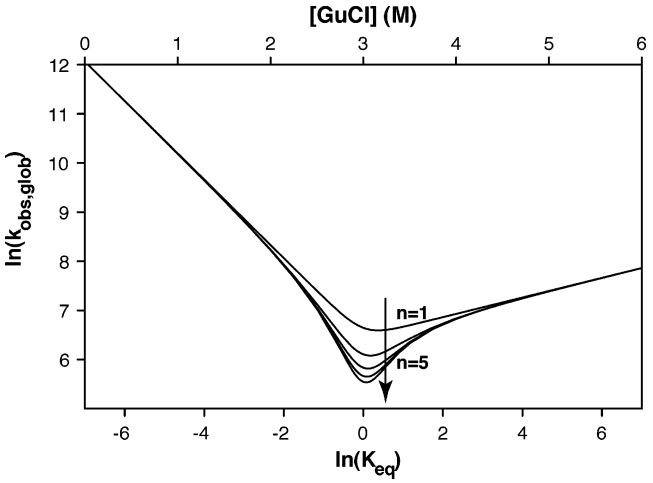


FIGURE 8: Predicted chevron plots for proteins with multiple (n = 1, 2, 3, 4, and 5) identical, uncoupled domains. The plots show significant perturbation with a change in n only under conditions where $K_{eq} \approx 1$. The top x-axis is calculated as described in the legend of Figure 6.

Since it is the sum of folding and unfolding rate constants ($k_{obs,glob} = k_{f,glob} + k_{u,glob}$) that is measured experimentally to determine the kinetics of protein folding, we also calculated the effect of addition of domain(s) on this quantity. A plot of the dependence of $\ln(k_{obs,glob})$ on denaturant concentration [or $\ln(K_{eq})$] is termed a chevron plot (*42*). As discussed in the previous paragraph, the global folding or unfolding rate constants of a MIFD protein with identical domains differ from those of the single-domain protein only under conditions where they contribute insignificantly to k_{obs}. Therefore, the chevron plots for proteins with different numbers of identical domains would deviate from each other only under experimental conditions where both the global folding and unfolding rate constants contribute significantly to the observed rate constant ($K_{eq} \approx 1$), where $k_{obs} = 2/n$ times the rate constant for folding or unfolding of the single domain (Figure 8). To experimentally observe such a chevron plot, it would be necessary to design a probe that detects the overall folding of the MIFD protein, rather than a probe such as Trp13, which is found in both domains of BBdpA. A fluorescence resonance energy transfer (FRET) pair located on the N- and C-terminal domains with an R_o matched to the distance between the domain termini in the fully native form might represent such a probe.

Biological Implications of the Uncoupled (Un)folding of the Antibody Binding Domains of Protein A. One of the

bottlenecks in the efficient translocation of proteins across the cell wall of Gram-positive bacteria is the protein unfolding (*43*, *44*). In *E. coli*, the unfolded, translocation-competent state of a newly synthesized protein to be secreted is maintained by the cytosolic secretion chaperone SecB (*45*). However, no SecB homologues have been found so far in *S. aureus*, raising the possibility that protein unfolding is a prerequisite for the secretion of SpA (*45*). The expected rate constant for the secretion of SpA (~60 s^{-1}) (*46*) is comparable to the unfolding rate constant of WT BdpA (~70 s^{-1}) (*25*). This rapid unfolding rate may have been optimized by evolution to produce maximum secretion efficiency. The even more rapid folding rate may be the consequence of evolutionary pressure to produce a stable protein with such a rapid unfolding rate. Coupling between domains would increase the size of the cooperative unfolding unit and thereby slow unfolding. The high degree of sequence similarity of the five domains of SpA suggests that they have stabilities and kinetics similar to those of the B domain and that they also fold independently. The calculations described above show that the global unfolding of this five-domain sequence would be ~12 orders of magnitude slower than that of a single domain. This decelerated unfolding rate constant would significantly impede the secretion efficiency if the simultaneous unfolding of all five domains is required for protein A to be secreted. On this basis, we suggest that only one domain is unfolded at a time during secretion, while the other domains, both the previously secreted N-terminal domains and the unsecreted C-terminal domains, are folded. Given the lack of coupling between domains, it is likely that this mechanism would hold even if each domain is actively unfolded, as observed in other protein translocation systems (*47*).

The uncoupled unfolding of the five domains of SpA could also be biologically advantageous for the prevention of protein aggregation when two or more of these domains are unfolded. A previous study has shown that the efficiency of co-aggregation between different domains increases markedly with an increase in the level of sequence identity (*48*). Figure 2 shows that the level of sequence identity between any two adjacent domain pairs of SpA is at least 75%, which is significantly higher than the proposed upper limit (~40%) that can be tolerated without significant aggregation (*48*). Therefore, SpA is expected to be considerably vulnerable to aggregation when two or more of its domains are unfolded. Despite this prediction, our in vitro studies showed no increased tendency of BBdpA to aggregate at concentrations as high as 1.5 mM. The calculations described in the previous section show that under cellular conditions the kinetics associated with unfolding of multiple domains in SpA are orders of magnitude slower than that for a single domain. Therefore, it seems likely that the very low probability of multiple domains in SpA being unfolded simultaneously in vivo could avoid aggregation that is initiated via interactions between a pair of unfolded domains. Moreover, the fast folding kinetics of BdpA (and SpA) under native cellular conditions could also facilitate the kinetic partitioning of protein molecules toward the monomeric folded state rather than the aggregate product. The fast, independent folding of five domains of SpA also allows the correct folding of each domain subsequent to its synthesis on ribosome, before the synthesis of downstream domains is complete. The formation of native structure in each domain would prevent the formation of non-native contacts that are likely to form in a multidomain protein with coupled domains. In these coupled multidomain proteins, the initially synthesized parts of the polypeptide chain, corresponding to one or more constituent domains, remain unfolded and await the completion of full-length protein synthesis before they can fold to the correct native structure. The fast, uncoupled folding of SpA domains compared to folding of the other two-domain proteins indicates that the protein is evolutionary optimized for rapid, cotranslational folding, as previously observed for other protein systems (*49*).

We have shown experimentally that two SpA B domains used in the same manner as two neighboring domains in the full-length protein have thermodynamically and kinetically uncoupled folding reactions. It would be interesting to try to introduce coupling between the domains of protein A by amino acid substitution or by addition of cell permeable small molecules and determine the effect on the efficiency of the secretion process of protein A. This approach might allow the development of antibiotics to reduce *S. aureus* virulence. Protein A has been demonstrated to cause inflammatory response in pulmonary epithelial cells through stimulation of the TNFα receptor (*50*).

In the calculations described here, we demonstrate the utility of flux equations in deriving rate laws for complex systems of multiple equilibria. This approach should be useful in analyzing and interpreting the experimental behavior of such proteins, which likely represent a large fraction of the eukaryotic proteome.

APPENDIX

The $(n - j)!j!$ statistical factor is used when adding the concentrations of fully folded (N_n) species and partially folded intermediates (eq 12) in the conservation of mass equation:

$$[P]_{total} = [N_n] + \sum_{i=1}^{n!}\sum_{j=1}^{n}\frac{[I_{i,j}]}{(n-j)!j!} = [N_n] + \sum_{i=1}^{n!}\sum_{j=1}^{n}\frac{[N_n]\prod_{k=1}^{j}K_{eq,j,k}}{(n-j)!j!} = [N_n]\left[1 + \sum_{i=1}^{n!}\sum_{j=1}^{n}\frac{\prod_{k=1}^{j}K_{eq,j,k}}{(n-j)!j!}\right] \quad \text{(A1)}$$

Since the fluxes through all i (=1 to $n!$) unfolding pathways for a MIFD protein with identical domains are equal, substituting eq 19 into eq 18 yields

$$F_{total} = \frac{1}{n!}\sum_{i=1}^{n!}\left(\sum_{j=1}^{n}\frac{1}{F_j}\right)^{-1} \quad \text{(A2)}$$

F_j in eq A2 can be determined using the definition of flux (eq 20) and substituting $k_{u,i,j}$ with k_u since all the constituent domains of a MIFD protein with identical domains unfold with the same rate constant:

$$F_{total} = \frac{1}{n!}\sum_{i=1}^{n!}\left(\sum_{j=1}^{n}\frac{1}{k_u[I_{i,j-1}]}\right)^{-1} \quad \text{(A3)}$$

Substituting the definition of $[I_{i,j-1}]$ (eq 16) and evaluating the first summation:

$$F_{\text{total}} = \frac{1}{n!}\sum_{i=1}^{n!}\left(\sum_{j=1}^{n}\frac{1}{k_u[N_n]K_{eq}^{j-1}}\right)^{-1} = \frac{k_u[N_n]}{\sum_{j=1}^{n}\frac{1}{K_{eq}^{j-1}}} = \frac{k_u[N_n]K_{eq}^{n-1}}{\sum_{j=1}^{n}K_{eq}^{j-1}} \quad \text{(A4)}$$

ACKNOWLEDGMENT

We thank Ace Hatch for assistance in the collection of NMR data and Harvey Sage for collection of analytical ultracentrifugation data and Jane Clark for helpful discussion.

SUPPORTING INFORMATION AVAILABLE

Representative example of stopped-flow fluorescence data for BdpA and BBdpA (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding Des. 3*, R81−91.
2. Scott, K. A., Steward, A., Fowler, S. B., and Clarke, J. (2002) Titin: A multidomain protein that behaves as the sum of its parts, *J. Mol. Biol. 315*, 819−29.
3. Batey, S., Randles, L. G., Steward, A., and Clarke, J. (2005) Cooperative folding in a multi-domain protein, *J. Mol. Biol. 349*, 1045−59.
4. Vermeer, A. W., and Norde, W. (2000) The thermal stability of immunoglobulin: Unfolding and aggregation of a multi-domain protein, *Biophys. J. 78*, 394−404.
5. Brandts, J. F., Hu, C. Q., Lin, L. N., and Mos, M. T. (1989) A simple model for proteins with interacting domains. Applications to scanning calorimetry data, *Biochemistry 28*, 8588−96.
6. Zhou, Z., Feng, H., Zhou, H., Zhou, Y., and Bai, Y. (2005) Design and folding of a multidomain protein, *Biochemistry 44*, 12107−12.
7. Privalov, P. L. (1982) Stability of proteins. Proteins which do not present a single cooperative system, *Adv. Protein Chem. 35*, 1−104.
8. Jaenicke, R. (1999) Stability and folding of domain proteins, *Prog. Biophys. Mol. Biol. 71*, 155−241.
9. Jaenicke, R. (1987) Folding and association of proteins, *Prog. Biophys. Mol. Biol. 49*, 117−237.
10. Teichmann, S. A., Chothia, C., and Gerstein, M. (1999) Advances in structural genomics, *Curr. Opin. Struct. Biol. 9*, 390−9.
11. Wenk, M., Herbst, R., Hoeger, D., Kretschmar, M., Lubsen, N. H., and Jaenicke, R. (2000) Gamma S-crystallin of bovine and human eye lens: Solution structure, stability and folding of the intact two-domain protein and its separate domains, *Biophys. Chem. 86*, 95−108.
12. Martin, A., and Schmid, F. X. (2003) The folding mechanism of a two-domain protein: Folding kinetics and domain docking of the gene-3 protein of phage fd, *J. Mol. Biol. 329*, 599−610.
13. Batey, S., Scott, K. A., and Clarke, J. (2006) Complex folding kinetics of a multidomain protein, *Biophys. J. 90*, 2120−30.
14. Mello, C. C., Bradley, C. M., Tripp, K. W., and Barrick, D. (2005) Experimental characterization of the folding kinetics of the notch ankyrin domain, *J. Mol. Biol. 352*, 266−81.
15. Main, E. R., Lowe, A. R., Mochrie, S. G., Jackson, S. E., and Regan, L. (2005) A recurring theme in protein engineering the design, stability and folding of repeat proteins, *Curr. Opin. Struct. Biol. 15*, 464−71.
16. Main, E. R., Stott, K., Jackson, S. E., and Regan, L. (2005) Local and long-range stability in tandemly arrayed tetratricopeptide repeats, *Proc. Natl. Acad. Sci. U.S.A. 102*, 5721−6.
17. Mello, C. C., and Barrick, D. (2004) An experimentally determined protein folding energy landscape, *Proc. Natl. Acad. Sci. U.S.A. 101*, 14102−7.
18. Tashiro, M., Tejero, R., Zimmerman, D. E., Celda, B., Nilsson, B., and Montelione, G. T. (1997) High-resolution solution NMR structure of the Z domain of *staphylococcal* protein A, *J. Mol. Biol. 272*, 573−90.
19. Mayr, E. M., Jaenicke, R., and Glockshuber, R. (1997) The domains in γB-crystallin: Identical fold-different stabilities, *J. Mol. Biol. 269*, 260−9.
20. Rudolph, R., Siebendritt, R., Nesslauer, G., Sharma, A. K., and Jaenicke, R. (1990) Folding of an all-β protein: Independent domain folding in γII-crystallin from calf eye lens, *Proc. Natl. Acad. Sci. U.S.A. 87*, 4625−9.
21. Parker, M. J., Spencer, J., Jackson, G. S., Burston, S. G., Hosszu, L. L., Craven, C. J., Waltho, J. P., and Clarke, A. R. (1996) Domain behavior during the folding of a thermostable phosphoglycerate kinase, *Biochemistry 35*, 15740−52.
22. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−50.
23. Vu, D. M., Peterson, E. S., and Dyer, R. B. (2004) Experimental resolution of early steps in protein folding: Testing molecular dynamics simulations, *J. Am. Chem. Soc. 126*, 6546−7.
24. Sato, S., Religa, T. L., Daggett, V., and Fersht, A. R. (2004) Testing protein-folding simulations by experiment: B domain of protein A, *Proc. Natl. Acad. Sci. U.S.A. 101*, 6952−6.
25. Myers, J. K., and Oas, T. G. (2001) Preorganized secondary structure as an important determinant of fast protein folding, *Nat. Struct. Biol. 8*, 552−8.
26. Dimitriadis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G., and Smith, A. D. (2004) Microsecond folding dynamics of the F13W G29A mutant of the B domain of *staphylococcal* protein A by laser-induced temperature jump, *Proc. Natl. Acad. Sci. U.S.A. 101*, 3809−14.
27. Arora, P., Oas, T. G., and Myers, J. K. (2004) Fast and faster: A designed variant of the B-domain of protein A folds in 3 μsec, *Protein Sci. 13*, 847−53.
28. Dill, K. A., and Bromberg, S. (2002) *Molecular Driving Forces*, Garland Science, New York.
29. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol. 185*, 60−89.
30. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry 6*, 1948−54.
31. Pace, N. C., and Scholtz, J. M. (1997) *Protein structure: A practical approach*, IRL Press, Oxford, U.K.
32. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods Enzymol. 131*, 266−80.
33. Huang, G. S., and Oas, T. G. (1995) Submillisecond folding of monomeric lambda repressor, *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−82.
34. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) Microsecond protein folding through a compact transition state, *J. Mol. Biol. 263*, 311−22.
35. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation, *Adv. Protein Chem. 24*, 1−95.
36. Karimi, A., Matsumura, M., Wright, P. E., and Dyson, H. J. (1999) Characterization of monomeric and dimeric B domain of Staphylococcal protein A, *J. Pept. Res. 54*, 344−52.
37. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−48.
38. Vasilkoski, Z., and Weaver, D. L. (2004) Diffusion-collision model algorithms for protein folding kinetics, *J. Comput. Chem. 25*, 1101−7.
39. Karplus, M., and Weaver, D. L. (1976) Protein-folding dynamics, *Nature 260*, 404−6.
40. Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D., and Weaver, D. L. (1988) Diffusion-collision model for the folding kinetics of myoglobin, *Proteins 4*, 211−27.
41. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) Protein folding dynamics: Quantitative comparison between theory and experiment, *Biochemistry 37*, 5337−43.
42. Matthews, C. R. (1987) Effect of point mutations on the folding of globular proteins, *Methods Enzymol. 154*, 498−511.
43. Wickner, W., and Schekman, R. (2005) Protein translocation across biological membranes, *Science 310*, 1452−6.

44. Bolhuis, A., Tjalsma, H., Smith, H. E., de Jong, A., Meima, R., Venema, G., Bron, S., and van Dijl, J. M. (1999) Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*, *Appl. Environ. Microbiol. 65*, 2934−41.
45. van Wely, K. H., Swaving, J., Freudl, R., and Driessen, A. J. (2001) Translocation of proteins across the cell envelope of Gram-positive bacteria, *FEMS Microbiol. Rev. 25*, 437−54.
46. Goder, V., Crottet, P., and Spiess, M. (2000) In vivo kinetics of protein targeting to the endoplasmic reticulum determined by site-specific phosphorylation, *EMBO J. 19*, 6704−12.
47. Matouschek, A. (2003) Protein unfolding: An important process in vivo? *Curr. Opin. Struct. Biol. 13*, 98−109.
48. Wright, C. F., Teichmann, S. A., Clarke, J., and Dobson, C. M. (2005) The importance of sequence diversity in the aggregation and evolution of proteins, *Nature 438*, 878−81.
49. Sanchez, I. E., Morillas, M., Zobeley, E., Kiefhaber, T., and Glockshuber, R. (2004) Fast folding of the two-domain semliki forest virus capsid protein explains co-translational proteolytic activity, *J. Mol. Biol. 338*, 159−67.
50. Gomez, M. I., Lee, A., Reddy, B., Muir, A., Soong, G., Pitt, A., Cheung, A., and Prince, A. (2004) *Staphylococcus aureus* protein A induces airway epithelial inflammatory responses by activating TNFR1, *Nat. Med. 10*, 842−8.

BI060923S