Folding pathways of proteins with increasing degree of sequence identities but different structure and function

Rajanish Giri^{a,1}, Angela Morrone^{a,1}, Carlo Travaglini-Allocatelli^a, Per Jemth^b, Maurizio Brunori^{a,2}, and Stefano Gianni^{a,2}

^aIstituto Pasteur-Fondazione Cenci Bolognietti, Istituto di Biologia e Patologia Molecolari del Consiglio Nazionale delle Ricerche, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Università di Roma "La Sapienza," Piazzale A. Moro 5, 00185 Rome, Italy; and ^bDepartment of Medical Biochemistry and Microbiology, Uppsala University, Biomedical Center Box 582, SE-75123 Uppsala, Sweden

Edited by William A. Eaton, National Institutes of Health-NIDDK, Bethesda, MD, and approved April 23, 2012 (received for review February 22, 2012)

Much experimental work has been devoted in comparing the folding behavior of proteins sharing the same fold but different sequence. The recent design of proteins displaying very high sequence identities but different 3D structure allows the unique opportunity to address the protein-folding problem from a complementary perspective. Here we explored by Φ -value analysis the pathways of folding of three different heteromorph pairs, displaying increasingly high-sequence identity (namely, 30%, 77%, and 88%), but different structures called G_A (a 3- α helix fold) and G_B (an α/β fold). The analysis, based on 132 site-directed mutants, is fully consistent with the idea that protein topology is committed very early along the pathway of folding. Furthermore, data reveals that when folding approaches a perfect two-state scenario, as in the case of the G_A domains, the structural features of the transition state appear very robust to changes in sequence composition. On the other hand, when folding is more complex and multistate, as for the G_B s, there are alternative nuclei or accessible pathways that can be alternatively stabilized by altering the primary structure. The implications of our results in the light of previous work on the folding of different members belonging to the same protein family are discussed.

kinetics | protein engineering | protein folding | protein G

The ultimate goal of a biophysical study is to extract general rules from the analysis of simple systems, a task that is particularly challenging in the case of protein folding. In fact, when considering different globular proteins, complexity stems by and large from the difference in sequence but also the multiplicity of structures of the native and denatured states. A suitable strategy to tackle the problem is to study proteins that differ in sequence but share the same overall fold, i.e., members of the same protein family (1–8). Over the past two decades, this approach allowed drawing some general conclusions about the correlation between 3D structure and sequence composition. In particular, it was reported that the mechanism of folding is, generally, conserved for members of the same protein family (9, 10) supporting the idea that native topology is often a main factor in controlling folding pathway and speed (11).

A sophisticated protein engineering approach allowed Bryan and co-workers to obtain pairs of proteins with an increasing degree of sequence identity (up to the extraordinary value of 95%), but different 3D structure and function (12, 13). The primary structure of two domains from the streptococcal protein G, sharing 16% sequence identity, was subjected to extensive site-directed mutagenesis cycles, leading to the synthesis of pairs of variants with an increasing level of sequence identity (30%, 77%, 88%, and 95%, respectively) (12, 13). The two wild-type protein domains are called G_A , displaying a three-helix bundle fold, and G_B , displaying a $\alpha + \beta$ ubiquitin-like fold. Therefore, the different variants (Fig. 1) were identified as G_{A30} , G_{A77} , G_{A88} , and G_{A95} (for the G_A fold), and G_{B30} , G_{B77} , G_{B88} , and G_{B95} (for the G_B fold), depending on their relative degree of sequence

identities. These five pairs represent a paradigmatic experimental model system to address the folding problem from an original perspective; how can similar sequences lead to very different folds?

We have recently analyzed the folding mechanisms of G_{A88} and G_{B88} at a variety of different conditions by experiments and simulations (14). We observed that despite the high level of identity of the primary structures of these two proteins (49 out of 56 residues), their folding pathways appear to diverge as early as in the denatured state, which in G_{B88} but not in G_{A88} displays a detectable residual structure. This surprising finding prompted us to carry out a systematic analysis of each of the heteromorph variants designed by Bryan and co-workers (12). The natural wild-type G_A domain contains no intrinsic fluorescent probe (i.e., no Trp) and differs by only three amino acids from the lower tier G_{A30} . Furthermore, we could not carry out a complete folding characterization of the variants G_{A95} and G_{B95} , which differ only in three positions (15), because of their low thermodynamic stability. Therefore, we have undertaken a detailed analysis of the folding pathways of the three remaining heteromorph pairs, which were subjected to an experimental investigation using the Φ -value analysis (16). This procedure, introduced and validated by Fersht et al. (ref. 17 and references therein), infers structural information on folding transition state(s) by comparing the kinetics and thermodynamics of folding of a given protein with those of a series of conservative mutants; in this work, 132 mutants were fully characterized.

The results show that when folding approaches a perfect two-state scenario, as in the case of G_A , the structural features of the transition state of G_{A30} , G_{A77} , and G_{A88} appear very robust to changes in sequence composition and are therefore similar. On the other hand, when folding is more complex, as in the case of the three-state folder G_B , there are multiple and alternative nuclei or accessible pathways that can be selectively stabilized by altering the primary structure. The implications of our results in the light of previous work on the folding of members of the same protein families are discussed and highlight the crucial physico-chemical features which bias the early commitment to the $\alpha + \beta$ fold of the G_B family.

Results

To test how sequence composition versus topology dictates the folding of proteins, we performed an extensive Φ -value analysis

Author contributions: M.B. and S.G. designed research; R.G. and A.M. performed research; R.G., A.M., C.T.-A., P.J., M.B., and S.G. analyzed data; and M.B. and S.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹R.G. and A.M. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: maurizio.brunori@uniroma1.it or stefano.gianni@uniroma1.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201794109/-DCSupplemental.

Fig. 1. Structures and sequence alignments of the different G_A and G_B variants. All engineered G_A and G_B proteins display structure and function similar to their respective natural wild-type domains G_A and G_B . For each protein, amino acid identities are shown in blue and nonidentities in gray. A complete folding characterization of the variants shown inside the red rectangle was performed.

on six different proteins (Fig. 1), three variants of the G_A domain of staphylococcal protein G (i.e., G_A30 , G_A77 , and G_A88), and three variants of the G_B domain (i.e., G_B30 , G_B77 , and G_B88). A total of 132 mutants were produced, purified, and characterized by equilibrium and kinetic folding experiments.

Performing a complete Φ -value analysis demands a careful selection of the experimental conditions. In fact, the protein of interest must be stable enough to allow accurate determination of the folding kinetics for its destabilized mutants, but it must not be too stable, otherwise its unfolding kinetics may be difficult to evaluate with the necessary accuracy. As detailed below, since the three pairs of proteins characterized in this work and their site-directed variants displayed widely different thermodynamic stabilities as well as different sensitivities to changes in ionic strength, we optimized the experimental conditions for each protein with regard to temperature and type of denaturant (i.e., urea or GdnHCl). In addition, in an effort to test the robustness of the folding pathway, a limited set of Φ values was measured for each protein at more than one experimental condition, while the full set of Φ values was obtained for each system at pH 7.2 (see *SI Text*).

G_A Proteins. Urea induced equilibrium transitions of G_A30 , G_A77 , and G_A88 were measured at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer. A typical equilibrium denaturation profile for each protein is reported in Fig. 2 *A–C*). While in the case of G_A77 and G_A88 we could observe monotonic sigmoidal transitions, G_A30 was found to be too stable and was not fully denatured even at very high urea concentrations; therefore, in the case of G_A30 we performed (un)folding experiments using GdnHCl. Because of the ionic nature of GdnHCl, we could not perform (un)folding experiments with this denaturant on G_A77 and G_A88 , since their stabilities display a pronounced dependence on ionic strength, as shown by folding experiments in the presence of sodium chloride.

We extensively studied the folding and unfolding kinetics of G_A30 , G_A77 and G_A88 by stopped-flow and temperature jump (T-jump) experiments. In the case of G_A77 and G_A88 , it was not possible to measure reliable folding and unfolding rate constants

Fig. 2. Equilibrium denaturation of the G_A (*A–C*) and G_B (*D–F*) proteins monitored by CD in 50 mM sodium phosphate buffer at pH 7.2. As described in the text, because of the different thermodynamic stabilities of the proteins, the equilibrium denaturations were performed at different experimental conditions: the unfolding experiment of G_A30 was performed at 25 °C using GdnHCl; G_A77 and G_A88 experiments were performed at 10 °C using urea; the unfolding denaturations of G_B30 and G_B77 were carried out at 25 °C using GdnHCl; and G_B88 experiment was performed at 25 °C using urea and 0.4 M sulfate. For all proteins, the pH was held at the constant value of 7.2 in the presence of 50 mM sodium phosphate.

at 25 °C over a wide range of denaturant concentration because reactions were too fast for our stopped-flow apparatus; thus, kinetic folding data for the two proteins were recorded at 10 °C. In all cases, the folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration. Semilogarithmic plots of the observed folding/unfolding rate constants of G_A30 , G_A77 , and G_A88 versus denaturant concentration (i.e., chevron plots) are presented in Fig. 3 *A–C*. All proteins displayed a V-shaped chevron plot, a hallmark of two-state folding. Two-state folding was further confirmed by the excellent agreement between the thermodynamic parameters obtained by equilibrium and kinetic data. These data parallel earlier studies on G_A88 (14) and confirm that this protein system folds via a two-state folding pathway with an unstructured denatured state.

We addressed the structural features of the transition state for folding of the G_A proteins by Φ -value analysis. A total of 50 mutants was produced: 14 for G_A30 , 17 for G_A77 , and 19 for G_A88 . Three mutants expressed poorly or were too unstable to be included in the analysis. The remaining 47 were subjected to equilibrium and kinetic folding experiments (Figs. S1–S3). In some cases, the folding and unfolding rate constants were too fast for the stopped-flow methodology and were determined using a capacitor-discharge T-jump apparatus. Fitted parameters are listed in Table S1.

Following a generally accepted convention (16, 18–20), the experimentally determined Φ -values were grouped in three different classes and mapped on the native structure of the G_A protein (Fig. 4): small values ($\Phi < 0.3$; red), intermediate values ($0.3 < \Phi < 0.7$; magenta), and large values ($\Phi > 0.7$; blue). Inspection

Fig. 3. Chevron plots of G_A and G_B variants. (A–C) panels: semilogarithmic plots of the observed rate constant for folding and unfolding of G_A30 , G_A77 , and G_A88 versus denaturant concentration measured at pH 7.2 and 10 °C. (D–F) panels: Semilogarithmic plots of the observed rate constant for folding and unfolding of G_B30 , G_B77 , and G_B88 versus denaturant concentration measured at pH 7.2 and 25 °C. The lines are the best fit to a two- or a three-state model.

of Fig. 4 clearly reveals that the folding nucleus is by and large conserved in all the G_A proteins, with the highest Φ values clustered at the interface between helix 1 and helix 2, suggesting that, in the case of such a two-state system, the dominant folding mechanism is very robust to perturbations of the primary structure.

G_B Proteins. The folding pathway of the G_B proteins is inherently more complex than that of the G_A partners. In fact, we have previously shown (21) that the wild-type G_B protein, commonly referred as GB1, a widely used model system for protein-folding studies, is characterized by the presence of an on-pathway intermediate, as indicated by a curvature in the unfolding arm of its chevron plot. Such a curvature, detected only at high concentrations of GdnHCl, allows addressing experimentally both the early and late folding events. Consequently, we resorted to study the folding of G_B30 , G_B77 , and G_B88 at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer, using GdnHCl as a denaturing agent. Unfortunately, however, in the case of G_B88 , many of its site-directed variants were poorly soluble at moderate concentrations of GdnHCl, and we could not obtain reliable folding data on this system with this denaturant. Furthermore, the low stability of this variant did not allow us to perform a complete Φ -value analysis in the absence of the stabilizing agent sodium sulfate. All folding studies of G_B88 were therefore carried out using urea as denaturant and in the presence of 0.4 M sodium sulfate.

The equilibrium unfolding transitions of G_B30 , G_B77 , and G_B88 are reported in Fig. 2 D–F. In all cases, we observed a monotonic sigmoidal transition, suggesting the absence of equilibrium intermediates.

The chevron plots for G_B30 , G_B77 , and G_B88 as a function of denaturant concentration, at 25 °C and pH 7.2, are reported in Fig. 3 D–F. It is evident that, in analogy to what previously

Fig. 4. Structural distribution of the measured Φ -values on the native structures of G_A30 , G_A77 , and G_A88 . The experimentally determined Φ values were divided into three categories and reported on the structure of G_A variants using the following color code: red, $0 < \Phi < 0.30$; magenta, $0.30 < \Phi < 0.70$; blue, $0.70 < \Phi < 1$. A conserved nucleus between the helices $\alpha 1$ and $\alpha 2$ (indicated by the red arrows) is clearly evident in all the G_A variants.

observed for GB1, both G_B30 and G_B77 display a pronounced curvature in their unfolding arms, as expected for a three-state folding mechanism (22, 23). In the case of G_B88 , while we could not detect such a curvature when performing the experiments in urea, data recorded in the presence of high concentrations of GdnHCl were similar to those of G_B30 and G_B77 (Fig. S4). This observation suggests that all the G_B variants appear to conform to a three state-folding mechanism.

We addressed the structural features of the early and late transition states for folding of the G_B proteins by Φ -value analysis. We produced a total of 82 mutants: 24 for G_B30 , 27 for G_B77 , and 31 for G_B88 . 16 variants expressed poorly or were too unstable to be included in the analysis. The remaining 66 were subjected to equilibrium and kinetic folding experiments (Figs. S5–S7). In the case of G_B30 and G_B77 , the data for each variant was globally fitted to a three-state equation with m -values assumed to be the same as those of the wild-type proteins. All data was in agreement with a three-state mechanism involving the presence of two folding transition states, a more denatured-like $TS1$ ($\beta = 0.77$ and $\beta = 0.69$ for G_B30 and G_B77 , respectively) and a native-like $TS2$ ($\beta = 0.97$ for both the proteins). On the other hand, in the case of G_B88 , because of the low solubility of its variants in the presence of GdnHCl, we could perform the experiments only in the presence of urea, obtaining structural information only on the early transition state $TS1$ ($\beta = 0.81$).

A graphic depiction of the structural distribution of the measured Φ -value for $TS1$ and $TS2$ of G_B30 and G_B77 and of $TS1$ of G_B88 is reported in Fig. 5. It is interesting to note that the distribution of the measured Φ -values for the first transition state (Φ_{TS1}), plotted on the native structure of G_B , show considerable differences among G_B30 , G_B77 , and G_B88 . In fact, looking at G_B30 and G_B88 , a shift of the medium-high Φ -values from the first β -hairpin to the second, with G_B77 displaying an intermediate behavior (Fig. 5), may be appreciated. This trend indicates that alternative folding nuclei, located at the hairpins between either $\beta 1$ – $\beta 2$ or $\beta 3$ – $\beta 4$, drive the folding to the G_B -topology. These nuclei may be selectively stabilized depending on amino acid composition. Remarkably, in the native-like transition state $TS2$, which we could infer only for G_B30 and G_B77 , both nuclei appear in the process of being folded, and the two transition states display a similar overall structure, indicating that the alternative folding pathways converge as the native state is approached. In order to compare the Φ -values obtained under different conditions, kinetic experiments on some mutants were performed with both guanidine and urea. The calculated Φ -values were approximately the same; the folding mechanism of the G_B proteins is therefore not affected by the nature of the denaturant agent used to obtain the data [Fig. S8 shows the structural distribution

Fig. 5. Structural distribution of the measured Φ -values for *TS1* (G_B30 , G_B77 , and G_B88) and for *TS2* (G_B30 and G_B77). The experimentally determined Φ values were divided into three categories and reported on the native structure of G_B variants using the same color code as that used for the G_A proteins. In the case of G_B30 and G_B88 , it is possible to observe a shift of the medium-high Φ -values from the first β -hairpin to the second (indicated by the red arrows), with G_B77 displaying an intermediate behavior.

of the measured Φ -values for *TS1* of G_B77 obtained in presence of GdnHCl (A) and urea (B)].

Discussion

The holy grail in protein-folding studies is to unveil the correlation between sequence information and reaction mechanisms. A classical approach to address this question has been to study proteins that differ in amino acid sequence but share the same fold (1–8). On the other hand, the design and production of proteins sharing high sequence identity, yet displaying a different structure and function, allows for the first time to approach the folding problem from a complementary perspective (12, 15, 22, 23). In this work, we have characterized the complete folding pathway of heteromorphic proteins originating from two domains of streptococcal protein G being either fully α -helical (called G_A) or largely β (called G_B). An extensive Φ -value analysis of G_A30 , G_A77 , G_A88 , and of G_B30 , G_B77 and G_B88 generates a “matrix” of proteins, which share the same topology (the two rows of Fig. 1) and display increasingly high sequence identities while keeping a different structure.

When does a protein commit to its native topology in its folding pathway? In a preliminary study, we compared the folding and unfolding kinetics of G_A88 and G_B88 at a variety of different experimental conditions (14). We observed that a detectable residual structure is present in the denatured state of G_B88 , while the denatured state of G_A88 is essentially unstructured. In the light of these observations, we hinted that the folding pathways of these two proteins diverge as early as in the denatured state. In the present work, by systematically describing the structural features of the folding transition states of G_A30 , G_A77 , G_A88 , as well as of G_B30 , G_B77 , and G_B88 , we have taken our comparative study on heteromorphic pairs to the next level of complexity addressing by Φ -value analysis the structure of the transition states. Remarkably, the structural distribution of the measured Φ -values in all these variants is reminiscent of what has been previously observed in many single domain proteins, with a weak nucleus displaying moderately high values of Φ (>0.7), and simultaneous formation of extensive native-like structure, which gradually tapers off from the nucleus to other regions of the protein (24, 25). Accordingly, the transition states of all the proteins considered here appear to resemble a distorted version of the corresponding native

states, with some polarization of structure at the N- or C-terminal β -hairpins in the case of G_B30 and G_B88 respectively. Furthermore, additional support for the presence of residual structure in the denatured state of G_B88 is represented by the presence of three nonstandard values of Φ (i.e., L20A, V21A, and Y33F; Table S2). Remarkably, these positions are located at the N- and C-terminal regions of the central helix of the G_B fold, which was observed by molecular dynamics simulation to be partially formed in the denatured state of G_B88 (14). Overall, despite the very high level of sequence identities, when comparing the folding of each heteromorphic pair, no common intermediate was detected, suggesting that they fold via completely independent paths. All these findings converge in indicating that proteins commit very early to their topology, and the structural features leading to their native state are, most likely, already imprinted in their denatured states.

A crude but reliable method to compare the folding pathways of these different proteins is to analyze the structural distribution of measured Φ -values. From this perspective, data indicates that, while in the case of G_A the mechanism appears rather robust to divergence in primary structure, in the case of G_B , folding is more malleable. In fact, when mapped onto the corresponding native structures (Fig. 4), the Φ -values reveal a conserved transition state among all the three G_A variants (G_A30 , G_A77 , G_A88); on the other hand, G_B30 and G_B88 clearly display a shift of the medium-high Φ -values from the first β -hairpin to the second, with G_B77 displaying an intermediate outlook (Fig. 5). This finding may be seen in the light of the proposal that the number of accessible pathways for folding is determined by the different nucleation motifs contained within a given native topology (26). For example, the structure of ribosomal protein S6 seems to be composed of two different nucleation patterns acting as independent cooperative units, each of which constitutes a separate entry to parallel folding trajectories (27). Accordingly, it may be suggested that in the case of G_B , the symmetrical organization of its 3D topology implies the presence of multiple nucleation motifs that permit alternative folding pathways.

For our purposes, it would be revealing to understand which structural determinants preclude the sequence of the G_A proteins to adopt the structure of G_B and vice versa. Recent molecular dynamics simulations on G_A88 and G_B88 suggested the long and stable helix in the central region of the sequence of G_B88 preventing the polypeptide chain to form the loop connecting helix 1 and helix 2 in G_A88 and thus to fold into a fully helical structure (14). This finding is further corroborated by the high helical propensity of the only α -helix of all the G_B proteins as predicted by AGADIR (28), when compared to the G_A counterparts. By following this view, it is interesting to note that structure selection in the G_A proteins seems to be initiated by the formation of the critical contacts between helix 1 and helix 2 (Fig. 4). In the case of the G_B proteins, on the other hand, the Φ -value analysis reported in this work, together with the previously published molecular dynamics simulations, suggests a scenario whereby folding is guided by the alternative docking of the N- or C-terminal hairpins on the central long helix, that might be (partially) preformed in the denatured state.

Although the protein-folding reaction involves the formation and breakage of a myriad of contacts, a typical feature of small single domain proteins is the ability to fold cooperatively (29–31). Many weak noncovalent bonds form simultaneously and, very often, only the fully native and fully denatured states may be experimentally detected. Yet, because cooperativity is never extreme, not all residues are equally important for folding, and one or more regions of the protein tend to display a selective propensity to fold independently (32, 33). These subdomains initiate the folding reaction and act as folding nuclei. Accordingly, when cooperativity decreases, the energetic coupling for folding of different regions of the same protein is also decreased, i.e., different regions of the protein may fold independently and folding

may become modular. In this perspective, it is of interest to compare the different robustness of the folding mechanisms for the G_A and G_B proteins. In fact, while G_A , a two-state system, appears to display a unique nucleus, in the case of G_B , where the tendency to populate folding intermediates is documented, alternative nuclei are present and folding is more sensitive to changes in sequence composition. Overall, our data reveals that pathway malleability is determined by the presence of multiple nuclei; the segregation of such nuclei results in stabilization of folding intermediates, whereas “perfect” two-state systems are characterized by a unique diffused nucleus and, therefore, by a robust folding pathway.

Materials and Methods

Site-Directed Mutagenesis and Protein Expression and Purification. G_A and G_B genes were cloned into the vector pG58 (provided by Philip N. Bryan, University of Maryland) which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein (34). These genes were used as templates to perform site-directed mutagenesis. All mutants were obtained by using the QuikChange mutagenesis kit (Stratagene), according to the manufacturer's instructions. All of the mutations were confirmed by DNA sequencing.

Proteins were expressed and purified as described previously (14).

Equilibrium Unfolding. Circular dichroism (CD) spectra were recorded between 250 and 200 nm using a Jasco spectropolarimeter (Jasco, Inc., Easton, MD, USA). Urea and guanidine denaturations were followed at 222 nm in a 1 cm path length quartz cuvette (Hellma, Plainview, NY, USA) at 10 °C or 25 °C. Protein concentration was typically 15 μ M. The buffer used was 50 mM sodium phosphate pH 7.2.

Stopped-Flow Measurements. Single mixing kinetic folding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instruments (Applied Photophysics). The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments,

performed at 25 °C and 10 °C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. The buffer used was 50 mM sodium phosphate pH 7.2. Final protein concentrations were typically 1 μ M. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μ M after mixing protein concentration), as expected from monomolecular reactions without effects due to transient aggregation (35).

T-Jump Fluorescence Spectroscopy. The relaxation kinetics were measured as a function of guanidine or urea by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech). Temperature was rapidly changed from 18 °C to 25 °C and from 4 °C to 10 °C with a jump size of 7 °C and 6 °C, respectively. 10 to 20 individual traces were averaged at given denaturant concentrations. Protein concentration was typically 20 μ M. The excitation wavelength was 280 nm, and the fluorescence emission was measured using a 320 nm cut-off glass filter.

Data Analysis. Equilibrium experiments. Data were fitted to a standard two-state denaturation. An equation that takes into account the pre- and post-transition baselines was used to fit the observed unfolding transition (36).

Kinetic experiments. Analysis was performed by nonlinear least squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted globally by numerical analysis based on a two- or a three-state model.

The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration. The observed chevron plots were fitted globally with shared m -values. The global fit was obtained with Prism software (Graphpad). Φ values were calculated from folding rate constants using standard equations (17).

ACKNOWLEDGMENTS. We thank Prof. Philip N. Bryan for inspiring this work. Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBRN07BMCT_007 to M.B.) Progetto di Interesse 'Invecchiamento' (S.G.), and from the Swedish Research Council (P.J.).

- Calosci N, et al. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc Natl Acad Sci USA* 105:19241–19246.
- Chi CN, et al. (2007) A conserved folding mechanism for PDZ domains. *FEBS Lett* 581:1109–1113.
- Clarke J, Cota E, Fowler SB, Hamill SJ (1999) Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. *Structure* 7:1145–1153.
- Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J Mol Biol* 36:293–305.
- Martínez JC, Serrano L (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 6:1010–1016.
- Riddle DS, et al. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 6:1016–1024.
- Travaglini-Allocatelli C, Gianni S, Brunori M (2004) A common folding mechanism in the cytochrome c family. *Trends Biochem Sci* 29:535–541.
- Travaglini-Allocatelli C, et al. (2003) Exploring the cytochrome c folding mechanism: Cytochrome c552 from *Thermus thermophilus* folds through an on-pathway intermediate. *J Biol Chem* 278:41136–41140.
- Travaglini-Allocatelli C, Ivarsson Y, Jemth P, Gianni S (2009) Folding and stability of globular proteins and implications for function. *Curr Opin Struct Biol* 19:3–7.
- Zarrine-Afsar A, Larson SM, Davidson AR (2005) The family feud: Do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* 15:42–49.
- Baker D (2000) A surprising simplicity to protein folding. *Nature* 405:39–42.
- Alexander PA, et al. (2007) The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc Natl Acad Sci USA* 104:11963–11968.
- He Y, Chen Y, Alexander P, Bryan PN, Orban J (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. *Proc Natl Acad Sci USA* 105:14412–14417.
- Morrone A, et al. (2011) The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J Biol Chem* 286:3863–3872.
- Alexander PA, et al. (2009) A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci USA* 106:21149–21154.
- Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782.
- Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 101:7976–7981.
- Geierhaas CD, Salvatella X, Clarke J, Vendruscolo M (2008) Characterisation of transition state structures for protein folding using 'high,' 'medium,' and 'low' (Φ)-values. *Protein Eng Des Sel* 21:215–222.
- Gianni S, et al. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* 17:1431–1437.
- Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Gianni S (2009) Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. *J Am Chem Soc* 131:11727–11733.
- Morrone A, et al. (2011) Gb1 is not a two-state folder: Identification and characterization of an on-pathway intermediate. *Biophys J* 101:1–8.
- Parker MJ, Spencer J, Clarke AR (1995) An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J Mol Biol* 253:771–786.
- Wildegger G, Kiefhaber T (1997) Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* 270:294–304.
- Abkevich VI, Gutin AM, Shakhnovich EI (1994) Specific nucleus as the transition state for protein folding: Evidence from the lattice model. *Biochemistry* 33:10026–10036.
- Itzhaki LS, Otzen DE, Fersht AR (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.
- Lindberg MO, Oliveberg M (2007) Malleability of protein folding pathways: A simple reason for complex behavior. *Curr Opin Struct Biol* 17:21–29.
- Haglund E, Lindberg MO, Oliveberg M (2008) Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J Biol Chem* 283:27904–27915.
- Munoz V, Serrano L (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41:495–509.
- Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195.
- Fersht AR (2008) From the first protein structures to our current knowledge of protein folding: Delights and scepticisms. *Nat Rev Mol Cell Biol* 9:650–654.
- Kubelka J, et al. (2006) Sub-microsecond protein folding. *J Mol Biol* 359:546–553.
- Fersht AR (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci USA* 92:10869–10873.
- Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of beta-hairpin formation. *Nature* 390:196–199.
- Ruan B, et al. (2004) Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. *Biochemistry* 43:14539–14546.
- Silow M, Oliveberg M (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc Natl Acad Sci USA* 94:6084–6086.
- Santorino MM, Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063–8068.