Published in final edited form as:

J Mol Biol. 2007 November 23; 374(2): 528–546. doi:10.1016/j.jmb.2007.09.024.

Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Gō-model simulation

Zhenyu Gu¹, Maithreyi K. Rao², William R. Forsyth^{1,†}, John M. Finke^{2,‡}, and C. Robert Matthews^{1,‡}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

²Department of Chemistry, Oakland University, Rochester, MI 48309, USA

Abstract

The structures of partially-folded states appearing during the folding of a $(\beta\alpha)_8$ TIM barrel protein, the indole-3-glycerol phosphate synthase from *S. solfataricus* (sIGPS), was assessed by hydrogen exchange mass spectrometry (HX-MS) and Gō-model simulations. HX-MS analysis of the peptic peptides derived from the pulse-labeled product of the sub-millisecond folding reaction from the urea-denatured state revealed strong protection in the $(\beta\alpha)_4$ region, modest protection in the neighboring $(\beta\alpha)_{1-3}$ and $(\beta\alpha)_{5\beta_6}$ segments and no significant protection in the remaining N- and C-terminal segments. These results demonstrate that this species is not a collapsed form of the unfolded state under native-favoring conditions nor is it the native state formed via fast-track folding. However, the striking contrast of these results with the strong protection observed in the $(\beta\alpha)_{2-5\beta_6}$ region after 5 s of folding demonstrates that these species represent kinetically-distinct folding intermediates that are not identical as previously thought. A re-examination of the kinetic folding mechanism by chevron analysis of fluorescence data confirmed distinct roles for these two species: the burst-phase intermediate is predicted to be a misfolded, off-pathway intermediate while the subsequent 5 s intermediate corresponds to an on-pathway equilibrium intermediate. Comparison with the predictions using a C_α Gō-model simulation of the kinetic folding reaction for sIGPS shows good agreement with the core of structure offering protection against exchange in the on-pathway intermediate(s). Because the native-centric Gō-model simulations do not explicitly include sequence-specific information, the simulation results support the hypothesis that the topology of TIM barrel proteins is a primary determinant of the folding free energy surface for the productive folding reaction. The early misfolding reaction must involve aspects of non-native structure not detected by the Gō-model simulation.

Keywords

Protein folding; kinetic folding mechanism; misfolded intermediate; topology; stable intermediate

Correspondence to: John M. Finke; C. Robert Matthews.

[†]Present address: William R. Forsyth, TransForm Pharmaceuticals, Inc. 29 Hartwell Ave. Lexington, MA 02421.

[‡]Questions regarding the experimental protein folding study should be addressed to C.R.M. E-mail address: c.robert.matthews@umassmed.edu, Phone: (508) 856-2251; FAX: (508) 856-8358 and questions regarding the Gō-model simulations should be addressed to J.M.F E-mail address: finke@oakland.edu, Phone: (248)370-3088; Fax: (248)370-2321

Introduction

Small proteins and autonomously-folding protein domains (< 100 amino acids) often fold rapidly from chemically-denatured states without populating partially-folded states.^{1,2} By contrast, larger proteins often acquire substantial far-UV ellipticity in the sub-millisecond time range, far faster than the appearance of the native state.^{3–5} The significance of this secondary structure to the folding mechanism has been a subject of controversy for several years. One view holds that this “burst-phase” species is simply the contraction of the ensemble of unfolded states in response to a solvent that favors the native state.^{6,7} Although the reduction in the accessible conformational space may favor the subsequent folding reaction, the protein remains in the unfolded thermodynamic state. The competing view holds that, in at least some cases, the product of the burst-phase reaction is a folding intermediate that corresponds to a distinctly different thermodynamic state.^{8–10}

A discriminating feature of these alternative explanations for the burst phase reaction is the prediction of a high energy barrier separating two discrete thermodynamic states. Experimentally, one would expect an exponential response in the microsecond time range. Although barriers have been observed for cytochrome c^{11,12} and some other proteins,^{13,14} random searches of conformational space could also display exponential behavior.¹⁵ Another, perhaps more robust test, involves the development of sufficient stability in the burst-phase species to provide protection against exchange with solvent for amide hydrogens involved in main-chain main-chain hydrogen bonds. Application of this approach to the all- α apomyoglobin,⁸ the β/α sheet dihydrofolate reductase¹⁶ and the $\alpha+\beta$ ribonuclease H¹⁷ has shown protection against HX within a few milliseconds for main chain amide hydrogens in discrete regions of each protein. The absence of protection throughout the protein, as observed for the native state, is consistent with the assignment of the burst-phase reaction to the formation of a folding intermediate and not to a parallel fast track to the native state.

The folding mechanisms of $(\beta\alpha)_8$ TIM barrel proteins provide a fascinating subject for the HX test of the properties of the products of their burst-phase reactions. TIM barrels are among the most common motifs in biology. The canonical barrel has eight parallel β -strands arranged in a consecutive and cylindrical fashion with hydrogen bonds between β_1 and β_8 to seal the structure. Eight α -helices alternate in sequence with the β -strands and form an amphipathic shell surrounding the barrel; loops between the C-termini of β -strands and the N-termini of the following α -helices form active sites that are capable of catalyzing a host of chemical reactions.^{18,19} The chemically-induced unfolding of TIM barrels invariably involves at least one stable intermediate that is rich in secondary structure,^{5,20} and the kinetic folding mechanism displays additional complexities.²¹ Notable among these is the appearance of 25–50% of the far-UV ellipticity within a few milliseconds, i.e., burst-phase species that are candidates for the HX test of their stabilities and, thereby, roles in folding.

Previous studies of the kinetic folding mechanisms of three TIM barrels, the α subunit of tryptophan synthase,²¹ indole-3-glycerol phosphate synthase⁵ and a protein of unknown function from yeast, IOLI²² assumed that the burst-phase species were folding intermediates. These assumptions were based upon the sigmoidal loss of the CD amplitude of the burst-phase with increasing final denaturant concentration in refolding jumps and, in the case of α TS, on the observation of a subsequent refolding phase that accelerated at increasing denaturant concentration.²¹ The latter observation was interpreted as the unfolding of an off-pathway misfolded burst-phase intermediate for α TS and implied a barrier between the unfolded and burst-phase intermediate states. The near coincidence of the urea-melting profiles for the far-UV ellipticities of the equilibrium intermediate in sIPGS and its burst-phase species led to the conclusion that they were one and the same species, i.e., the burst-phase reaction led directly to the on-pathway equilibrium intermediate.⁵ A third type of response is displayed by IOLI,

whose unfolded form is thought to partition between off- and on-pathway intermediates in the millisecond time frame.²²

Previous HX analyses of the equilibrium folding intermediates in α TS²³ and sIGPS,²⁴ employing pepsin digestion at acidic pH where mass-labeled amides can be examined by MALDI mass spectrometry, provided insights into the structures of their respective partially-folded forms. In the case of sIGPS, the HX-MS data were sufficiently rich to allow the determination of the structures of a pair of stable intermediates predicted by kinetic studies of the folding reaction.⁵ Differences in the structures of these equilibrium intermediates were attributed to sequence-specific hydrophobic clusters formed by isoleucine, leucine and valine side chains.²⁴ Complications with the folding properties of α TS, i.e., four parallel folding channels, and IOLI, i.e., a weakly-folded and self-associating native state, led to the choice of sIGPS as the initial target of an HX-MS analysis of protection in its burst-phase species.

The product of the sub-millisecond folding reaction for sIGPS was found to provide significant protection against HX for a subset of its amide hydrogens that differs markedly from that derived from previously-identified on-pathway equilibrium intermediates.²⁴ A re-examination of the kinetic folding mechanism revealed an off-pathway burst-phase intermediate whose unfolding under refolding conditions leads to the on-pathway intermediates. Complementary G δ -model simulations of a C α representation of sIGPS support the conclusion that the topology of a TIM barrel protein dictates its funnel-shaped folding free energy surface and the presence of the on-pathway intermediates.

Results

Stability and structure of the product of the apparent burst phase intermediate for sIGPS

Full-length IGPS from *S. solfataricus*, containing 248 residues, has proven to be an intractable candidate for folding studies because it spontaneously aggregates upon refolding from its chemically-denatured state.^{25,26} However, the deletion of the first 26 residues, containing the α 00 α -helix, results in a fully-reversible unfolding reaction that enabled a previous HX analysis of the structures of its equilibrium intermediates.²⁴ This truncated form of sIGSP also served as the target of the present study of its sub-millisecond folding reaction.

The kinetic folding mechanism of truncated sIGPS (Scheme 1) was previously postulated based on the results of stopped-flow and manual-mixing circular dichroism (CD) experiments.⁵ The urea-denatured state, U, is presumed to fold within milliseconds to a highly-structured intermediate, I_a, that is also highly populated under equilibrium conditions at 5 M urea. I_a converts to another stable intermediate, I_b, via a slow, ~100 s, urea-independent conformational rearrangement reaction before proceeding to the native state, N. The I_b → N reaction is another slow but urea-dependent step consistent with the burial of additional surface area required to achieve the full TIM barrel fold. Given the extensive protection against exchange observed for I_a in a previous HX-MS study of the equilibrium intermediates,²⁴ the products of a quenched-flow pulse-labeling were expected to display similar pattern of protection in the same ($\beta\alpha$)₂₋₅ β ₆ region of sIGPS within the first few milliseconds of folding. If, however, the product of the burst-phase reaction is a collapsed form of the unfolded state, the protection should be minimal throughout the sequence.

The HX test for the protection pattern of the product of the sub-millisecond apparent folding reaction for sIGPS was executed by stopped-flow dilution of fully deuterated urea-denatured protein to induce folding, pulse-labeling with ²H₂O at pH 9.5 for 50 ms at 25 °C and at various folding times in the hundreds of milliseconds time range before a final quenching step with ice-cold trifluoroacetic acid to pH 2.7 to halt exchange. Following pepsin digestion, the resulting 19 peptides covering 90% of the sequence (Figure 1) were mass-analyzed by MALDI

MS. Similar measurements on the native and fully deuterated states of sIGPS provided normalization controls and, thereby, access to the percent protection of each peptide as a function of folding time. Representative MS spectra for four different peptides at several folding times in the millisecond time range are shown in Figure 2, and the percent protection for each of the 19 peptides after 75 ms of folding is shown in Table 1. Peptide 36–47 shows weak protection that appears to decrease as the folding time increases from 75 to 214 ms (Figure 2A), peptide 100–115 shows strong protection that increases with folding time (Figure 2B), peptide 143–151 shows weak protection that does not change obviously over the folding time examined (Figure 2C), and peptide 197–204 does not offer significant protection during the first 214 ms of folding (Figure 2D). Similar to peptides 36–47 and 143–151 (Figures 2A and 2C), the mass spectra of most of the peptides at different refolding intervals consist of a pair of isotope envelopes corresponding to folded reference and unfolded reference respectively. This behavior is characteristic of peptides that exchange via the EX1 mechanism, whereby refolding from the exchange-competent state is slower than exchange of the exposed amide hydrogen with solvent.^{27,28} In such cases, the retardation of exchange, i.e., the protection factor, is proportional to the rate of unfolding and not to the relative stability of the two states. Several peptides, e.g., peptide 100–115 (Figure 2B), yielded mass spectra that appear to reflect exchange by both EX1 and EX2 mechanisms. For these peptides, rapid local unfolding/refolding reactions may couple with slower, larger scale reactions.

A continuum of protection from 0 to 73% is displayed by the 19 peptides as the folding time was progressively increased from 75 to 214 ms prior to labeling (Figure 3). At 75 ms, peptides from the N- and C-termini, e.g., 1–8, 9–15, 162–179, 180–196, 197–203 and 211–220, showed no measurable protection, $\leq 10\%$ (Table 1). The corresponding segments, which comprise α_0 (an additional N-terminal helix) α_6 , β_7 , α_7 and α_8 (a peptide containing β_8 was not recovered), are evidently not involved in stabilizing the product of the sub-millisecond folding reaction. Proceeding towards the center of the sequence from both termini, weak protection, 10–40%, was observed for peptides 24–35 (β_1), 36–47 (α_1), 48–55 (α_1/β_2), 69–83 (α_2/β_3), 87–97 (α_3), 123–134 (α_4/β_5), 135–141 (β_5/α_5), 143–151 (α_5/β_6) and 146–159 (α_5/β_6). Strong protection, $> 40\%$, was observed for peptides, 100–115 (α_3/β_4), 107–120 (β_4/α_4) and 109–121 (α_4).

The percent protection of five peptides, 87–97, 100–115, 107–120, 109–121 and 123–134, which represents a continuous stretch of residues covering α_3 , β_4 , α_4 and β_5 , are distinguished from the remaining peptides in that the protection appears to increase with folding time. Curiously, the protection appears to decrease for peptide 36–47 (α_1) and, possibly, 1–8 (α_0), unlike the nearly constant protection displayed by the remaining peptides over this time frame. With the exception of peptides 87–97 and 100–115, the apparent increases and decreases cited above for peptides 1–8, 36–47, 107–120, 109–121 and 123–134 are comparable to the estimated error in the percent protection, $\pm 5\%$, and must be viewed with caution.

Although the protection pattern is complex, the HX-MS data clearly show that the product of the burst-phase reaction for sIGPS is not simply the unfolded state under native-favoring conditions. Nor is this species the native TIM barrel structure, which shows strong protection throughout the sequence. Rather, a partially-folded state with significant structure in the center and progressively diminishing structure towards both termini is populated within a few milliseconds after folding is initiated by dilution of denaturant.

Strikingly, a comparison of these data with HX-MS data reported previously for a mixture of the I_a and I_b intermediates populated at 5 M urea (Table 1 and Figure 4)²⁴ shows that the products of the burst-phase CD reaction and the on-pathway I_a intermediate are not identical as originally predicted.⁵ The equilibrium HX-MS data were interpreted to mean that the I_a species displays strong protection against exchange for a continuous stretch of residues spanning $(\beta\alpha)_{2-5}\beta_6$.²⁴ At 5 M urea, the minor I_b species, which extends protection to α_1 and

α_6 , β_7 and α_7 , accounts for the ~25% protection observed in peptides 36–47, 162–179 and 180–196.²⁴ The discrepancies between the protection pattern displayed after 75 ms, i.e., for the I_{bp} species, and that expected for the I_a species motivated a re-examination of the kinetic folding mechanism of sIGPS.

Kinetic folding mechanism of sIGPS revisited

A previous study of the folding mechanism of sIGPS employed manual-mixing and stopped-flow CD spectroscopy to monitor the time-dependent formation and disruption of the global secondary structure.⁵ The folding mechanism in Scheme 1 was deduced from these data and the results of an equilibrium denaturation analysis of the thermodynamic properties.

To complement the previous study, stopped-flow fluorescence spectroscopy was used to provide insights into the development/disruption of tertiary structure during the early stages of folding of sIGPS. The eight tyrosines sequestered in the N-terminal half of the sequence were excited at 280 nm and their combined emission was monitored above 320 nm with a cut-off filter. A representative kinetic trace over the first 1000 s of folding is shown in Figure 5A as fluorescence intensity vs log folding time. The fluorescence intensity decreases over the first few hundred milliseconds before increasing to reach an equilibrium value in the several hundred seconds time range. A fit of these data revealed three exponentials with time constants of 0.6 s, 7.3 s and 80 s at 0.7 M urea and pH 7.8 at 25 °C; the amplitude of the 0.6 s (~47%) phase is opposite to those for the 7.3 s (28%) and the 80 s (72%) phase. Although the slowest phase was also observed by CD spectroscopy, the two faster phases were not. As the final urea concentration for refolding jumps was increased, the time constants for the fastest refolding phase decreased and the medium refolding phase increased, respectively (Figure 5B). The time constant for the slowest detected refolding phase by stopped-flow fluorescence was independent of denaturant below 1 M urea and increased above 2 M urea, precisely as observed by CD measurements.⁵ The urea-independent refolding phase with a time constant of ~700 s observed by CD is beyond the detection limit of stopped flow apparatus and was not monitored in the present study. This phase is thought to reflect the dissociation of a dimeric form of the intermediates⁵ and was not considered further.

The acceleration of the fastest refolding phase at increasing final denaturant concentration is a signature of an off-pathway intermediate whose unfolding under refolding conditions limits access to subsequent on-pathway intermediates. By implication, the U state folds to the off-pathway I_{bp} state within the dead time of mixing, < 5 ms. The I_{bp} state then unfolds in the hundreds of milliseconds time range to return to the productive folding pathway *via* the U state. A revised kinetic folding mechanism consistent with the stopped-flow fluorescence data is shown in Scheme 2. The deceleration of the subsequent refolding phase at increasing denaturant concentration is consistent with its attribution to a refolding reaction. A comprehensive analysis of these data showed that this phase corresponds to the $I_b \rightarrow N$ step in Scheme 2; the results of this analysis will be presented elsewhere (Gu, Kathuria and Matthews, unpublished results). The roll-over in the time constant for the slowest refolding phase at ~2 M urea, from urea-dependent above to urea-independent below that denaturant concentration, suggests that two different microscopic steps limit folding over this range of conditions. Under moderate refolding conditions, the rate-limiting reaction is the urea-dependent refolding of the I_b species to the N species. Under strongly folding conditions, the preceding urea-independent interconversion of I_a to I_b is rate-limiting. The assignment of the observed refolding time constants to specific steps in the revised kinetic mechanism (Scheme 2) is shown in Figure 5. Note that the observed time constants are composites of microscopic rate constants; the assignments shown in Figure 5 reflect the dominant but not sole contribution to the observed time constant.

The revised mechanism predicts that the product of the sub-millisecond folding reaction, I_{bp} , is distinct from the subsequent intermediate populated after the first few seconds of folding, I_a , as shown in Figure 4. Further, the stable structure in this transient version of the I_a species, whose lifetime is ~ 100 s under strongly folding conditions (Figure 5B), should be identical to that previously determined by HX-MS analysis of the I_a species populated at equilibrium.²⁴ This prediction was tested by determining the protection pattern against hydrogen exchange after 5 s of folding when the I_a species should dominate the population. Thus, a 5 s pulse of $^2\text{H}_2\text{O}$ after 5 s of folding at pH 7.8 and 25 °C was used to provide a pulse intensity very similar to that experienced by the burst-phase intermediate for a 50 ms pulse at pH 9.5 at 25 °C.²⁹ The percent protection for the same peptides derived from the I_{bp} intermediate, the intermediate populated after 5 s and the equilibrium intermediates populated at 5 M urea²⁴ is shown in Table 1 and plotted in Figure 4. Note that the protection pattern for the equilibrium intermediates contains contributions from both I_a and I_b .²⁴

The strong protection, $\geq 40\%$, observed in peptides 48–55, 69–83, 87–97, 100–115, 107–120, 109–121, 123–134, 135–141 and 143–151, which represent a nearly continuous stretch of residues spanning $\beta_2(\beta\alpha)_{3-5}\beta_6$ for both the 5 s kinetic intermediate and the equilibrium intermediate, I_a , populated at 5 M urea confirms their identity and validates the revised mechanism. The only exception to the strong protection is displayed by the peptide corresponding to residues 59–68, a string of primarily polar side chains between β_2 and α_2 that are not protected under any conditions. Close inspection of the data in Figure 4 and Table 1 shows moderate protection, $\sim 25\%$, in peptides 36–47, 162–179 and 180–196 in the mixture of I_a and I_b present at 5 M urea, but little or no protection in these same peptides in the species populated after 5 s of folding, i.e., I_a . The complex HX properties of the peptides derived from the mixture of the I_a and I_b species at 5 M urea previously led to the conclusion that I_a is structured in the $\beta_2(\beta\alpha)_{3-5}\beta_6$ region and I_b has an expanded structure that includes α_1 , α_6 and β_7 .²⁴ The different protection observed for peptides 36–47 (α_1), 162–179 (α_6) and 180–196 (β_7/α_7) between the 5 s intermediate and the equilibrium intermediate at 5 M urea are precisely those predicted to be protected in I_b but not I_a . The implied structures for the product of the burst-phase reaction, I_{bp} , the product of 5 s of refolding, I_a and previously studied I_b ²⁴ are portrayed on ribbon diagrams in the folding scheme of sIGPS in Figure 6.

To test whether these differences reflect the different HX pulse regimes, the equilibrium intermediate at 5 M D_4 -urea, as a control, was also labeled with the same quenched-flow parameters as the burst-phase intermediate, i.e., a 50 ms pulse-label at pH 9.5 and 25 °C. The same protection pattern was obtained as with the 5 s pulse at pH 7.8 (Figure 4). The small, $\sim 10\%$, but consistently higher percent protection for all of the peptide including those from N- and C-termini may reflect the necessity to pulse-label the pre-deuterated intermediate with H_2O . The stabilization of sIGPS in $^2\text{H}_2\text{O}$ could result in the appearance of a small fraction of the native state and uniform protection of all peptides at 5 M urea.^{30,31} Thus, the product of the burst-phase folding reaction is an off-pathway misfolded form that represents a thermodynamic state distinct from the unfolded state (Scheme 2). However, the burst-phase intermediate is not the on-pathway equilibrium intermediate, I_a , as previously concluded.⁵

Gō-model simulation of sIGPS folding mechanism

The relative contributions of the sequence and the main-chain topology to the structures of sIGPS folding intermediates can be assessed by comparison of these experimental results with the predictions of a native-centric C_α Gō-model simulation of the folding reaction. Although a Gō-model might be expected to eliminate intermediates and fold directly to the native conformation, variations in loop lengths between elements of secondary structure, variations in packing density, etc., might lead to differential stabilization of sub-domains and, thereby, folding intermediates. A previous Gō-model analysis of the folding mechanism of αTS ³²

revealed partially-folded states that strongly resembled those identified by HX-MS analysis of peptic peptides²³ and native-state HX-NMR analysis of the full-length protein (Vadrevu, Wu and Matthews, manuscript in preparation).

Both equilibrium and kinetic folding Gō-model simulations were performed on a C_{α} -model built on the crystal structure of the truncated version of sIGPS, PDB code: 2c3z, examined in the complementary HX-MS study. Equilibrium folding simulations were attempted to determine the folding temperature, T_m , at which the protein, which by repeated folding and unfolding, spends equal amounts of time in both states. Unfortunately, the folding/unfolding rate became sufficiently slow between 370K and 380K, that it was not possible to determine that the simulation had reached equilibrium. T_m was estimated from the observation that trajectories of unfolded sIGPS would refold irreversibly below 370K and trajectories of native sIGPS would unfold irreversibly above 380K. With this limitation, the present analysis focused on kinetic refolding simulations at 340K ($\sim 0.9T_m$) in lieu of equilibrium simulations. This temperature also facilitated the acquisition of a statistically-significant number of trajectories each covering a feasible number of simulation time steps (45×10^6). The kinetic folding simulations have the advantage of more closely approximating the data obtained from the HX-MS experiments on the sIGPS kinetic refolding mechanism.

A series of 30 independent refolding kinetic simulations were conducted as described previously,³² and 3 representative trajectories are shown in Figure 7A. Under refolding conditions, i.e., at $T = 340K$, most C_{α} -models of sIGPS fold within 45×10^6 time steps. As is evident in Figure 7A, the number of native contacts Q stops at different levels, categorized as incremental ranges of $100Q$ denoted as folding stages A–F, before achieving the native fold at $Q \approx 500$ contacts. As shown in Figure 7A, the 30 trajectories fall into 3 types: (1) 9 fast productive trajectories which move through each intermediate stage without significant pauses (black trace); (2) 14 slow productive trajectories which fold only after numerous aborted attempts to fold from a $Q \approx 375$ intermediate ensemble in folding Stage D (green trace); and (3) 7 unproductive trajectories which remain in the $Q \approx 375$ intermediate ensemble (yellow trace). The successful completion of 23 of the 30 simulations provided enough sampling of the complete sIGPS folding pathways to permit a comprehensive analysis. Unless stated otherwise, simulation data on sIGPS folding in the present study are derived from the 23 successfully refolded trajectories and do not include the 7 simulations which did not reach the native state.

The significance of each folding Stage A–F can be quantified by compiling the results of all 23 productive simulations into a probability of trajectories sampling each value of Q (Figure 7B). Although small peaks appear near Q values of 150, 225 and 325, their low probabilities render their significance ambiguous. By contrast, the high probability for the set of native contacts with a maximum value of Q at ≈ 375 contacts suggests the presence of a partially-folded state.

The Gō-model simulation results in the form of predicted percent folded for segments corresponding to the peptic peptides can be compared to the observed percent protection results for I_{bp} , the 5 s intermediate, I_a , and the mixture of I_a and I_b at 5 M urea (Figure 4). In general, there is very good agreement between the core of structure in the I_a and I_b intermediates defined by protection against HX and that predicted by the Gō-model simulation. Strong protection, $> 40\%$, was observed in experiments and predicted by the simulation for peptides spanning the region from residues 48–159, $(\beta\alpha)_{2-5}\beta_6$. The simulation overestimated the protection at the N-terminus for residues 9–47, $\alpha_0(\beta\alpha)_1$ and near the C-terminus for residues 146–179 $(\beta\alpha)_6$. The most apparent discrepancy exists at the N-terminus where the simulation predicts very strong protection in peptide 1–8 and the HX-MS results show just the absence of protection. Considering that the simulation only contained native-like information on the main-chain topology, the agreement between the segments with strong protection against exchange, i.e.,

the observed and predicted structures for the on-pathway equilibrium intermediates, is quite remarkable (Figure 4).

Discussion

Burst-phase intermediate or collapsed unfolded state for sIGPS?

HX-MS analysis of protection of main-chain amide hydrogens against solvent exchange for the product of the millisecond denaturant-dilution reaction for sIGPS provides definitive evidence that this species is not a collapsed form of the unfolded state. The time constant for exchange from an unprotected amide hydrogen under these conditions, ~10 ms, and the EX1 exchange behavior displayed by the peptides imply that the lifetime of the species providing protection must exceed 10 ms. When considered with the observed several hundred millisecond unfolding reaction under refolding conditions (Figure 5B), these results are not consistent with the expectations of a collapsed unfolded state. Rather, sIGPS initially misfolds to an off-pathway species with substantial secondary structure that is slowest to unfold (and, thereby, offers protection against exchange) in the central $(\beta\alpha)_4$ region; flanking regions appear to unfold more rapidly (Figure 4).

The apparent increase in protection over the first few hundred millisecond of folding for a continuous string of residues from 87–134 (Figure 3), corresponding to $\alpha_3(\beta\alpha)_4\beta_5$ in the native barrel structure, may reflect a folding reaction that cannot be detected by either fluorescence or CD. This behavior cannot be due to the ~600 ms back-tracking reaction because only a subset of the peptides that are strongly protected in the product of that reaction, I_a , display the enhanced protection with time. The apparent decrease in protection for peptides 36–47 (α_1) and, possibly, 1–8 (α_0) in the same time range might reflect the same process, however, the changes are sufficiently small that this conjecture must be viewed as very tentative. The full development of the misfolded off-pathway intermediate, I_{bp} , might, therefore, require more than a single kinetic step.

Structure of the folding intermediates in sIGPS

The HX-MS determination of the structures of the transient and stable folding intermediates for sIGPS, i.e., I_{bp} , I_a and I_b , provides insights into a process by which the information encoded in the 1D amino acid sequence is translated into the 3D structure of this $(\beta\alpha)_8$ protein. Within a few milliseconds, unfolded sIGPS acquires strong protection near the center of the sequence, the $(\beta\alpha)_4$ region, and weaker protection in the flanking regions spanning α_1 to β_6 . Comparison with the protection patterns observed for the intermediate highly populated after 5 s of folding, I_a , and for the composite mixture of equilibrium intermediates, I_a and I_b , (Figure 4) shows that I_a and I_b offer strong protection from $\beta_2(\beta\alpha)_3\text{--}\beta_6$. I_b also protects α_1 , α_6 , α_7 and β_7 from exchange; the reduced level of protection in these regions, ~25%, reflects the lesser population of I_b relative to I_a .²⁴ The strong protection in I_{bp} for a smaller segment of the sequence than I_a suggests that the partitioning between I_{bp} and I_a (Scheme 2) favors the off-pathway reaction because it requires the organization of a smaller structure centered around $(\beta\alpha)_4$. The implied requirement for the more extensive development of structure in the transition state to access the I_a intermediate would make that pathway less likely in the face of competition with the off-pathway reaction. From the perspective of the folding free energy surface, the transition state free energy penalty engendered by the incomplete compensation of activation enthalpy and entropy is apparently smaller for folding to I_{bp} than for I_a .

Although the $U \rightarrow I_{bp}$ reaction is kinetically preferred over the $U \rightarrow I_a$ reaction, the I_{bp} species lacks the large-scale cooperativity characteristic of globular proteins. Specifically, the protection pattern in the I_{bp} species does not display the uniform level expected for a single unfolding unit via the EX1 mechanism. Only the central $(\beta\alpha)_4$ segment is strongly protected

after 75 ms of folding. By contrast, both the 5 s intermediate, i.e., I_a , and the subsequent on-pathway intermediate, I_b , show strong and approximately uniform protection in their core regions from α_3 to α_5 under conditions where both I_a and I_b exchange by a EX1 mechanism. The flanking regions, containing β_2 , α_2 , α_5 and β_6 , are somewhat less strongly protected implying that these segments are not tightly coupled to the $\alpha_3(\beta\alpha)_{4-5}$ core. The EX1 exchange behavior of these intermediates precludes a comparison of their thermodynamic properties.²⁴ However, the marginal stability of I_{bp} estimated from stopped-flow CD measurements,⁵ clearly demonstrate that I_{bp} is not sufficiently stable that it would compete with I_a , I_b or N as the final product of the folding reaction of sIGPS. I_{bp} re-equilibrates with I_a , I_b and, ultimately, N over the several hundred second time range to achieve a stable, fully-folded $(\beta\alpha)_8$ barrel responsible for biological function. Productive folding begins with a minimal $\beta_2(\beta\alpha)_{3-5}\beta_6$ template (I_a), adds the adjacent α_1 , α_6 and β_7 segments (I_b) and then adds the remaining terminal segments, α_0 , α_7 , β_8 and α_8 to close the barrel (N).

The strong protection observed in the I_{bp} intermediate for the $(\beta\alpha)_4$ module may be correlated with the high density of non-polar side chains in this segment (Figure 1). Eleven of the 22 residues spanning positions 100–121 are isoleucines (2), leucines (7) and valines (2). The 100–113 segment constitutes the maximum on a hydropathy plot (data not shown) as calculated by the method established by Kyte and Doolittle,³³ and the 2D contact map shows that this segment has the highest density of 4.2 Å contacts between non-polar side chains in sIGPS.²⁴ Thus, the strong protection against HX for $(\beta\alpha)_4$ likely reflects the rapid formation of a hydrophobic cluster from these side chains. Either the concomitant formation of H-bonds and/or the exclusion of solvent from the associated amide hydrogens is responsible for the retardation of exchange observed by HX-MS.

The observation that strong protection is limited to this small segment suggests that the segment may adopt a non-native fold, incapable of propagating native-like structure to the neighboring $\beta\alpha$ -repeats. The weaker protection displayed by these neighboring segments, along with the substantial far-UV ellipticity of the I_{bp} species,⁵ could mean that these segments adopt weakly-folded $\beta\alpha$ -hairpins that cannot couple tightly with a non-native $(\beta\alpha)_4$ conformation. Alternatively, the weak protection in the same region where strong protection is displayed by I_a , $\beta_2(\beta\alpha)_{3-5}\beta_6$ (Figure 4), could be a result of the formation of a small fraction of I_a within 75 ms by the partitioning of the U state between I_{bp} (the major fraction) and I_a (the minor fraction). In the latter case, only the $(\beta\alpha)_4$ segment would be strongly protected in I_{bp} . A recent global analysis of the folding mechanism of IOLI²² found precisely this type of partitioning between off- and on-pathway intermediates during the earliest events in folding. A similar global analysis of sIGPS is in progress and will test for partitioning at the earliest stage of folding (Gu, Kathuria and Matthews, unpublished results).

A caveat for the structural interpretation of I_{bp} from the HX-MS data is the possibility that the alkaline pH required to effect exchange of solvent-exposed amide hydrogens within 50 ms unfolds marginally-stable elements of structure that exist at neutral pH.³⁴ Thus, the structure for I_{bp} implied by the data in Figure 4 and Table 1 may represent a minimal view. The good agreement between the results of the manual-mixing labeling experiment at 5 M urea and pH 7.8 and the quenched-flow alkaline pH labeling experiment of I_a and I_b equilibrated at 5 M D₄-urea (Figure 4) demonstrates that both I_a and I_b are resistant to possible alkaline unfolding.

Comparison with the folding intermediate predicted by Gō-model simulations

Many theoretical studies have investigated the determinants of protein folding.^{35–39} A model of protein folding which has yielded a number of recent successes is the Gō-model.^{32, 40–42} The Gō-model is predicated on the assumption that only native contacts are attractive.⁴³ The sIGPS Gō-model used in the present study also assumes that all native contacts are equal in their attractive energy. These assumptions rest on the hypothesis that evolution has selected

for these native contact energies and de-selected against non-native contact energies, either through direct selective pressure or through co-evolution with other properties of the protein.⁴⁴ If this hypothesis is true for sIGPS, then one would expect good agreement between the folding simulations of the G δ -model and experimental measurements of the structures of folding intermediates for sIGPS. If other factors, such as non-native attractive energies, play a role then one would not expect a good agreement between the G δ -model simulations and experimental measurements of the folding of sIGPS.

The agreement between the predicted percent protection for the Q $\approx$ 375 species and observed percent protected for the peptic peptides derived from the mixture of the I_a and I_b intermediates for sIGPS is quite good (Figure 4). All 10 of the peptides derived from the central region of sIGPS, residues 48–159, are predicted and observed to be strongly protected against HX in intermediates that are highly populated during the folding reaction, i.e., from I_a and I_b. The strong protection predicted for peptide 162–179 mimics that expected for the I_b equilibrium intermediate and suggests that the simulation captures its properties more closely than those of the I_a species. Although the simulation tends to over-estimate the percent protection for folded regions and incorrectly predicts the folding of the N-terminus (residues 1–8), the simulation correctly predicts 10 of the 11 strongly-protected peptides (and possibly all 11 in the I_b species).

Agreement between the sIGPS G δ -model and HX experimental data is a consequence of the fact that the attractive energy between native contacts has evolved to be much stronger than the energy of non-native contacts. In this situation, the energy will always decrease as the protein folds to the native state; hence the term “folding energy funnel.”⁴⁴ Alternatively, the energy may not be funneled in some regions of a protein or possibly throughout the entire protein. One particularly interesting possibility is that residues required for the function of the protein are not necessarily those that would favor the funnel landscape. Conserved functional residues may introduce “frustration” in protein folding in two ways: (1) “energetic frustration”, i.e., - functional residues may form stable non-native contacts more readily than non-functional residues; (2) “topological frustration”, i.e., the requirement to place functional residues in an optimal spatial position may result in a complex local topology. Because G δ -models do not include non-native contacts, energetic frustration caused by functional residues can be identified by difference between the G δ -model simulation and experimental data. Topological frustration can be directly observed by the formation of native contacts that inhibit further native contacts from forming. Topological frustration can manifest itself either as off-pathway species, which requires global unfolding to correct, or as a “backtracking” reaction, which requires a decrease in contact propensity in some areas to enable contacts in other regions to form and promote productive folding. In the G δ -model simulation a decrease in the overall Q value is required to correct the first type of topological trap (global unfolding). In the second type of trap, back-tracking involves a decrease in a small number of contacts as the overall folding progresses (i.e. total Q increases). Off-pathway intermediates can be clearly observed in plots of total Q versus time (Figure 7A). On the other hand, detection of backtracking is more subtle and requires detailed quantitation of individual contacts throughout increasing stages of folding. If any contact decreases its probability while the overall Q of the protein increases, then this contact is defined as backtracking.

The totality of folding pathways from the 30 sIGPS trajectories is indicated in Figure 8, where the fraction of folding for all 18 peptic fragments is indicated during simulations in folding stages A–F. In Figure 8, for the 23 productive trajectories, colors indicate intermediates which are biased towards N-terminal folding (red), central folding (black), and C-terminal folding (blue) within each folding stage. These colors are assigned qualitatively to aid the eye and may change for a given trajectory as its structure changes throughout later folding stages. Also, intermediate structures of the 7 unproductive trajectories are shown in yellow.

For the largely unfolded structures in Stage A (Figure 8A), all but a pair of structures are essentially the same, indicating little topological heterogeneity. There appears to be no structural difference between productive (black) and unproductive (yellow) trajectories. Within the early folding intermediates in Stage B, a high degree of heterogeneity exists as three separate groups of folding pathways emerge, each involving either N-terminal contacts (7 trajectories), central contacts (10 trajectories), and C-terminal contacts (13 trajectories). Each of these groups contains at least one trajectory which did not completely refold by the end of the simulation (yellow), indicating that the structural determinants of unproductive trapping in Stage D have not emerged at this folding stage. Further in folding in Stage C, this heterogeneity decreases with fewer intermediates having N-terminal (4) and central (3) structures and more C-terminal (23) structures. As in Stage B, the unproductive trajectories do not completely correlate with one particular intermediate group. By stage D, all trajectories are either central (18) or C-terminal (5) and appear much more homogeneous. However, unproductive trajectories in stage D all have structures more similar to the central structures (black) and not C-terminal structures (blue). In Stage E, all 23 productive trajectory structures (black) are highly similar to each other but are all slightly dissimilar from the 7 unproductive structures (yellow). Their primary difference lies in increased contacts in the central region for nonproductive trajectories and fewer contacts at the termini. In Stage F, all 23 folded structures appear similar as expected (the 7 unproductive trajectories did not reach this stage).

In Figures 8A–8E, it should be noted that some stages (A, D and F) represent structures with the characteristics of intermediates, since they are much more populated (Figure 6B). Other less populated stages (B, C and E) will have properties more similar to transition states. Since HX experiments can only probe populated species, experiments in the present study can only be used to compare with highly populated stages (A, D and F) of sIGPS folding. Less populated states (B, C and E) may be more easily studied with experiments which probe transition states, such as ϕ -values.⁴⁵

Are nonproductive trajectories the result of topological frustration? Observing the nonproductive (yellow trace) trajectory in Figure 7A, this possibility seems likely since the trajectory appears to make short attempts forward to Q~425 but then falls back to Q~375. Thus, ~50 native contacts are being made from the Q~375 state which are later forced to break apart since they are not correctly ordered. Until this topological frustration is alleviated, folding to the native state will not occur. For the fast productive trajectory in Figure 7A (black trace), this topological frustration does not appear to exist since it continues through Stage D without any reversal of the overall Q. Also, as shown by the slow productive trace in Figure 7A (green trace), some topologically frustrated sIGPS Gō-models can correct themselves and reach the native state. Since the intermediates in Stage D do not appear to unfold globally to correct the topological frustration, one would expect backtracking to be the mechanism by which the topological frustration would be alleviated.

Is backtracking a general mechanism which helps the protein reach the native state for the sIGPS simulations? Figure 8E supports this explanation since productive trajectories (black) all have lower central contacts than nonproductive (yellow) trajectories in Stage E. Because Stage E is analogous to a low-population transition state (Figure 7B), the formation of the final contacts in the terminal regions (particularly α_7) may face an insurmountable free energy barrier if fully-folded central residues present excessive steric exclusion.

Figure 9 further supports a backtracking mechanism in sIGPS. Figure 9 shows the average fraction of folded residues in each peptide fragment for folding Stages B–E, normalized such that Stage A peptides are 0 and Stage F peptides are 1. Only the 23 productive trajectories are included in each average. For reference, the data shown for Stage D in Figure 9 are the simulation data used to compare with experiments in Figure 4. While no backtracking occurs

early in folding between Stages B, C, and D, Stage E distinctly shows lower contact probabilities (within statistical errors) in peptides 84–97 (α_3), 107–120 (β_4/α_4) and 109–121 (α_4), than the prior Stage D. Because these differences are greater than the statistical error, backtracking largely appears to be a general feature of the 23 productive sIGPS Gō-model folding trajectories. A analysis of individual trajectories reveals that all 7 nonproductive trajectories do not backtrack (data not shown). However, backtracking is not always necessary since 6 of the 23 productive trajectories do not backtrack as well, indicating that, while helpful, backtracking is not obligatory to folding (data not shown). For comparison, the exact contacts which backtrack are overlayed on the full 2D contact map of sIGPS in Figure 10. From Figure 10, one can observe that backtracking contacts are predominantly local across the α_3 - α_4 region. Validating these backtracking contacts experimentally will require additional studies since the present HX experiments are not capable of probing backtracking in “transition-state-like” stages of folding, such as the low probability Stage E in the sIGPS Gō-model.

Figure 9 also indicates the location of sIGPS active site residues K28, R157, G187, I207, G208, and S209. Residue K28 falls within peptide 36–47 (α_1), R157 falls within peptide 146–159 (α_5/β_6), G187 falls within peptide 180–196 (β_7/α_7), and residues I207–S209 fall between peptides 197–203 (α_7) and 211–220 (α_8). An interesting question is whether these active site residues impose topological and/or energetic frustration in regions of sIGPS required for activity. Topological frustration in simulations, indicated by backtracking in peptides 87–97, 107–120, and 109–121, does not correlate with the active site residues. Frustration in the HX exchange experiments is indicated by the decreasing contact probabilities in peptides 1–8, 24–35 (β_1), and 36–47 (α_1) between 75 msec and 5 sec, although the experiments do not reveal whether this frustration is topological or energetic. Energetic frustration is evidenced by deviation between the Gō-model simulations and HX exchange experiments, which correspond to sIGPS peptides 1–8, 9–15, 24–35, and 36–47 (Figure 4). This peptide correlation between (a) destabilized contacts in the $I_{bp} \rightarrow I_a$ reaction and (b) simulation-experiment discrepancy, suggests that the frustrated contacts in I_{bp} are due to energetic frustration at the N-terminus. Regardless, only one active site residue (K28) correlates with the regions of this energetic frustration.

Thus, for sIGPS, there is not a strong correlation between active site residue location and regions of topological or energetic frustration. A re-examination of past TIM barrel Gō-model simulations shows mixed agreement with this finding.^{32,46} For α TS, there was almost no energetic frustration evident as the agreement between experiment and simulation was nearly perfect.³² Topological frustration in α TS was evident as both simulations and experiments have captured an N-terminal off-pathway intermediate, I_{bp} , with structure between regions β_1 – β_3 .³² In these simulations, I_{bp} has the β_2 - α_2 loop incorrectly wrapped around the β_3 - α_3 loop such that the α_2 contacts cannot be fully made – a true topological trap.³² Some backtracking in the on-pathway stages of α TS Gō-model folding is also observed at particular contacts in the β_1 – β_3 region during a final folding stage (Stage D to Stage E), although backtracking of contiguous groups of contacts was not observed. A number of α TS active site residues that bind the indole group (F22, L100, Y102) and, more notably, residues which are essential for catalysis (E49 and D60) are located in this topologically-frustrated/backtracking off-pathway β_1 – β_3 region. However, other residues which bind the indole and the phosphate of the substrate lie outside this region (L127, A129, I153, Y175, S235). Thus, α TS appears to have a higher correlation than sIGPS between topological frustrated regions and active site residues.

A recent computational study of yeast triose phosphate isomerase (γ TIM) folding shows that the γ TIM Gō-model folds without off-pathway intermediates and in a very similar manner to experimental exchange studies.⁴⁶ A small degree of backtracking was noted within the final stages of folding (Stage D to E) at select sites between residues 25–110. However, none of the three conserved active site residues (K12, H95, E165) was shown to backtrack. Thus, like

sIGPS, γ TIM does not have high correlation between active site residues and energetic/topological frustration.

Recent computational studies with members of the β -trefoil protein class have supported the idea that conservation of functional residues imposes topological frustration in the folding of these proteins.^{47,48} Why is this correlation not found in TIM barrels? Perhaps the reason lies in the fact that the β -trefoil proteins studied are not enzymes but have a signaling function involving binding to other proteins.^{47,48} Since binding interfaces between proteins generally require a much larger surface area than active sites, they involve more residues. Thus, binding proteins involve constraints on many residues while enzymes, like TIM barrels, only need to conserve a few key active site residues. For TIM barrels, it appears that there is sufficient freedom for evolutionary pressure to select residues, neighboring the active site residues, that minimize the local energetic and topological frustration.

Topology vs sequence in the folding of $(\beta\alpha)_8$ barrel proteins

Based upon the revised folding mechanism for sIGPS shown in Scheme 2, it is apparent that the thermodynamic and kinetic folding mechanisms of three $(\beta\alpha)_8$ proteins are far better conserved than their amino acid sequences. With pair-wise sequence identities of less than 13%,²² α TS, IOLI, and sIGPS, display several common features on their folding free energy surfaces. All have stable intermediates that retain ~50% of the far-UV CD ellipticity and whose conversion to the native state is one of the rate-limiting steps in folding. In the case of sIGPS, a pair of stable intermediates is involved, I_a and I_b , and the interconversion between them is rate-limiting under strongly folding conditions. The revised kinetic mechanism for sIGPS is now consistent with those for α TS²¹ and IOLI²² in that all three initially misfold within milliseconds to an off-pathway species that is also rich in secondary structure. The unfolding of the I_{bp} species controls access to the subsequent on-pathway equilibrium intermediate via the unfolded state. The most distinguishing differences between the folding mechanisms involves α TS, where cis/trans isomerization reactions at P28, P217 and P261⁴⁹ lead to parallel folding channels, and IOLI, where only a single stable intermediate has been detected at moderate protein concentrations. As for sIGPS and α TS, it is possible that a second equilibrium intermediate for IOLI exists and might be detected with an appropriate probe. Only a fortuitously placed histidine, H92, is sensitive to the presence of the I2 intermediate in α TS.⁵⁰ Thus, with the exception of the unique role for three prolines in α TS, the thermodynamic and kinetic folding mechanisms of sIGPS, α TS and IOLI are remarkably similar. The common topology of these three $(\beta\alpha)_8$ barrels is evidently more important than their respective sequences in defining the distinctive features of their folding free energy surfaces. The same conclusion can be drawn from the good agreement between the predicted protection against HX from the sequence-blind G δ -model simulation and the observed results for the equilibrium on-pathway intermediates for sIGPS.

The conservation of the folding free energy surface, however, does not extend to the structures of the stable intermediates that populate this surface at moderate urea concentrations. A previous comparison of the HX protection patterns offered by the I_a and I_b intermediates from sIGPS and the I1 intermediate from α TS^{23,24} revealed that the secondary structures are not identical. The I1 species in α TS displays strong protection in the $(\beta\alpha)_{1-4}$ segment, while the composite I_a and I_b species in sIGPS display strong protection in the $\beta_2(\beta\alpha)_{3-5}\beta_6$ and the $\alpha_1(\beta\alpha)_{2-7}$ segments, respectively. In both cases, the protection patterns correlate with the presence of large hydrophobic clusters dominated by the branched aliphatic side chains from isoleucine, leucine and valine residues.^{24,51} Thus, as the sequences of these $(\beta\alpha)_8$ barrels evolved, the location of the structured regions defining the equilibrium intermediate moved with the location of the ILV-dominated hydrophobic clusters.

Why are the on- and off-pathway intermediates for $(\beta\alpha)_8$ barrel proteins conserved?

With billions of years to evolve sequences that fold as efficiently as possible, i.e., via a two-state mechanism that avoids off- and on-pathway intermediates, why are both types of intermediates preserved in the three $(\beta\alpha)_8$ barrels examined thus far? The sub-millisecond misfolded species would not interfere with folding during synthesis because it would appear and dissipate far faster than the rate of synthesis on a ribosome at 10–20 amino acids per second^{52,53} or the subsequent folding reactions. Its conservation might reflect the $(\beta\alpha)_n$ -repeat architecture of these barrels and, thereby, the opportunity for local and rapid folding reactions to forms of sufficient stability to be transiently populated during folding. As has been previously speculated,⁵¹ this structure may not be able to propagate to the remainder of the sequence because it adopts a non-native local fold or because its edges adopt a non-native fold that enhances its interactions with solvent before the propagation reaction can occur.

The equilibrium intermediate would, however, be a candidate for evolutionary pressure because it persists for 10's to 100's of seconds for this set of barrels. An extensive mutational analysis of the equilibrium intermediate for α TS^{54,55} demonstrated that a tightly-packed N-terminal region, $(\beta\alpha)_{1-4}$, is not tightly coupled to a molten globule-like C-terminal region containing $(\beta\alpha)_{5-8}$. This differential behavior was attributed to misfolding at the boundaries of the $(\beta\alpha)_{1-4}$ region that precluded propagation to the C-terminal region in the I1 intermediate. The possibility of misfolding at the periphery of an N-terminal core is supported by the results of another mutational analysis of a pair of long-range H-bonds in the N-terminal region between aspartic acid side chains at the C-termini of two helices and main chain amide hydrogens near the N-termini of their preceding β -strands.⁵⁶ These $\beta\alpha$ -hairpin clamps are disrupted in the I1 intermediate. However, the very strong protection of these amide hydrogens against exchange, > weeks-months at pH 7.8 and 25 °C, implies that both have found alternative partners in the I1 state and only exchange through other very high energy intermediates. Extending the argument presented for the I_{bp} species, the existence of partially-misfolded equilibrium intermediates might reflect the $(\beta\alpha)$ -repeat architecture and the opportunity for regional folding. The core of the structure in the equilibrium intermediates might be sufficiently similar to that for the native conformation that natural variations in the sequence which might destabilize and, thereby, reduce the population of the intermediate might also destabilize the native conformation. Precisely this response has been observed previously for α TS.^{49,54}

This hypothesis is supported by the good agreement between the HX protection pattern of the on-pathway equilibrium intermediates for sIGPS and the predictions of the kinetic Gō-model simulations. The Gō-model, which does not contain sequence information and only accounts for native-like interactions between C α positions, was surprisingly accurate at predicting the location of the structure in the composite I_a and I_b intermediates in sIGPS. By implication, the core of structure in these species is very similar to that for the native conformation. The discrepancies at the boundaries might reflect non-native structure in the actual protein that can not be detected by the Gō-model simulation. The generally good agreement between experiment and prediction implies, as was the case for the conservation of the mechanism for diverse sequences, that the topology of the $(\beta\alpha)_8$ barrel is more important than the sequence in determining the basic features of the folding free energy surface.

Although the $(\beta\alpha)_8$ barrels are symmetrical to a first approximation, the prediction of structure in a subset of the $\beta\alpha$ -repeat units for both sIPGS and α TS³² suggests that the sequences do influence the positioning and/or packing of the underlying main-chains. The correlation between the protection against HX and the predicted structure in the intermediate may reflect the enhanced exclusion of solvent from peptide bonds under isoleucine, leucine and valines. The resulting decrease in the volume of the peptide bond⁵⁷ could both retard HX and draw the side chains closer together in their resident hydrophobic cluster. In turn, the contraction of the cluster would more effectively exclude solvent from the peptide bond and strengthen its

associated H-bond to another main chain peptide bond. The synergy between the tertiary structure defined by the hydrophobic cluster and the secondary structure defined by the β -barrel strands could serve as the source of the cooperativity in the stable folding intermediates of $(\beta\alpha)_8$ barrel proteins.

Materials and Methods

Reagents

Ultrapure urea was purchased from ICN Biomedicals, Inc. (Auroa, OH). Pepsin immobilized on cross-linked 6% agarose gel was obtained from Pierce Chemicals (Rockford, IL). Deuterium oxide (99.9%) was obtained from Sigma-Aldrich. The α -cyano-4-hydroxycinnamic acid matrix for MALDI-TOF mass spectrometry was obtained from Aldrich Chemicals, and trifluoroacetic acid (peptide synthesis grade) and acetonitrile (HPLC grade) were obtained from Fisher Scientific (Pittsburgh, PA).

Preparation of samples pulse-labeled by manual mixing methods for mass spectrometric analysis

The expression and purification of sIGPS were performed as described previously.²⁴ The refolding of urea-denatured sIGPS (incubated in 8 M urea containing a 10 mM phosphate buffer, pH 7.8, overnight at 25 °C) was initiated by rapid 11-fold dilution in refolding buffer (10 mM phosphate at pH 7.8 and 25 °C). After 5 s of refolding, the sample was pulse-labeled with deuterated buffer for 5 s and then quenched with trifluoroacetic acid (TFA) to pH 2.7, as described previously.²⁴ A non-deuterated sample (the folded reference) for determining the percent protection in sIGPS and the peptic fragments was prepared in the same fashion by exposing the native protein to the protonated buffer for 5 s in the absence of urea. The unfolded reference sample was prepared by incubating sIGPS in 8 M D₄-urea and deuterated buffer overnight at pH 7.8 and 25 °C.

Pulse-labeling of burst-phase species for mass spectrometric analysis by quenched-flow mixing

Quenched-flow hydrogen-exchange was performed according to previously published procedures^{58–61} using a Biologic SFM-400/Q for rapid mixing and quenching. The fully deuterated unfolded protein was prepared by dissolving lyophilized sIGPS in 8 M D₄-urea at pH 7.8 for at least 24 h at room temperature. The refolding at 25 °C was initiated by rapid mixing of the unfolded and fully-deuterated sIGPS (in syringe 1) with 10 volumes of refolding buffer (10 mM phosphate buffer in H₂O at pH 7.8) in syringe 2. At different refolding time intervals, the pH was increased to 9.5 by adding an appropriate volume of the solution in syringe 3 (KOH solution, pH 11) for 50 ms to label unfolded regions with protium. Exchange was then quenched by decreasing the pH to 2.7 with an ice-cold quench solution in syringe 4 (0.05% TFA, 2% acetonitrile). The collected sample was then immediately added to the immobilized pepsin gel on ice for digestion as described previously²⁴. The folded and unfolded reference for the burst-phase species was obtained by digestion of the fully deuterated and protonated sIGPS, respectively, prepared with the same quenching method (without the refolding and labeling steps) in the SFM-400/Q.

Validation of quenched-flow results

To test the validity of the results from the quenched-flow labeling experiment on the burst-phase species, a sample of sIGPS in 5 M urea was labeled for 50 ms at pH 9.5 for comparison with previous results from labeling the equilibrium intermediates for 5 s at pH 7.8. The similar pulse intensities were expected to yield the same protection patterns and percent protections if the quenched-flow experiment is an accurate probe of the protection against HX. The fully

deuterated sIGPS was equilibrated in 5 M D₄-urea, the pH was increased to 9.5 for 50 ms and the sample was quenched to pH 2.7 with ice-cold TFA in the quenched-flow apparatus.

Pepsin digestion and MALDI-TOF MS analysis

The procedures for pepsin digestion, identification and sample preparation for MALDI-TOF mass spectrometric analysis have been described previously.²⁴ Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra of peptide fragments were obtained on a Micromass M@LDI-LR instrument in the reflectron mode with the same parameters as previously described.²⁴

Analysis of spectral data from MALDI-TOF MS

The percent protection for peptides was calculated according to Equation 1.

$$\text{Present protection} = 100 \times (M - M_{\text{unfolded ref}}) / (M_{\text{folded reference}} - M_{\text{unfolded reference}}) \quad 1$$

For the manual mixing experiment, the definition of the parameters in equation (1) has been described previously²⁴. For the quenched-flow study, $M_{\text{unfolded ref}}$ is the mass of the peptide derived from the native state protein, and $M_{\text{folded ref}}$ is the mass of the peptide derived from fully-exchanged protein. The mass was measured by determining the centroid of the isotopic envelope with an algorithm provided by the manufacturer of the spectrometer. This approach has previously been used to determine the relative percentage protection of peptides after HD exchange.^{23,24,62}

Kinetic Refolding Simulations of the sIGPS Gō-model

The sIGPS Gō-model was designed on the 2c3z PDB structure, an IGPS construct lacking the N-terminal 26 residues. The kinetic folding simulations were conducted as described previously³² with the following differences. Molecular dynamics simulations were performed as kinetic refolding of sIGPS, where protein chains, unfolded at high temperature, are allowed to refold at lower temperature.

The initial coordinates used for MD simulations of γ TIM were obtained from simulated annealing, using the 2c3z protein data bank (PDB) coordinates as an initial structure. The starting coordinates of each refolding trajectory are obtained from the final structure of a short simulation at 999K of a randomly determined length (500–1500 psec) and random initial velocities, followed by 100 psec at 400K and random initial velocities. These starting 400K coordinates are immediately placed into a 340K simulation temperature with random initial velocities and run for 30 nsec (20×10^6 time steps).

Simulation Analysis

Throughout the entirety of each 45 nsec folding simulation, 30,000 structures were obtained (1 structure every 1.5 psec). For each structure sampled throughout the simulations, the total number of native contacts (Q) formed was calculated where each native contact was determined to be formed if it falls within 1.5 times the native distance. The maximum number of tertiary native contacts (folding contacts within the monomer) possible for any sIGPS protein conformation is 560.

To identify the structures of intermediate stages of folding, the simulation structures were grouped by similar values of Q contacts. Group “A” consists of structures with Q between 0–100; Group “B” with Q between 101–200; Group “C” with Q between 201–300; Group “D” with Q between 301–400; Group “E” with Q between 401–500; and Group “F” with Q between

501–560. Within groups A–F, the probability that each residue i is folded is calculated through equations 2a and 2b:

$$\text{ProbabilityFolded}(i) = \frac{\sum_{n=1}^N \sum_{t=1}^T \sum_{p=0}^P \delta_{D(i,n,t,p), D(i,X,p)}}{NTP} \quad 2a$$

Where

$$\delta_{jk} = \begin{cases} 1, & \text{if } D(i, n, t, p) \leq 1.5 * D(i, X, p) \\ 0, & \text{if } D(i, n, t, p) > 1.5 * D(i, X, p) \end{cases} \quad 2b$$

In equation 2a and 2b, N is the total number of simulations in each group (10), T is the length of each simulation (30,000), P is the total number of contact pairs involving residue i (varies from 0–16), $D(i, n, t, p)$ is the distance of C_α atoms of residue i and contact partner, defined by p , in the simulation at a given set of n, t values, and $D(i, X, p)$ is the distance of C_α atoms of residue i and contact partner, defined by p , in the reduced PDB structure 2c3z (denoted as X). For residues in siGPS, i was varied between 1–222.

To estimate the percent protection, i.e., the percent of folding, of each residue within the intermediate stages in the simulations, equation 3 was used.

$$\text{Present protection} = 100 \times (P - P_{\text{unfolded ref}}) / (P_{\text{folded ref}} - P_{\text{unfolded ref}}) \quad 3$$

Where P is the percent of folding of the residue averaged over all Q states in one of the intermediate folding stages (B, C, D, or E), $P_{\text{unfolded ref}}$ is the percent of folding of the residue averaged over all Q states in stage A, and $P_{\text{folded ref}}$ is the percent of folding of the residue averaged over all Q states in stage F.

For each simulation condition studied, the probability that the simulation structure has a value of $Q = i$ was computed through equations 4a and 4b:

$$\text{Probability}(Q = i) = \frac{\sum_{n=1}^N \sum_{t=1}^T \delta_{Q(n,t), i}}{NT} \quad 4a$$

Where

$$\delta_{Q(n,t), i} = \begin{cases} 1, & \text{if } Q(n, t) = i \\ 0, & \text{if } Q(n, t) \neq i \end{cases} \quad 4b$$

In equations 4a and 4b, N is the total number of simulations in each group (10) and T is the length of each simulation (30,000). For native contacts, i was varied between 0–560.

To obtain statistically significant results, 30 refolding trajectories were collected. The average properties of each group of 10 trajectories, 1–10, 11–20, and 21–30, were calculated separately and the three average values used to determine the overall average and standard deviation of each measured value presented from the simulations in this work.

Abbreviations

α TS, alpha subunit of tryptophan synthase from *E. coli*
 CD, circular dichroism
 HD, hydrogen-deuterium
 HX-MS, hydrogen-deuterium exchange mass spectrometry
 I_a and I_b , partially-folded equilibrium intermediates in sIGPS
 I_{bp} , off-pathway sub-millisecond kinetic intermediate of sIGPS
 IOLI, a TIM barrel protein of unknown function from yeast
 MALDI, matrix-assisted laser desorption ionization
 D_4 -urea, fully deuterated urea
 2H_2O , deuterated water
 MD, molecular dynamics
 MS, mass spectrometry
 N, native state
 sIGPS, indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*
 TIM, triose phosphate isomerase
 TFA, trifluoroacetic acid
 U, unfolded state

Acknowledgements

This work was supported by N.I.H. grant GM23303 to C.R.M. J.M.F. is the recipient of a Dreyfus Start-Up Award from the Camille and Henry Dreyfus Foundation. We are grateful to J. Evans, B. Evans, K. Green and J. Patel of the Proteomics and Mass Spectrometry Core Facility at the University of Massachusetts Medical School for training and use of the mass spectrometry instruments. We thank S. Kathuria for assistance with the analysis of the chevron data, and we also thank Drs. R. Simler, R. Vadrevu, J.A. Zitzewitz and Y. Wu for many insightful discussions.

References

- Bai Y, Karimi A, Dyson HJ, Wright PE. Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci* 1997;6:1449–1457. [PubMed: 9232646]
- Jackson SE. How do small single-domain proteins fold? *Fold Des* 1998;3:R81–R91. [PubMed: 9710577]
- Ballew RM, Sabelko J, Gruebele M. Observation of distinct nanosecond and microsecond protein folding events. *Nat Struct Biol* 1996;3:923–926. [PubMed: 8901868]
- Ballew RM, Sabelko J, Gruebele M. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc Natl Acad Sci U S A* 1996;93:5759–5764. [PubMed: 8650166]
- Forsyth WR, Matthews CR. Folding Mechanism of Indole-3-glycerol Phosphate Synthase from *Sulfolobus solfataricus*: A Test of the Conservation of Folding Mechanisms Hypothesis in $(\beta\alpha)_8$ Barrels. *J Mol Biol* 2002;320:1119–1133. [PubMed: 12126630]
- Qi PX, Sosnick TR, Englander SW. The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nat Struct Biol* 1998;5:882–884. [PubMed: 9783747]
- Sosnick TR, Shtilerman MD, Mayne L, Englander SW. Ultrafast signals in protein folding and the polypeptide contracted state. *Proc Natl Acad Sci U S A* 1997;94:8545–8550. [PubMed: 9238013]
- Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 1993;262:892–896. [PubMed: 8235610]
- Sauder JM, Roder H. Amide protection in an early folding intermediate of cytochrome c. *Fold Des* 1998;3:293–301. [PubMed: 9710575]
- Khorasanizadeh S, Peters ID, Roder H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* 1996;3:193–205. [PubMed: 8564547]
- Sosnick TR, Mayne L, Hiller R, Englander SW. The barriers in protein folding. *Nat Struct Biol* 1994;1:149–156. [PubMed: 7656032]

12. Tezcan FA, Findley WM, Crane BR, Ross SA, Lyubovitsky JG, Gray HB, Winkler JR. Using deeply trapped intermediates to map the cytochrome c folding landscape. *Proc Natl Acad Sci U S A* 2002;99:8626–8630. [PubMed: 12084923]
13. Walkenhorst WF, Green SM, Roder H. Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry* 1997;36:5795–5805. [PubMed: 9153420]
14. Oliveberg M, Tan YJ, Silow M, Fersht AR. The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J Mol Biol* 1998;277:933–943. [PubMed: 9545382]
15. Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. *Protein Sci* 2000;9:452–465. [PubMed: 10752607]
16. Jones BE, Matthews CR. Early intermediates in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange and NMR. *Protein Sci* 1995;4:167–177. [PubMed: 7757007]
17. Chamberlain AK, Handel TM, Marqusee S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 1996;3:782–787. [PubMed: 8784352]
18. Sjogren T, Hajdu J. Structure of the bound dioxygen species in the cytochrome oxidase reaction of cytochrome cd1 nitrite reductase. *J Biol Chem* 2001;276:13072–13076. [PubMed: 11278884]
19. Farber GK, Petsko GA. The evolution of alpha/beta barrel enzymes. *Trends Biochem Sci* 1990;15:228–234. [PubMed: 2200166]
20. Gualfetti PJ, Bilsel O, Matthews CR. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci* 1999;8:1623–1635. [PubMed: 10452606]
21. Bilsel O, Zitzewitz JA, Bowers KE, Matthews CR. Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* 1999;38:1018–1029. [PubMed: 9893998]
22. Forsyth WR, Bilsel O, Gu Z, Matthews CR. Topology and sequence in the folding of a TIM barrel protein: Global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a ($\beta\alpha$)₈ barrel of unknown function from *B. subtilis*. *J Mol Biol*. In press
23. Rojsajjakul T, Wintode P, Vadrevu R, Robert Matthews C, Smith DL. Multi-state unfolding of the alpha subunit of tryptophan synthase, a TIM barrel protein: insights into the secondary structure of the stable equilibrium intermediates by hydrogen exchange mass spectrometry. *J Mol Biol* 2004;341:241–253. [PubMed: 15312776]
24. Gu Z, Zitzewitz JA, Matthews CR. Mapping the Structure of Folding Cores in TIM Barrel Proteins by Hydrogen Exchange Mass Spectrometry: The Roles of Motif and Sequence for the Indole-3-glycerol Phosphate Synthase from *Sulfolobus solfataricus*. *J Mol Biol* 2007;368:582–594. [PubMed: 17359995]
25. Sanchez del Pino MM, Fersht AR. Nonsequential unfolding of the alpha/beta barrel protein indole-3-glycerol-phosphate synthase. *Biochemistry* 1997;36:5560–5565. [PubMed: 9154940]
26. Andreotti G, Tutino ML, Sannia G, Marino G, Cubellis MV. Indole-3-glycerol-phosphate synthase from *Sulfolobus solfataricus* as a model for studying thermostable TIM-barrel enzymes. *Biochim Biophys Acta* 1994;1208:310–315. [PubMed: 7947963]
27. Englander SW, Kallenbach NR. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* 1983;16:521–655. [PubMed: 6204354]
28. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. Detection of transient protein folding populations by mass spectrometry. *Science* 1993;262:896–900. [PubMed: 8235611]
29. Englander SW, Poulsen A. Hydrogen-Tritium Exchange of the random chain polypeptide. *Biopolymers* 1969;7:379–393.
30. Efimova YM, Haemers S, Wierczinski B, Norde W, van Well AA. Stability of globular proteins in H₂O and D₂O. *Biopolymers* 2007;85:264–273. [PubMed: 17143859]
31. Cioni P, Strambini GB. Effect of heavy water on protein flexibility. *Biophys J* 2002;82:3246–3253. [PubMed: 12023248]

32. Finke JM, Onuchic JN. Equilibrium and kinetic folding pathways of a TIM barrel with a funneled energy landscape. *Biophys J* 2005;89:488–505. [PubMed: 15833999]
33. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157:105–132. [PubMed: 7108955]
34. Wildegger G, Kiefhaber T. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* 1997;270:294–304. [PubMed: 9236130]
35. Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* 1992;89:8721–8725. [PubMed: 1528885]
36. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding--a perspective from simple exact models. *Protein Sci* 1995;4:561–602. [PubMed: 7613459]
37. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* 1987;84:7524–7528. [PubMed: 3478708]
38. Karplus M, Sali A. Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol* 1995;5:58–73. [PubMed: 7773748]
39. Camacho CJ, Thirumalai D. Denaturants can accelerate folding rates in a class of globular proteins. *Protein Sci* 1996;5:1826–1832. [PubMed: 8880906]
40. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys J* 2002;83:3525–3532. [PubMed: 12496119]
41. Clementi C, Jennings PA, Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc Natl Acad Sci U S A* 2000;97:5871–5876. [PubMed: 10811910]
42. Shea JE, Onuchic JN, Brooks CL 3rd. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci U S A* 1999;96:12512–12517. [PubMed: 10535953]
43. Go N. Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 1983;12:183–210. [PubMed: 6347038]
44. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545–600. [PubMed: 9348663]
45. Tsuji T, Chrnyk BA, Chen X, Matthews CR. Mutagenic analysis of the interior packing of an alpha/beta barrel protein. Effects on the stabilities and rates of interconversion of the native and partially folded forms of the alpha subunit of tryptophan synthase. *Biochemistry* 1993;32:5566–5575. [PubMed: 8504078]
46. Patel B, Finke JM. Folding and unfolding of γ TIM monomers and dimers. *Biophys J* 2007;93
47. Chavez LL, Gosavi S, Jennings PA, Onuchic JN. Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc Natl Acad Sci U S A* 2006;103:10254–10258. [PubMed: 16801558]
48. Gosavi S, Chavez LL, Jennings PA, Onuchic JN. Topological frustration and the folding of interleukin-1 beta. *J Mol Biol* 2006;357:986–996. [PubMed: 16469330]
49. Wu Y, Matthews CR. Parallel channels and rate-limiting steps in complex protein folding reactions: prolyl isomerization and the alpha subunit of Trp synthase, a TIM barrel protein. *J Mol Biol* 2002;323:309–325. [PubMed: 12381323]
50. Saab-Rincon G, Froebe CL, Matthews CR. Urea-induced unfolding of the alpha subunit of tryptophan synthase: one-dimensional proton NMR evidence for residual structure near histidine-92 at high denaturant concentration. *Biochemistry* 1993;32:13981–13990. [PubMed: 8268176]
51. Wu Y, Vadrevu R, Kathuria S, Yang X, Matthews CR. A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol* 2007;366:1624–1638. [PubMed: 17222865]
52. Kjeldgaard, NO.; Gausing, K. Regulation of biosynthesis of ribosomes. In: Nomura, M.; Tissieres, A.; Lengyel, P., editors. *Ribosomes*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1974. p. 369-392.
53. Kennell D, Riezman H. Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J Mol Biol* 1977;114:1–21. [PubMed: 409848]

54. Wu Y, Matthews CR. Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of Trp synthase, a TIM barrel protein. *J Mol Biol* 2003;330:1131–1144. [PubMed: 12860133]
55. Wu YF, Vadrevu R, Yang X, Matthews CR. Specific structure appears at the N terminus in the sub-millisecond folding intermediate of the alpha subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol* 2005;351:445–452. [PubMed: 16023136]
56. Yang X, Vadrevu R, Wu Y, Matthews CR. Long-range Side chain-Backbone Interactions Play Crucial Roles in Stabilizing the ($\beta\alpha$)₈ Barrel Motif of the Alpha Subunit of Tryptophan Synthase. *Protein Sci.* In press
57. Schell D, Tsai J, Scholtz JM, Pace CN. Hydrogen bonding increases packing density in the protein interior. *Proteins* 2006;63:278–282. [PubMed: 16353166]
58. Yang H, Smith DL. Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry* 1997;36:14992–14999. [PubMed: 9398224]
59. Udgaonkar JB, Baldwin RL. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 1988;335:694–699. [PubMed: 2845278]
60. Mullins LS, Pace CN, Raushel FM. Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange-two-dimensional NMR spectroscopy. *Biochemistry* 1993;32:6152–6156. [PubMed: 8512924]
61. Roder H, Elove GA, Englander SW. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 1988;335:700–704. [PubMed: 2845279]
62. Simler BR, Levy Y, Onuchic JN, Matthews CR. The folding energy landscape of the dimerization domain of Escherichia coli Trp repressor: a joint experimental and theoretical investigation. *J Mol Biol* 2006;363:262–278. [PubMed: 16956620]

Figure 1.

The peptide map of sIGPS for HX-MS analysis. The elements of secondary structure and the 19 peptides resulting from pepsin digestion at pH 2.7 are displayed on the amino acid sequence of IGPS from *S. sulfataricus*. The α-helices are shown in magenta, and the β-strands in blue above the sequence, and the peptides are shown as black lines below the sequence. This figure was taken from Gu et al., (2007)²⁴ with minor modifications.

Figure 2.

Representative MALDI-TOF spectra of four peptides displaying different types of behavior over the first few hundred milliseconds of folding of sIGPS: (A) peptide 36–47; protection appears to decrease with time (B) peptide 100–115; protection increases with time (C) peptide 143–151; protection is largely independent of time and (D) peptide 197–203; no protection is apparent at any of these folding times. The spectra obtained at various refolding time intervals prior to quenched-flow labeling and both the folded and unfolded reference spectra for each peptide are shown in each panel. The refolding reaction was initiated from 8.0 M urea by 11 times dilution to 10 mM potassium phosphate buffer, pH 7.8, at 25 °C.

Figure 3. Percent protection against exchange of amide hydrogens for solvent deuterium for 19 peptides of sIGSP derived from pepsin digestion at pH 2.7 and 0 °C as a function of refolding time. The estimated errors are $\pm 5\%$.

Figure 4.

Percent protection for 18 of the 19 peptic peptides derived from sIGPS for quenched-flow labeling of the burst-phase intermediate with a 50 ms pulse at pH 9.5 after 75 ms of folding (green), manual-mixing labeling of the equilibrium intermediates, I_a and I_b , at 5 M urea with a 5 s pulse of deuterated buffer at pH 7.8²⁴ (black), manual-mixing labeling of the I_a intermediate highly-populated after 5 s of folding with a 5 s pulse of deuterated buffer at pH 7.8 (red), quenched-flow labeling of the equilibrium intermediates, I_a and I_b , at 5 M D_4 -urea with a 50 ms pulse of protonated buffer at pH 9.5 (blue) and Gō-model simulation (cyan). The data for peptide 59–68 were not included because this peptide does not display protection against HX under any conditions.

Figure 5.

(A) The fluorescence-detected refolding reaction of sIGPS via stopped-flow mixing from 8 M to 1.2 M urea at pH 7.8 and 25 °C. The trace was best fit by three exponentials with time constants and relative amplitudes of 0.6 s and -47%, 7.3 s and 28% and 80 s and 72%; the amplitudes were normalized to the sum of the two phases of increasing intensity. The residuals are shown in the lower part of panel A. The data density was reduced by 10-fold to illustrate the quality of the fit. (B) A chevron plot of the refolding (open symbols) and unfolding (filled symbols) time constants as a function of the final urea concentration. Stopped-flow fluorescence detected phases have symbols with cross-marks inside. The CD data (symbols without cross-marks) are taken from Figure 7 in Forsyth et al., (2002)⁵. The assignments of

the time constants to individual steps in the folding mechanism, Scheme 2, are also shown. The unfolding reactions were initiated in 0 M urea and ended at the indicated urea concentrations; the refolding jumps were initiated at 8 M urea and ended at the indicated urea concentrations. The buffer in all experiments contained 10 mM phosphate buffer at pH 7.8 and 25 °C.

Figure 6.

Implied structures and protection patterns for the off-pathway I_{bp} specie at 75 ms, on-pathway I_a (5s) and I_b 24 species are displayed on ribbon diagrams along with the folding scheme of sIGPS. Regions with strong percentage protection, $\geq 40\%$, are shown in blue, regions with moderate percentage protection, $10\% < \text{protection} < 40\%$ are shown in green, regions with no significant protection, $\leq 10\%$, are shown in red, the disordered region (peptide 59–68) are shown in grey, and regions with no peptide coverage are shown in orange. For reference, the unfolded and native structures are shown in all red and all blue respectively.

Figure 7.

Gō-model simulation trajectory of sIGPS folding. (A) Number of native contacts, Q , at various time steps in folding for a fast productive trajectory (black), slow productive trajectory (green), and an unproductive trajectory (yellow). (B) Probability of refolding trajectories sampling each value of Q . The grey lines indicate the one standard deviation level of uncertainty.

Figure 8.

Structural profiles of all 30 sIGPS simulation trajectories in folding stages A–F, displayed as the fraction folded in each of the peptic fragments. Trajectories with a bias of folded contacts in the center are highlighted in black, trajectories with a bias of folded contacts at the N-terminus are highlighted in red and, trajectories with a bias of folded contacts at the C-terminus are highlighted in blue. The 7 unproductive trajectories which did not reach the native state within the time of the simulation are highlighted in yellow.

Figure 9. Average fraction folded of the 23 productive trajectories in folding stages B (orange), C (blue), D (green) and E (black). The fraction folded was normalized such that stage A (unfolded) values were set to 0 and stage F values (folded) were set to 1. Errors indicate standard deviations. The locations of the 6 primary active site residues are highlighted with purple vertical lines.

Figure 10.

Contact map of the 2c3z sIGPS structure displaying each of the 571 contacts (black squares). Backtracking contacts, which decrease in probability as the overall Q progresses from folding stages D to E, are shown as yellow squares. For reference, secondary structure elements are shown along each axis.

scheme 1.

scheme 2.

Table 1Percent protection of sIGPS peptides derived from transient and stable folding intermediates^a

Peptide	Secondary Structure ^b	Percentage Protection ^c		
		Labeling Time for Transient Intermediate ^d		Equilibrium Intermediates ^e
		75 ms	5 sec	5 M Urea
1–8		8	0	8
9–15	α_0	3	10	11
24–35	β_1	13	0	4
36–47	α_1	26	11	23
48–55	$(\alpha_1)/\beta_2$	12	49	71
59–68	disordered	0	0	0
69–83	α_2/β_3	16	60	49
87–97	α_3	26	89	88
100–115	$\beta_4/(\alpha_4)$	43	71	54
107–120	β_4/α_4	60	88	83
109–121	α_4	65	90	78
123–134	α_4/β_5	16	85	80
135–141	β_5/α_5	32	89	70
143–151	α_5/β_6	20	55	61
146–159	$(\alpha_5)/\beta_6$	13	37	49
162–179	α_6	3	5	27
180–196	β_7/α_7	5	2	25
197–203	α_7	0	0	0
211–220	α_8	0	0	0

^aThe labeling procedures are described in Materials and Methods.^bParenthesis indicates a small contribution from the element of secondary structure.^cThe estimated errors are $\pm 5\%$.^dRefolding was initiated by dilution from 8 M urea to a final concentration of 0.72 M urea in 10 mM phosphate buffer, pH 7.8 and 25°C.^eData taken from Gu et al.²⁴