The Denatured State Dictates the Topology of Two Proteins with Almost Identical Sequence but Different Native Structure and Function

Received for publication, June 18, 2010, and in revised form, November 19, 2010. Published, JBC Papers in Press, November 29, 2010. DOI 10.1074/jbc.M110.155911

Angela Morrone†, Michelle E. McCully‡, Philip N. Bryan§, Maurizio Brunori‖, Valerie Daggett¶, Stefano Gianni** and Carlo Travaglini-Allocatelli††

From the †Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche “A. Rossi Fanelli,” Università di Roma “La Sapienza,” 500185 Rome, Italy, the ‡Biomolecular Structure and Design Program and Department of Bioengineering, University of Washington, Seattle, Washington 98195, and the §Institute for Bioscience and Biotechnology Research/Department of Bioengineering, University of Maryland, Rockville, Maryland 20850.

The protein folding problem is often studied by comparing the mechanisms of proteins sharing the same structure but different sequence. The recent design of the two proteins GA88 and GB88, displaying different structures and functions while sharing 88% sequence identity (49 out of 56 amino acids), allows the unique opportunity for a complementary approach. At which stage of its folding pathway does a protein commit to a given topology? Which residues are crucial in directing folding mechanisms to a given structure? By using a combination of biophysical and computational techniques, we have characterized the folding of both GA88 and GB88. We show that, contrary to expectation, GA88, characterized by a native α+β fold, displays in the denatured state a content of native-like helical structure greater than GB88, which is all-α in its native state. Both experiments and simulations indicate that such residual structure may be tuned by changing pH. Thus, despite the high sequence identity, the folding pathways for these two proteins appear to diverge as early as in the denatured state. Our results suggest a mechanism whereby protein topology is committed very early along the folding pathway, being imprinted in the residual structure of the denatured state.

Understanding the rules that govern the folding of proteins is one of the main unsolved problems in modern science (1). Current knowledge on the protein folding reaction has been extrapolated by comparing folding processes of different members of a given protein family. Generally, proteins with significant sequence similarity are expected to have a similar fold. In fact, analysis of the Protein Data Bank (PDB) reveals that a sequence similarity of 40% nearly always leads to a similar fold (12). This observation provoked Rose and Creamer (13) in 1994 to issue the “Paracelsus Challenge,” whereby the protein folding community was confronted with the task of designing two proteins that are at least 50% identical but possessed different folds. Amazingly, this goal was fully achieved in only 3 years, when Regan and co-workers (14) designed a sequence that, despite being 50% identical to a mostly β-sheet protein, folded into a four-helix bundle. Since then, others have achieved similarly impressive results (15). Recently, ambitious work by Bryan and co-workers (16, 17) led to the design of pairs of proteins with an extraordinarily high degree of sequence identity but different folds and different functions. In particular, the sequences of two domains from streptococcal protein G were subjected to an iterative design of heteromorphic pairs, leading the authors to produce two protein G variants, called Gα88, which is mostly α-helical (the 3-helix bundle protein A fold), and Gβ88, displaying the α+β protein G fold (Fig. 1). These two proteins share 88% sequence identity (49 out of 56 amino acids), yet they display two different structures and functions that are similar to the respective wild-type proteins. In parallel with studies on protein families, this protein engineering achievement offers the unique opportunity for a complementary study on protein folding mechanism addressing two key questions. 1) At which stage of its folding pathway does a pro-
tein commit to a given topology? 2) Which residues are crucial in directing folding to a given native structure? Here we present for the first time an extensive characterization of the folding mechanisms of GA88 and GB88 by experiment and molecular dynamics (MD) simulation. The results obtained under a variety of solvent conditions suggest the presence in the denatured state of GB88 of pH-sensitive residual structure, as indicated by a pH dependence of its $m_D$ value, which is not observed for GA88. The MD simulations are consistent with these findings, showing that for GB88 the non-polar solvent-accessible surface area decreases markedly at low pH. Interestingly, in analogy with earlier observations on a similar heteromorphic protein A/protein G pair sharing 59% sequence identity (i.e. A219 and G311) (18), the extent of native-like helical structure in the denatured state of GB88 (the protein G fold) is greater than that of denatured GA88, which is all-$\alpha$ in its native conformation.

Both our current and our earlier studies of this system suggest that protein topology is committed very early along the folding pathway being “imprinted” in the residual structure of the denatured state; this weak, loosely defined topology is sufficient to dictate the pathway of folding. The significance of these observations from the perspective of previous work on the folding of proteins with the same topology but very different sequences is discussed.

**EXPERIMENTAL PROCEDURES**

The buffers used were 50 mM sodium phosphate from pH 8.0 to 6.3, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0, and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. In an effort to check whether the buffer composition had any effect on the folding experiments, we performed additional control experiments in the presence of 50 mM Tris at pH 7.2 and 8.0, 50 mM Mes at pH 6.2 and 5.5, and 50 mM sodium phosphate (first protonation) at pH 2.0. Both the folding and the unfolding reactions were essentially independent from buffer composition. All reagents were analytical grade.

**Protein Expression and Purification**

Bryan and co-workers (16) cloned GA88 and GB88 genes into the vector pG58, which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein. Proteins were purified employing an affinity chromatography previously developed (19). Soluble cell extract of pro-domain fusion proteins was injected on a 5-ml Bio-Scale Mini Profinity eXact cartridge at 5 ml/min to allow binding and then washed with 10 column volumes of 100 mM NaPO$_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of 100 mM NaPO$_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped, and the column was incubated for 30 min to allow complete cleavage. After elution, the purified proteins were then dialyzed to remove potassium fluoride.

**Equilibrium Unfolding**

Equilibrium denaturations were followed on a JASCO circular dichroism (CD) spectropolarimeter (JASCO, Inc., Easton, MD), in a 1-cm quartz cuvette (Hellma). CD spectra of GA88 and GB88 were recorded between 250 and 200 nm. Protein concentrations were typically 6 $\mu$M.
The Denatured State Dictates the Topology of Two Proteins

Stopped-flow Measurements

Single mixing kinetic folding experiments were carried out on a Pi-star stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm, and the fluorescence emission was measured using a 320 nm cutoff glass filter. In all experiments, performed at 25 and 10 °C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM), as expected from monomolecular reactions without effects due to transient aggregation (20).

Data Analysis

Equilibrium experiments—Assuming a standard two-state model, the urea-induced denaturation transitions were fit to the equation

\[ \Delta G_d = m_D - m_N (D - D_{50}) \]  (Eq. 1)

where \( m_{D-N} \) is the slope of the transition (proportional to the increase in solvent-accessible surface area ongoing from the native to the denatured state) and \( D_{50} \) is the midpoint of the denaturation transition. An equation that takes into account the pre- and post-transition baselines was used to fit the observed unfolding transition (21).

Kinetic Experiments—Analysis was performed by non-linear least squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted, using the Kaleidagraph software package, by numerical analysis based on a two-state model following the equation

\[ k_{obs} = k_F + k_U \]  (Eq. 2)

where \( k_F \) and \( k_U \) represent the folding and unfolding rate constants, respectively. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (22).

Molecular Dynamics Simulations

We performed 11 all-atom, explicit solvent MD simulations for each protein, G₈88 and G₉88: 298 K (30 ns), 498 K neutral and low pH (3 × 50 ns, 2 × 5 ns), for a total of 700 ns (0.7 μs) of simulation time. The starting structures were taken from the published NMR ensembles (17) (G₈88, PDB id 2jws, Model 1; G₉88, PDB id 2jw2, Model 3 [298 K], Model 1 [498 K]). Low pH systems were created by protonating all aspartate and glutamate residues (neither protein contains a histidine). The simulations were all performed using our in-house MD software, in lucem molecular mechanics (ilmm), with the Levitt et al. (24) all-atom force field and the microcanonical ensemble (N, constant number of particles, system volume, and total energy). Non-bonded terms were treated with an 8 Å (298 K) force-shifted cutoff. The proteins were minimized and solvated with explicit F3C flexible waters (25) using our standard protocol (26). Briefly, the protein was treated in vacuo for 1000 steps of steepest descent minimization. Pre-equilibrated F3C water was added within 1.8 Å of the protein to a box extending at least 10–12 Å from the protein on all sides. The water was then minimized for 1000 steps and then subjected to 500 ps of dynamics with 2-fs time steps and an additional 500 steps of steepest descent minimization. Finally, the protein was minimized for 500 steps.

The denatured state ensemble was defined as the final 30 ns of the three longer (50 ns) 498 K simulations. The Ca root mean square deviation (r.m.s.d.) was calculated over all 56 residues as well as for just the core residues. The core was defined as the consecutive residues beginning with the N-terminal residue of the first secondary structure element as defined by He et al. (17) to the C terminus of the final element (G₈88, residues 9–51; G₉88, residues 1–55). The percentages of helix and solvent-accessible surface area (SASA) were calculated using our in-house implementations of the Dictionary of Protein Secondary Structure (DSSP) (27) and Lee and Richards (28) algorithms, respectively. The percentage of helix or β-structure was reported over the total 56 residues as the number of structured residues as defined by DSSP, and non-polar SASA was reported as the sum of the SASA for all non-polar residues (Ala, Val, Phe, Pro, Met, Ile, Leu, Trp, and Gly). When values were reported relative to the native state, the average value over the 30-ns 298 K simulation was used as the native value.

Multidimensional scaling of the all-against-all Ca r.m.s.d. matrix was performed using the R statistical package to assign conformational ensembles. The exit from the native-like cluster, or point of no return, in the three-dimensional projection of the multidimensional scaling was defined as the transition state (TS), and the preceding 5 ps was defined as the transition state ensemble, as described previously (29, 30). The average Ca r.m.s.d. to the starting structure was reported for each TS structure, as was the average pairwise Ca r.m.s.d. between all TS structures in the ensemble.

The fraction of time that residues were in contact was calculated for the native ensembles (all 30 ns of the native state), the TS ensembles (5 ps from each 498 K unfolding simulation), and the denatured state (the last 30 ns of the three long 498 K simulations). Two residues were considered in contact if they contained carbon atoms that were <5.4 Å apart or any other two non-hydrogen atoms that were <4.6 Å apart. Hydrogen bonds were measured in the denatured state (last 30 ns of the three 50-ns 498 K simulations) for specific residue pairs using the following criteria: 1) the distance between the donor hydrogen and acceptor atom was ≤2.6 Å; 2) the donor-hydrogen-acceptor angle was within 45° of linearity; and 3) the charges on the donor and acceptor atoms were <−0.3, and the charge on the hydrogen was >0.3.

RESULTS

Equilibrium Unfolding of G₈88 and G₉88—To study the folding mechanism of G₈88 and G₉88, we carried out both equilibrium and kinetic (un)folding experiments. The results of urea-induced equilibrium denaturation of G₈88 and G₉88 measured at 10 °C, pH 7.2 in 50 mM sodium phosphate buffer,

---

4 D. A. C. Beck, D. O. V. Alonso, and V. Daggett, in lucem molecular mechanics (ilmm), University of Washington, Seattle, WA.
The Denatured State Dictates the Topology of Two Proteins

FIGURE 2. Equilibrium denaturation of GA88 (A) and GB88 (B) monitored by CD in 50 mM sodium phosphate buffer at pH 7.2 and 10 °C. Lines are the best fit to a two state unfolding mechanism. mdeg, millidegrees.

as monitored by far-UV circular dichroism (CD) spectroscopy, are provided in Fig. 2 (representative spectra for native and denatured GA88 and GB88 are also provided in supplemental Fig. S1). The observed transitions follow a simple two-state behavior, suggesting the absence of stable equilibrium intermediates for both proteins and indicating that these designed variants are capable of cooperative (un)folding reactions. Furthermore, the reaction was fully reversible under all conditions explored.

At physiological pH and in the absence of denaturant, the unfolding free energy of GA88 derived from a two-state analysis and a global fit of 60 wavelengths (from 250 to 220 nm) is ΔG_{D→N} = 3.00 ± 0.18 kcal/mol, with a m_{D→N} value of 0.62 ± 0.04 kcal mol^{-1} M^{-1}. In the case of GB88, ΔG_{D→N} = 2.35 ± 0.30 kcal/mol, and the m_{D→N} value is 1.10 ± 0.08 kcal mol^{-1} M^{-1}. Considering that the GB88 construct contains ~10 structured residues more than GA88, both m_{D→N} values are consistent with those expected for proteins of this size, according to the BPPred database (31). Hence, the seeming difference in cooperativity, as reflected by the different m_{D→N} values, can be accounted for by the difference in the number of structured residues between the two proteins.

Folding and Unfolding Kinetics—We carried out extensive fluorescence kinetic experiments on both proteins under a variety of different experimental conditions. In particular, the folding and unfolding kinetics were investigated at several pH values, ranging from 10 to 2. In the case of GA88, it was not possible to measure reliable folding and unfolding rate constants at 25 °C over a wide range of denaturant concentrations because the rates were too fast for our stopped-flow apparatus. Thus, kinetic folding data for the two proteins were recorded at 10 °C to slow the process for GA88 and to obtain values under the same conditions for comparison in the case of GB88. In all cases, the folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration (representative folding and unfolding time courses are reported as supplemental Fig. S2).

Semi-logarithmic plots of the observed unfolding rate constants for GA88 and GB88 versus denaturant concentration (i.e. chevron plots) at pH 7.2 are presented in Fig. 3. Both proteins displayed a V-shaped chevron plot, a hallmark of two-state folding (22). In the case of GA88, there was excellent agreement between the thermodynamic parameters obtained by equilibrium and kinetic data; however, in the case of GB88, there was a minor deviation, the m_{D→N} value being 1.10 ± 0.08 kcal mol^{-1} M^{-1} from equilibrium experiments and 0.90 ± 0.05 kcal mol^{-1} M^{-1} from chevron plot analysis. As observed previously for other small single domain proteins (32–34), a significant deviation of m_{D→N} from equilibrium and kinetic data suggests that there are residual structure and/or changes in the exposure of non-polar residues in the denatured state of the protein. Because the small deviation observed for GB88 is at the limit of experimental detection, we performed additional experiments under various conditions as well as MD simulations to further investigate these options, as described below.

A powerful method to address the global properties of folding transition and denatured states is the analysis of chevron plots recorded under different experimental conditions or on various site-directed variants (35). In fact, because the m-values (slopes of the unfolding and refolding arms of the chevron plots) reflect the change in accessible surface area upon (un)folding, analysis of their dependence on reaction conditions may be of diagnostic value to identify transition state movements along the reaction coordinate, as well as denatured state collapse or residual structure. In this study, we compared the folding kinetics of G_A88 and G_B88 at various pH values, ranging from 10 to 2. Inspection of Fig. 4 reveals that although both the stability and the m-values of G_A88 are insensitive to pH and this protein is fully native even at pH 2.0
The Denatured State Dictates the Topology of Two Proteins

A semi-logarithmic plot of the observed rate constant for folding and unfolding of G88 versus [urea] measured at different pH values and 10 °C (pH 10.0, ○ pH 8.0, ● pH 7.2, □ pH 4.0), and (pH 3.0, △ pH 2.0). B, semi-logarithmic plot of the observed rate constant for folding and unfolding of G88 versus [urea] measured at different pH values and 10 °C (○ pH 8.0, ● pH 7.2, ○ pH 6.2, X pH 5.5, □ pH 4.7, △ pH 4.3, □ pH 4.0). The lines are the best fit to a two-state model.

(supplemental Fig. S3), G88 is destabilized at pH < 5. However, the low stability of the latter protein and the poor definition of the observed refolding arms prevented a quantitative analysis of the m-values. Therefore as detailed below, we resorted to investigating the folding of G88 under stabilizing conditions.

Certain inorganic salts, such as phosphates and sulfates, favor compact protein conformations because of the preferential exclusion of solvent from the protein surface (36); this makes them potent stabilizers of both the native and the partially folded states. The chevron plots of G88 measured at different pH values in the presence of 0.4 M sodium sulfate at 25 °C (● pH 8.0, △ pH 7.2, ○ pH 6.7, □ pH 6.2, + pH 5.9, ◄ pH 5.5, X pH 5.0, △ pH 4.7, ○ pH 4.3, ● pH 4.0); lines are the best fit to a two-state model. B, dependence of mF (●), mU (●), and mD−U (●) on pH; lines are the best fit to an equation implying a single protonation site with pKa ∼ 5.

FIGURE 4. Chevron plots of GA88 and GB88 at different pH values. A, semi-logarithmic plot of the observed rate constant for folding and unfolding of GA88 versus [urea] measured at different pH values and 10 °C (∇ pH 10.0, ○ pH 8.0, ● pH 7.2, □ pH 4.0, + pH 3.0, △ pH 2.0). B, semi-logarithmic plot of the observed rate constant for folding and unfolding of GB88 versus [urea] measured at different pH values and 10 °C (∇ pH 10.0, ● pH 7.2, ○ pH 6.2, X pH 5.5, □ pH 4.7, △ pH 4.3, □ pH 4.0). The lines are the best fit to a two-state model.

FIGURE 5. A, semi-logarithmic plot of the observed rate constants for folding and unfolding of G88 versus [urea] measured at different pH values in presence of 0.4 M sodium sulfate at 25 °C (● pH 8.0, △ pH 7.2, ○ pH 6.7, □ pH 6.2, + pH 5.9, ◄ pH 5.5, X pH 5.0, △ pH 4.7, ○ pH 4.3, ● pH 4.0); lines are the best fit to a two-state model. B, dependence of mF (●), mU (●), and mD−U (●) on pH; lines are the best fit to an equation implying a single protonation site with pKa ∼ 5.
change in the Ca r.m.s.d. upon lowering the pH for G\textsubscript{A88}, in the case of G\textsubscript{N88}, the Ca r.m.s.d. increased at low pH. The overall residual helical content in the denatured state at neutral pH was similar for G\textsubscript{A88} and G\textsubscript{N88}. The helix content of both proteins increased when the pH was lowered. Interestingly, the helical content in the denatured state of G\textsubscript{N88} at low pH was surprisingly high, being 104% of the native extent of helix content. In contrast, G\textsubscript{A88} contained 39% of its native helix content. Although we could not directly address the helical content of the denatured states, we experimentally observed the \( m_{D-N} \) value of G\textsubscript{N88} to decrease with decreasing pH (Fig. 5B). This observation is consistent with the MD simulations, suggesting the denatured state of G\textsubscript{N88} to be more structured at acidic than at neutral pH. Importantly, such a dependence was not observed in G\textsubscript{A88}, whose \( m_{D-N} \) value was found insensitive to pH.

The relative compaction of G\textsubscript{N88} as compared with G\textsubscript{A88} can also be seen in Fig. 6. This effect is reflected in the non-polar SASAs of the two denatured states relative to their control native states (Table 1). The non-polar SASA at neutral pH was ~1000 Å\(^2\) lower in the starting structures (1330 ± 83 and 1185 ± 72 Å\(^2\) for G\textsubscript{A88} and G\textsubscript{N88}, respectively). Relative to the native state, the non-polar SASA of the denatured state at neutral pH increased by 182% in the case of G\textsubscript{A88} and by 174% for G\textsubscript{N88}. Although both proteins had some reduction of non-polar SASAs when the pH was lowered, G\textsubscript{N88} was more sensitive to acidification, and the increase in non-polar exposure upon unfolding was reduced relative to neutral pH. This finding parallels the experimental observations, which clearly indicate the denatured state of G\textsubscript{N88} to be more compact at acidic conditions, as reflected by a decreased \( m_{D-N} \) value (Fig. 5B).

TS ensembles were identified for each of the simulations described above, as well as for the two shorter simulations for each protein at both pH conditions (Fig. 7). The representative TS structures for G\textsubscript{N88} show that their helical contents and overall size were quite similar at both pHs. On the other hand, the G\textsubscript{A88} TS ensemble was very sensitive to pH, and the structures at neutral pH were much more native-like than those at low pH. The B1/B2 hairpin was more robust than the B3/B4 hairpin in the TS. Furthermore, the two hairpins did not appear to interact directly and instead behaved as different entities physically separated by the central helix.

Interestingly, the heterogeneity of the G\textsubscript{N88} and G\textsubscript{A88} TS ensembles at neutral pH were similar, particularly when the effect of the unstructured N terminus of G\textsubscript{N88} was accounted for (compare the core Ca r.m.s.d. with the “To self” values in Table 2). Both ensembles are ~5 Å Ca r.m.s.d. from their respective starting structures. The pH sensitivity of the G\textsubscript{N88} TS ensemble was dramatic; the Ca r.m.s.d. to the starting structure increased by 1.5 Å, and the average Ca r.m.s.d. between structures within the TS ensemble also increased by 1.4 Å. The core Ca r.m.s.d. was 6.2 Å for the G\textsubscript{N88} TS ensemble at low pH, and it was quite distorted with a Ca r.m.s.d. of 7.0 Å.

![Diagram](image)

### TABLE 1

| pH and temperature | Ca r.m.s.d. | % of \( \alpha \)-helix\(^a\) | % of native \( \alpha \)-helix\(^a\) | NP SASA\(^b\) | % of native NP SASA\(^b\) |
|-------------------|----------|------------------|------------------|-------------|------------------|
| G\textsubscript{A88} | N, 298 | 3.4 (0.4) | 67 (3) | 100 | 1330 (83) | 100 |
| N, 498 | 11.4 (1.3) | 15.8 (10.7) | 24 (16) | 2424 (204) | 182 (19) |
| L, 498 | 11.8 (2.5) | 26.4 (11.8) | 40 (18) | 2309 (224) | 174 (20) |
| G\textsubscript{N88} | N, 298 | 2.5 (0.4) | 26 (1) | 100 | 1185 (72) | 100 |
| N, 498 | 11.3 (1.5) | 15.5 (7.6) | 59 (29) | 2062 (198) | 174 (20) |
| L, 498 | 12.7 (1.2) | 27.4 (13.8) | 104 (52) | 1867 (191) | 158 (19) |

\(^a\) All properties were averaged over the final 30 ns of the three long 498 K simulations, and the S.D. is indicated in parentheses. The values for the 298 K simulations are defined as 100% native.

\(^b\) \( \alpha \)-Helix was defined using DSSP (27).

\(^c\) Percentage of native \( \alpha \)-helix was calculated relative to the percent of \( \alpha \)-helix in the 30-ns 298 K native simulation.

\(^d\) Nonpolar SASA (NP SASA) was defined as the sum of the SASA for all hydrophobic residues (see "Experimental Procedures").

\(^e\) Percentage of native nonpolar SASA (NP SASA) was calculated relative to the nonpolar SASA in the 30-ns 298 K native simulation.
The Denatured State Dictates the Topology of Two Proteins

FIGURE 7. Structures of the transition state at neutral and low pH. G_{A88} (above) and G_{B88} (below) are colored from red to blue from the N terminus to the C terminus. The NMR structure of the native states is shown in an ellipse in the middle for each of the two proteins. For each set, runs 1–5 are ordered clockwise from top left, to top right, to middle, to bottom left, and to bottom right. Transition state times for G_{A88} at neutral pH are at the following simulation times: 486, 483, 249, 331, and 225 ps for runs 1–5, respectively. Times for G_{A88} at low pH are at: 362, 227, 842, 77, and 165 ps. Times for G_{B88} at neutral pH are at: 422, 449, 248, 276, and 338 ps. Times for G_{B88} at low pH are at: 159, 102, 164, 305, and 147 ps. Transition states were selected as the first cluster exit in the three-dimensional multidimensional scaling of the all-against-all Cα r.m.s.d. matrix. For more details, see “Experimental Procedures.”

TABLE 2

A summary of the transition state from neutral and low pH simulations at 498 K

| pH | Core r.m.s.d. | Core r.m.s.d. |
|----|--------------|--------------|
|    | To start | To self | To start | To self |
| G_{A88} | N 5.5 (0.7) | 6.7 (1.0) | 4.2 (0.5) | 4.7 (0.4) |
|      | L 4.9 (0.5) | 5.5 (0.9) | 3.5 (0.6) | 3.7 (0.5) |
| G_{B88} | N 5.5 (1.1) | 4.8 (0.9) | 5.4 (1.1) | 4.7 (0.9) |
|      | L 7.0 (1.6) | 6.2 (1.4) | 6.9 (1.5) | 6.2 (1.4) |

* Core is residues 9–51 for G_{A88} and 1–55 for G_{B88}. The core value removes the effect of the unstructured N terminus in G_{A88}. Also note that G_{B88} low versus neutral pH is 4 Å and that the value comparing the two pH regions for G_{B88} is 6 Å.

The pairwise Cα r.m.s.d. to the starting structure was reported for all five transition state structures (n = 5).

The pairwise Cα r.m.s.d. between all five transition state structures was reported (n = 10).

from its starting structure (Fig. 7). In contrast, the G_{A88} TS remained within ~4 Å of its starting structure, and the spread within a TS ensemble was ~4.5 Å (excluding the N terminus of G_{A88}).

The overall folding pathways from representative simulations (run 1 in each case) of G_{A88} and G_{B88} are presented in Fig. 8 as the reverse of the simulated unfolding process. It appears that the topology, residual structure, and interactions in the denatured state direct whether the protein will fold into the helical G_{A88} structure or the mixed α/β G_{B88} structure. In G_{A88}, the protein has some dynamic residual structure, whereas the main chain is fairly fluid with different main chain interactions occurring over time within the collapsed state; the productive interactions with respect to folding are local along the sequence, with folding of kernels of helical structure that then dock together and consolidate in the TS ensemble. In contrast, in the case of G_{B88}, the approximate topology of the native state was already apparent in the denatured state, with segregation of the two hairpin regions by the central helix. This helix was more stable in G_{A88} than in G_{B88} due to improved packing interactions between residues 30 and 33, which in G_{A88} are both Ile, whereas in the case of G_{B88}, they are Phe and Tyr. Moreover, substitution of Gly to Ala at residue 24 and Ile to Thr at residue 25 in G_{B88} also increased the helical propensity of the region.

Interestingly, although the sequence of the N-terminal region is identical in the two proteins, there was a tendency for the region to form a helix in G_{A88} and a loose hairpin in G_{B88} (Figs. 7 and 8). This difference was, in large part, due to a hydrogen bond between Thr1 and Glu19, which was present in 99.8% of the G_{B88} denatured state structures at neutral pH. At low pH, however, the loose β1/β2 hairpin was not present because Glu19 was protonated. Although positions 1 and 19 are identical in G_{A88} and G_{B88}, in the case of the latter, we did not observe this hydrogen bond in the denatured state at neutral pH. Instead, Glu19 tended to interact with the solvent, and Thr1 either interacted with the solvent or formed hydrogen bonds with Asp47 and Glu46 (over 67 and 60% of the denatured state, respectively). The backbone φ/ψ angles of residues 47 and 48 when interacting with Thr1 were compatible with forming α3 in G_{A88}. These observations indicate that long range interactions play a critical role in the residual structure in the denatured state of these proteins.

The β3/β4 turn was also fractionally present in the denatured state of G_{A88} due to the presence of two side chain hydrogen bonds: Asp47–Lys50 and Asp47–Tyr45 (57 and 14% of the time at neutral pH, respectively). Of note, these hydrogen bonds were not present in the low pH denaturing simulations of G_{B88} due to the protonation of Asp47, nor were they present in the denatured state of G_{A88} where both Lys50 and Tyr45 are mutated to Leu. This interaction appears to stabilize the β3/β4 turn in G_{B88}, thus preventing these residues from assuming an α helical structure, as they do in G_{A88}.

DISCUSSION

Critical insights on many problems in biology have been classically achieved using simplified model systems. Although a comprehensive understanding of the folding of large multidomain proteins is still an aspiration, the successful design of two heteromorphic proteins sharing 88% sequence identity, called G_{A88} and G_{B88} (16), provides a unique opportunity to unveil the mechanism whereby a few key residues commit the polypeptide chain to its characteristic and functionally competent native topology. Here we have characterized the folding and unfolding kinetics of G_{A88} and G_{B88} by experiment...
The key findings show that both engineered proteins appear to fold via a two-state mechanism, and protein topology is committed very early along the folding pathway.

Understanding the Mechanism by Which Proteins Commit to Their Topology Highlights the Role of Long Range Interactions in Denatured States—For the purpose of this study, it is essential to understand the mechanism whereby a few key residues univocally determine native topology, i.e. which structural determinants preclude the sequence of GA88 from adopting the GB88 topology and vice versa? Experimentally, we clearly detected a difference in denatured state properties of the two proteins. In fact, we observed the mD/H11002N value of GB88 to decrease with decreasing pH (Fig. 5B). Transition state contacts are reported for all transition state ensembles (n = 30); denatured state contacts are plotted for the last 30 ns of the three long 498 K simulations (n = 90,000). Two residues were considered in contact if they contained carbon atoms that were ≥5.4 Å apart or any other non-hydrogen atoms that were ≤4.6 Å apart. Contacts were colored from white (never occurred) to black (present 100% of the time).

and simulation. The key findings show that both engineered proteins appear to fold via a two-state mechanism, and protein topology is committed very early along the folding pathway.

Understanding the Mechanism by Which Proteins Commit to Their Topology Highlights the Role of Long Range Interactions in Denatured States—For the purpose of this study, it is essential to understand the mechanism whereby a few key residues univocally determine native topology, i.e. which structural determinants preclude the sequence of GA88 from adopting the GB88 topology and vice versa? Experimentally, we clearly detected a difference in denatured state properties of the two proteins. In fact, we observed the mD/H11002N value of GB88 to decrease with decreasing pH (Fig. 5B). This observation is consistent with the MD simulations suggesting the denatured state of GB88 to be more structured at acidic than neutral pH, as mirrored both by its native helical content and by its solvent-accessible surface area (Table 1). Importantly, such a dependence was not observed in GA88, whose mD/H11002N value was found experimentally to be insensitive to pH. In summary, although only 7 of the 56 amino acids are different between GA88 and GB88, only the latter displays in its denatured state a detectable residual structure. Such a structure may be tuned by changing pH, as reflected both by analysis of mD/H11002N values as a function of pH (apparent pK_a ~5) and by comparison of the MD simulations at neutral and low pH. Surprisingly, the residues that are different between the two proteins (Fig. 1) do not include amino acids titrating below neutral pH, suggesting that the observed compaction of the denatured state of GB88 originates from non-local effects. Indeed, the MD simulations of GB88 highlight the presence of side chain hydrogen bonds in the β3/β4 hairpin turn, in the denatured state at neutral pH. These hydrogen bonds involve residue Asp47, which is protonated in the low pH simulations and does not form hydrogen bonds with either Tyr45 or Lys50 as it does at neutral pH. Interestingly, residues 45 and 50 are
both mutated to Leu in $G_A^{88}$ and are part of the $\alpha$3 helix. However, Asp$^{47}$ along with Glu$^{48}$ tends to interact with Thr$^1$ in the denatured state of $G_A^{88}$. This interaction favored backbone dihedral angles, which were compatible with $\alpha$-helix rather than the $\beta3$/4 turn, as in $G_B^{88}$.

According to AGADIR (37) and our MD simulations, the 7-residue difference in sequence between $G_A^{88}$ and $G_B^{88}$ yields an increased helical propensity for $G_B^{88}$. In particular, the 4 residues forming the loop connecting $\alpha1$ and $\alpha2$ in $G_A^{88}$ adopt a helical conformation in the structure of native $G_B^{88}$. Furthermore, MD reveals that substitution of Gly to Ala at residue 24 and Ile to Thr at residue 25 increases the helical propensity of the region. Thus, experiments and simulations converge in supporting the hypothesis that a longer, more stable $\alpha$-helix in $G_B^{88}$ prevents the latter sequence from folding to the $G_A^{88}$ structure. Interestingly, it was recently shown that a switch between the $G_A$ and $G_B$ structures may be obtained even with a single amino acid substitution (38). Under conditions where the $G_A$ fold is $>90\%$ populated, mutation of Leu$^{65}$ into Tyr shifts the population to $>90\%$ of the $G_B$ fold. Surprisingly, position 45 is not in the loop where $G_A^{88}$ and $G_B^{88}$ display a different helical propensity (Fig. 1), indicating that the greater helical content of the denatured state of $G_B^{88}$ is affected by long range interactions.

Overall, comparison of the folding of $G_A^{88}$ and $G_B^{88}$ highlights a conundrum; although only a few residues (or even a single one) are responsible for the selective stabilization of the two alternative topologies, information on the folding mechanism indicates that no single residue appears to act as a unique gatekeeper in the selection of protein topology. Both experiments and simulations on the folding of $G_A^{88}$ and $G_B^{88}$ suggest that native topology might be prescued in the denatured state, where incipient nuclei are present. Stabilization of such nuclei is affected by long range interactions, and commitment to the native fold occurs by selective stabilization of these incipient nuclei rather than by actively blocking alternative pathways.

Do Engineered Proteins Display Cooperative Folding?—An intriguing general question is whether folding is under evolutionary pressure. This issue was recently discussed by Baker and co-workers (39) in a study on the folding of a de novo designed protein, Top7, characterized by a novel non-natural topology. Although most small naturally occurring proteins fold in a cooperative manner (40), Top7 displays a non-cooperative folding mechanism, suggesting that cooperativity may be a result of natural selection. In this context, it may seem somewhat puzzling that both $G_A^{88}$ and $G_B^{88}$ fold in a cooperative manner, displaying single exponential folding and unfolding kinetics as well as V-shaped chevron plots, a hallmark of two-state folding (22). However, a possible explanation to reconcile these apparently conflicting observations is that although the structure of Top7 is non-natural (41), both $G_A^{88}$ and $G_B^{88}$ have been engineered starting from naturally occurring frameworks (16).

It was reported recently that malleability of protein folding pathways stems from the existence of multiple nuclei within a given protein structure (42). Implicit in this view, a natural topology would contain one or more nucleation motifs, representing the minimal units to encode for the final structure (43). Nucleation of the motif is the basis for cooperative folding, and the number of accessible pathways is related to the number of nucleation motifs within a protein (44). On the basis of the folding pathways observed for Top7, $G_A^{88}$, and $G_B^{88}$, it is tempting to speculate that because naturally occurring topologies contain nucleation motifs, productive folding of these substructures would result in cooperative transitions. In contrast, non-natural proteins may not contain stable nucleation motifs and, thus, they appear to fold non-cooperatively. We may conclude that evolution does not directly select for cooperative folding, but rather it selects for topologies that can fold in a cooperative manner.

Comparison with Studies on Protein Families—The study of homologous proteins has represented a powerful approach to obtain insight into protein folding (3–11, 34, 45, 46), especially when combined with structural information on intermediate or transition states. In a recent study, we addressed the structural features of the early and late transition states of two homologous three-state proteins, PSD-95 PDZ3 and PTP-BL PDZ2 (3). For different PDZ domains, we observed that the late folding transition states (TS2) are more similar to each other than the early transition states (TS1). This observation would suggest that although native topology defines the late stages of folding in a unique way, significant freedom in creating structural contacts is observed for the early events. In this perspective, it is of interest to compare the cases of $G_A^{88}/G_B^{88}$, whose folding appears to diverge early in the denatured state, with that of the PDZ family, where a strong native bias is seen only at the late stages of folding. A plausible scenario to reconcile these apparently contrasting results would imply the presence of multiple nucleation motifs at the early stages of PDZ folding. In fact, the apparent structural divergence of TS1 for PDZ2 and PDZ3 most likely arises from selective stabilization of alternative nuclei in the two denatured proteins, which may then appear to explore distinct early folding pathways. When folding proceeds to the native state, the alternative nuclei all consolidate in a native-like conformation, and folding pathways appear to converge. Support for this hypothesis comes from circular permutation experiments on PDZ2, whereby alternative nuclei may be selectively stabilized and alter the early events of folding without affecting the late ones (47, 48). On the other hand, $G_A^{88}$ and $G_B^{88}$ do not appear to contain alternative nuclei; moreover, because they display two completely different structures, their respective nucleation motifs may be completely independent such that their folding pathways diverge from the very early stages. Future work based on protein engineering experiments, MD simulations, and $\phi$-value analysis (23) will further address structural determinants in the folding of this heteromorphic protein pair in an effort to identify crucial residues for the stabilization of these two alternative topologies and their respective folding motifs and to extend the experimental and theoretical work to more complex multidomain systems.

REFERENCES

1. Editorial (2005) Science 309, 78 – 102

2. Fersht, A. R. (2008) Nat. Rev. Mol. Cell Biol. 9, 650 – 654
The Denatured State Dictates the Topology of Two Proteins

3. Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engrström, A, Ekdlund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M., and Jemth, P. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 19241–19246
4. Chi, C. N., Gianni, S., Calosci, N., Travaglini-Allocatelli, C., Engrström, K., and Jemth, P. (2007) FEBS Lett. 581, 1109–1113
5. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefan, M., and Dobson, C. M. (1999) Nat. Struct. Biol. 6, 1005–1009
6. Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999) Structure 7, 1145–1153
7. Friell, C. T., Capaldi, A. P., and Radford, S. E. (2003) J. Mol. Biol. 293–305
8. Martinez, J. C., and Serrano, L. (1999) Nat. Struct. Biol. 6, 1010–1016
9. Riddle, D. S., Balasubramanian, S., and Regan, L. (1997) J. Mol. Biol. 278, 41136–41140
10. Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2007) Proc. Natl. Acad. Sci. U.S.A. 104, 11963–11968
11. He, Y., Chen, Y., Alexander, P., Bryan, P. N., and Orban, J. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 14412–14417
12. Scott, K. A., and Daggett, V. (2007) Biochemistry 46, 1545–1556
13. Ruan, B., Fisher, K. E., Alexander, P. A., Doroshko, V., and Bryan, P. N. (2004) Biochemistry 43, 14539–14546
14. Siu, M., and Oliveberg, M. (1997) Proc. Natl. Acad. Sci. U.S.A. 94, 6084–6086
15. Santoro, M. M., and Bolen, D. W. (1988) Biochemistry 27, 8063–8068
16. Jackson, S. E., Fersht, A. R. (1991) Biochemistry 30, 10428–10435
17. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) J. Mol. Biol. 224, 771–782
18. Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995) Comput. Phys. Commun. 91, 215–231
19. Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E., and Daggett, V. (1997) J. Phys. Chem. B 101, 5051–5061
20. Beck, D. A., and Daggett, V. (2004) Methods 34, 112–120
21. Kabsch, W., and Sander, C. (1983) Biopolymers 22, 2577–2637
22. Lee, B., and Richards, F. M. (1971) J. Mol. Biol. 55, 379 – 400
23. Li, A., and Daggett, V. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 10430–10434
24. Li, A., and Daggett, V. (1996) J. Mol. Biol. 257, 412–429
25. Geierhaas, C. D., Nickson, A. A., Lindorff-Larsen, K., Clarke, J., and Vendruscolo, M. (2007) Protein Sci. 16, 125–134
26. Mayor, U., Gudosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V., and Fersht, A. R. (2003) Nature 421, 863–867
27. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M., and Fersht, A. R. (2005) Nature 437, 1053–1056
28. White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R., and Daggett, V. (2005) J. Mol. Biol. 350, 757–775
29. Sanchez, I. E., and Kiefhaber, T. (2003) J. Mol. Biol. 325, 367–376
30. Timasheff, S. N. (1993) Annu. Rev. Biophys. Biomol. Struct. 22, 67–97
31. Munoz, V., and Serrano, L. (1997) Biopolymers 41, 495–509
32. Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2009) Proc. Natl. Acad. Sci. U.S.A. 106, 21149–21154
33. Watters, A. L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., and Baker, D. (2007) Cell 128, 613–624
34. Tanford, C. (1970) Adv. Protein Chem. 24, 1–95
35. Hubner, I. A., Lindberg, M., Haglund, E., Oliveberg, M., and Shakhnovich, E. I. (2006) J. Mol. Biol. 359, 1075–1085
36. Haglund, E., Lindberg, M. O., and Oliveberg, M. (2008) J. Biol. Chem. 283, 27904–27915
37. McCallister, E. L., Alm, E., and Baker, D. (2000) Nat. Struct. Biol. 7, 669–673
38. Wenzl, B. G., Gørtner, M., Choo, W. X., Batey, S., and Clarke, J. (2009) J. Mol. Biol. 390, 1074 – 1085
39. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M., and Gianni, S. (2009) J. Am. Chem. Soc. 131, 11727–11732
40. Ivarsson, Y., Travaglini-Allocatelli, C., Morea, V., Brunori, M., and Gianni, S. (2008) Protein Eng. Des. Sel. 21, 155–160