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 f...

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MAY 2007

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2007.02.027 · Source: PubMed

3 AUTHORS:

Zhenyu Gu
University of Massachusetts Medical School
8 PUBLICATIONS **319** CITATIONS
SEE PROFILE

Jill A Zitzewitz
University of Massachusetts Medical School
21 PUBLICATIONS **997** CITATIONS
SEE PROFILE

Charles Robert Matthews
University of Massachusetts Medical School
146 PUBLICATIONS **5,902** CITATIONS
SEE PROFILE

Published in final edited form as:
J Mol Biol. 2007 April 27; 368(2): 582–594.

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 *S. solfataricus*

Zhenyu Gu, **Jill A. Zitzewitz**, and **C. Robert Matthews***
Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

Abstract

To test the roles of motif and amino acid sequence in the folding mechanisms of TIM barrel proteins, hydrogen-deuterium exchange was used to explore the structure of the stable folding intermediate for the of indole-3-glycerol phosphate synthase from *S. solfataricus* (sIGPS). Previous circular dichroism and fluorescence studies of the urea denaturation of sIGPS revealed the presence of an intermediate that is highly populated at ~4.5 M urea and contains ~50% of the secondary structure of the native state. Kinetic studies showed that this apparent equilibrium intermediate is actually comprised of two thermodynamically-distinct species, I_a and I_b. To probe the location of the secondary structure in this pair of stable on-pathway intermediates, the equilibrium unfolding process of sIGPS was monitored by hydrogen-deuterium (HD) exchange mass spectrometry. The intact protein and pepsin-digested fragments were studied at various urea concentrations by electrospray and MALDI-TOF mass spectrometry, respectively, after deuterium labeling and quenching in acid. Intact sIGPS strongly protects 54 amide hydrogens from HD exchange in the intermediate states, demonstrating the presence of stable folded cores. When the protection patterns and the exchange mechanisms for the peptides are considered with the proposed folding mechanism, the results can be interpreted to define the structural boundaries of I_a and I_b. Comparison of these results with previous HD exchange studies on another TIM barrel protein of low sequence identify, the alpha subunit of tryptophan synthase (αTS), indicates that the thermodynamic states corresponding to the folding intermediates are better conserved than their structures. Although the TIM barrel motif appears to define the basic features of the folding free energy surface, the structures of the partially-folded states that appear during the folding reaction depend on the amino acid sequence. Markedly, the good correlation between the HD exchange patterns of sIGPS and αTS with the locations of hydrophobic clusters defined by isoleucines, leucines, and valines suggests that branch aliphatic side chains play a critical role in defining the structures of the equilibrium intermediates.

Keywords

equilibrium intermediate; hydrophobic clusters; modular assembly; protein folding; thermodynamics of folding reactions

*Corresponding author. E-mail address: c.robert.matthews@umassmed.edu, Phone: (508) 856-2251; FAX: (508) 856-8358

Introduction

The $(\beta\alpha)_8$, TIM, barrel motif is one of the most common in biology, accounting for approximately 10% of the folds in all three super-kingdoms of life. Its capacity to catalyze five of six categories of biochemical reactions, including many essential metabolic reactions, implies that this motif arose early in evolution and flourished because of its catalytic adaptability.[1-4] Examination of the structures accessible to these βα-repeat proteins shows that, although pseudo-barrels with as few as 6 and expanded barrels with as many as 9 βα repeat units exist, the most common barrels contain 8 repeat units.[3] The repeats are sequentially-ordered into parallel-stranded arrays, with hydrogen bonds between strands 1 and 8 forming a cylindrical barrel. The anti-parallel amphipathic helices, alternating in the sequence with the strands, form a continuous shell around the largely nonpolar strands. The active sites are invariably defined by the loops at the C-termini of the strands, enabling the conservation of the structural core while offering a multitude of possibilities for substrate and cofactor recognition and for catalytic function.[5] In this sense, $(\beta\alpha)_8$ barrels are analogous to antibodies, where the anti-parallel β sandwich cores provide a stable platform for the somatic mutations in the loops that enable exquisite recognition of a plethora of antigenic determinants.

Although the $(\beta\alpha)_8$ motif is a single structural domain, its folding free energy surface is more complex than that expected for a simple, highly-cooperative unfolding transition to the denatured state. Studies on several $(\beta\alpha)_8$ barrel proteins, including the alpha subunit of tryptophan synthase from *E. coli*, αTS,[6] the indole-3-glycerol phosphate synthase from *S. solfataricus*, sIGPS,[7] and from *E. coli*, eIGPS,[8] and the N-(5′-phosphoribosyl)anthranilate isomerase from *E. coli*, PRAI,[9] have been shown to have stable intermediates with substantial thermodynamic stability that retain significant secondary structure (~50%) when probed by far-UV CD. Kinetic studies have shown that the folding of this intermediate to the native state can be the rate-limiting step in a complex reaction scheme.[7,9,10] Given the common occurrence of the equilibrium intermediate and its crucial role in folding, this species has been the subject of several fragmentation studies of its structure.

Early fragment complementation studies on αTS and PRAI were interpreted to mean that these proteins fold via a 6+2 model, with the $(\beta\alpha)_{1\text{-}6}$ folding nucleus corresponding to the equilibrium intermediate for both proteins.[11-13] However, a more comprehensive fragmentation analysis of αTS demonstrated that the equilibrium unfolding properties of this TIM barrel are more complex than revealed by the chemical denaturation experiment. The results led to the hypothesis that folding involves the modular assembly of βαβ supersecondary structural elements.[14] N-terminal fragments containing as few as 4 βα repeats can fold cooperatively. The addition of the $\beta_5\alpha_5\beta_6$ module confers enhanced stability, consistent with a refined (4+2) +2 model. A more recent characterization of a series of fragments and circularly permuted variants of PRAI, combined with molecular dynamics simulations, led to the proposal of a 5 +1+2 folding model in which the $(\beta\alpha)_{1\text{-}6}$ segment corresponds most closely to the equilibrium intermediate.[15] By contrast, the recovery of catalytic activity upon the complementation of the N- and C-terminal halves of the HisF barrel[16] and the retention of structure and stability in the C-terminal half barrel[17] support a 4+4 mechanism of folding. The $(\beta\alpha)_{5\text{-}8}$ segment would presumably serve as the platform defining the equilibrium intermediate observed in a chemical denaturation experiment (Z. Gu and C. R. Matthews, unpublished results). These fragmentation studies have provided strong evidence for the modular assembly of $(\beta\alpha)_8$ proteins. Relevant to the present study, the different models for assembly suggest that the structures of the corresponding intermediates are strongly influenced, if not controlled, by the amino acid sequence.

Hydrogen exchange mass spectrometry (HX-MS) has recently been used to obtain structural information on folding intermediates of several TIM barrel proteins, including αTS,[18] the

multimeric aldolase[19] and triose phosphate isomerase, TIM.[20] Brief exposure of transient or stable partially-folded forms to ^2H_2O, quenching with acid and digestion with pepsin enables the identification by MS of segments where the protection of amide hydrogens against solvent exchange implies structure. The most stable core of the equilibrium unfolding intermediate for αTS observed by HX-MS at 3 M urea is comprised of the $(\beta/\alpha)_{1\text{-}3}\beta_4$ segment, consistent with previous fragmentation studies and a (4+2)+2 folding model.[18] Although the dimeric rabbit TIM does not appear to populate a stable intermediate under equilibrium conditions, a partially-folded state involving the $(\beta\alpha)_{5\text{-}8}$ segment was observed in HX-MS studies of a late intermediate in the refolding reaction.[20] The results were interpreted in terms of a 4+4 C-terminus driven folding mechanism, similar to HisF.[16] Two rare partially-folded states in equilibrium with native dimeric yeast TIM were detected by a complementary technique, misincorporation proton-alkyl exchange (MPAX).[21] By monitoring the transient exposure of cysteines to water-soluble sulfhydryl reagents, the most stable core, i.e. segments resistant to modification, was shown to involve strands β_2-β_4. The presumed folding model is (1+(3)+2) +2 and is similar to that for αTS. The various combinations of βα modules observed to define the structures of partially-folded states in TIM barrel proteins show that the conservation of the discrete thermodynamic state for the intermediate detected by chemical denaturation has not been realized in the conservation of its structure nor, possibly, in its singularity.

To further explore the relationship between the motif, the amino acid sequence and the structure of the stable intermediate in $(\beta\alpha)_8$ barrels, the secondary structures of the pair of corresponding intermediates in sIGPS have been examined by HX-MS. The locations of the elements of secondary structure in the two equilibrium intermediates for sIGPS differ from that for the single species in αTS, consistent with the hypothesis that the structures of these intermediates are less conserved than their existence. Strikingly, the protection patterns in both sIGPS and αTS appear to correlate with the presence of hydrophobic clusters comprised of branched aliphatic side chains in the native conformation.

Results

Equilibrium unfolding reaction of sIGPS

A previous thermodynamic analysis of the urea-induced unfolding reaction of a truncated version of sIGPS, lacking the N-terminal 26 amino acids, revealed the presence of a single apparent intermediate that is highly populated at 4.5 M urea and retains ~50% of the ellipticity at 222 nm observed for the fully-folded TIM barrel.[7] This truncated version had been previously shown to avoid aggregation problems encountered in an earlier study of the folding properties of the full-length protein.[8,22] Kinetic studies of the folding reaction of the truncated protein, however, demonstrated that this intermediate is actually comprised of two thermodynamically-distinct, slowly-interconverting species, designated as I_a and I_b, of comparable stability (Scheme 1).[7]

Pulse-labeling HX-MS experiments on sIGPS were designed to probe the protection of backbone amide hydrogens from exchange with solvent under equilibrium conditions that vary from favoring the N state at 0 M urea, the I_a and I_b states at ~5 M urea and the U state at 8 M urea. By exposing sIGPS to a pulse of deuterated urea/^2H_2O for 5 s at pH 7.8 and 25 °C and then quenching exchange on ice to pH 2.7, backbone amide hydrogens whose protection factors against exchange are ~100-fold greater than that for solvent-exposed amide hydrogens can be detected. The mass spectra of the +25 charge state of pulse-labeled sIGPS at various urea concentrations are shown in Figure 1; the mass and abundance of each peak are summarized in Table 1.

The molecular mass for the 0% reference, which is the native state of sIGPS that has not been exposed to ^2H_2O, yielded the mass expected for the sequence of the truncated protein

(observed: 25,577 Da; calculated: 25,576.5 Da). The only peak detected at 0 and 2 M urea after the pulse-quench experiment (Figure 1), with a mass of ~25,640 daltons (Table 1), corresponds to the native state of sIGPS. The increase in the mass compared to the native state in H_2O, 62 daltons, reflects the exchange of backbone hydrogens that are weakly protected or unprotected from exchange during the 5 s pulse labeling with ^2H_2O at pH 7.8 and 25 °C. At 4 M urea, the native peak decreased in amplitude and a major new peak appeared at ~25,715 daltons, reflecting the presence of the stable intermediates, I_a and I_b. Apparently, the masses of the I_a and I_b species are sufficiently similar that individual peaks cannot be resolved by the ESI-MS technique. Above 4.5 M urea, the peak for the intermediates decreased in amplitude and a third peak, corresponding to the unfolded state, appeared at ~25,763 daltons.

The clear separation between the peaks for the native and the pair of intermediate states shows that the HD exchange reaction is governed by EX1 kinetics (Figure 1), i.e. the rate of refolding of the I_a and I_b species to the N state is slower than the exchange with solvent deuterium, ~100 s^{-1} under these conditions. This conclusion is consistent with the rate constants for the $I_a \rightarrow I_b$ and $I_b \rightarrow N$ reactions, which are both < 0.01 s^{-1} below 4.5 M urea.[7] The conversion of the intermediates to the unfolded state is less distinct and may have components of EX1 and EX2 mechanisms, the latter of which implies that refolding from U (or another less well-folded state) to the I states is faster than exchange. Although EX1 behavior precludes determination of thermodynamic parameters,[23] useful information about the extent of protection and the sites of protection in partially-folded forms can be obtained.

The number of amides protected in the native and pair of intermediate states was estimated from the molecular masses of N, I and U (Table 2). After correction for back-exchange, 85 amide hydrogens were observed to exchange with deuterium when sIGPS unfolded from N to I, and an additional 54 deuteriums were introduced when I unfolded to U (Table 2). Thus, the intermediate states, which are maximally (~50%) populated at ~4.5-5 M urea, were demonstrated to protect at least 54 and perhaps as many as 131 amide hydrogen atoms from exchange with solvent deuterium (Table 2). The upper limit, determined from the difference between the total number of exchangeable amide hydrogens in sIGPS and the number of hydrogens lost when N is converted to I, includes the amide hydrogens whose protection factors are less than ~100 at 5 M urea. The results indicate that a significant core of amide backbone hydrogens are folded in the intermediate states, and the regions responsible for this observed protection can be more precisely elucidated by examining the exchange behavior at the peptide level.

Identification of the protected regions in the folding intermediates of sIGPS by HD exchange and pepsin digestion

To determine the location of the amide hydrogens protected against exchange in the pair of stable intermediates in sIGPS, the pulse-labeled protein at 4 and 5 M urea was subjected to pepsin digestion at pH 2.7 on ice. The native and intermediate states are both well populated at 4 M urea, and the population of the native state is very small at 5 M urea. Comparison of the results for these two denaturant concentrations highlights the exchange behavior of the intermediate states. Peptic digestion resulted in 19 fragments that covered all of the secondary structural elements in sIGPS with the exception of β_8 (Figure 2). The digested samples were immediately spotted on an α-cyano-4-hydroxycinnamic acid matrix maintained on ice, dried under vacuum and analyzed by MALDI-TOF mass spectrometry to minimize back exchange.[24] Peptide labeling was also examined at 0 and 8 M urea to provide normalization controls and to examine the extent of protection provided by the native barrel conformation at 0 M urea.

With the exception of the 59-68 peptide, which corresponds to a disordered loop in the native conformation,[25] all of the other 18 peptides covering almost every element of secondary structure in sIGPS showed significant average protection against exchange in the absence of

urea (Table 3). The composite nature of the exchange reaction in peptides with multiple amide hydrogens limits the quantification to the average degree of protection relative to the native and unfolded states. The native barrel structure is evidently very stable in solution, consistent with the 8.3 kcal mol^{-1} difference in free energy between the native and intermediate states obtained from a urea-denaturation analysis.[7] Four different types of exchange behavior were observed for the labeled peptides derived from the pair of stable intermediates for sIGPS at 5 M urea where the native state is not populated (Figure 1 and Table 1). Examples of each type of behavior at 5 M urea are shown in Figure 3, and the deuterium levels in each peptide at 0, 4, 5 and 8 M urea are shown in Table 3. The first set, comprised primarily of peptides from the N- and C-termini (residues 1-8, 9-15, 24-35, 197-203 and 211-220), offered no significant average protection against exchange at 5 M urea (Figure 3a). As expected, the unstructured 59-68 peptide also did not offer protection against exchange, as was evident by the coincidence of the isotope envelope at all urea concentrations (data not shown). The second set, comprised of peptides 36-47, 162-179 and 180-196, corresponding to α_1, α_6 and β_7/α_7, respectively, offered ~25% average protection and exchanged by an EX1 mechanism (Figure 3b). The spectrum of a peptide exchanging via the EX1 mechanism displays two peaks, one at the mass expected for the protected peptide and another at the mass of its exchanged counterpart.[26] The third set, comprised of peptides 48-55, 143-151 and 146-159, corresponding to β_2, α_5 and β_6, offered higher extents of protection, ≥ 49%, and exchanged via an EX2 mechanism (Figure 3c). Mechanistically, these segments refold from an exchange-competent state more rapidly than their amide hydrogens can exchange with solvent.[23] The mass spectrum for a peptide that obeys this mechanism shows a single peak that is intermediate in mass to those of the fully-protected and fully-exchanged peaks.[26] The fourth set, comprised of a continuous string of peptides from residues 69-141, corresponding to α_2/β_2-β_5/α_5, also offered a higher percent protection, > 54%, but exchanged via an EX1 mechanism (Figure 3d). The mechanism of HD exchange for each peptide at 5 M urea is also shown in Table 3. Comparison with the results for the same peptides derived from the sample equilibrated at 4 M urea shows uniform enhancements in the percent protected that reflects the presence of the native state (Table 3).

Although it is clear from the data in Table 3 that the I_a and I_b intermediates for sIGPS experience significant protection against HD exchange in the region from α_1 to α_7, the complex response in terms of levels of average protection and mechanisms of exchange suggests that this region is not behaving as a single kinetic or thermodynamic unit. The interpretation of these results in terms of the species and properties of the folding mechanism shown in Scheme 1 reveals remarkable insights into the structures of I_a and I_b (see Discussion).

Discussion

Analysis of HD exchange results for sIPGS

The results of a previous thermodynamic and kinetic analysis of the folding of sIGPS demonstrated that the stable intermediate observed at equilibrium is actually comprised of two distinct states that interconvert via a slow, urea-independent rearrangement reaction (Scheme 1).[7] Because the $U \rightarrow I_a$, $I_a \rightarrow I_b$ and $I_b \rightarrow N$ folding reactions are all much slower at 5 M urea than the HD exchange time for solvent-exposed amide hydrogens,[7] the exchange behavior of the intact sIGPS and the derived peptic peptides is expected to obey the EX1 limit. The results displayed in Figure 1 and Table 3 are largely consistent with this expectation. However, the distinctly different levels of average protection observed at 5 M urea for the 36-47, 162-179 and 180-196 peptides, ~25%, and those for the all of the intervening peptides (with the exception of the disordered loop, peptide 59-68), >49% suggests that more than one kinetic opportunity for exchange is operative. This point is reinforced by the observation that three of the peptides at the boundaries of the more highly protected region, 48-55, 143-151 and 146-159, exchange via an EX2 mechanism. This complex mass spectrometry response, when considered

in terms of the folding mechanism (Scheme I), provides the first evidence for the structural differences between the thermodynamically-distinct I_a and I_b species.

The common lower extent of protection for the 36-47, 162-179 and 180-196 peptides, ~25%, likely reflects the behavior of that fraction of the population occupying the I_b state. That is, I_b represents at least 25% of the total population of the two intermediates in equilibrium at 5 M urea, and HD exchange through the I_a species is governed by an EX1 mechanism. This interpretation is consistent with a recent kinetic refolding HX-MS experiment (Z. Gu and C. R. Matthews, unpublished results), which showed no significant protection for peptides 36-47, 162-179 and 180-196 after refolding for 5 s, when only I_a is populated.[7] The higher level of protection observed for all peptides spanning residues 48 and 159 (but not the disordered loop peptide, 59-68) suggests that this region is folded in I_a and exchanges through less well-structured higher energy states. The boundaries of this region, peptides 48-55, 143-151 and 146-159, undergo rapid reversible localized folding reactions that enable exchange via an EX2 mechanism through higher energy states in the I_a manifold. The central portion, residues 69-141, is more stable and only exchanges globally through the unfolded state, U.

Structural implications for the stable folding intermediates

This complex HD exchange behavior can be understood in terms of the modular assembly of the native sIGPS TIM barrel under equilibrium folding conditions. The core of structure in the I_a species is comprised of the $\alpha_2(\beta\alpha)_{3\text{-}5}$ module. However, the partial protection observed in the boundary peptides, 69-83 and 135-141, and the EX2 behavior of the 143-151 peptide that also contains part of α_5 make it unclear as to whether α_2 and α_5 are firmly integrated into this structure. The neighboring β_2 and $\alpha_5\beta_6$ elements of secondary structure display EX2 behavior, consistent with the presumption that they fluctuate through disordered forms within the manifold of states corresponding to the I_a intermediate, $(\beta\alpha)_{2\text{-}5}\beta_6$. Finally, the conversion of I_a to I_b anneals α_1, α_6, β_7 and α_7 to $(\beta\alpha)_{2\text{-}5}\beta_6$ to produce $\alpha_1(\beta\alpha)_{1\text{-}7}$, an open form of the barrel lacking structure at the N- and C-termini. To a first approximation, the equilibrium folding reaction of sIGPS begins with a minimal $(\beta\alpha)_{3\text{-}4}\beta_5$ core with fluctuating β_2 and $\alpha_5\beta_6$ boundaries, I_a, integrates adjacent α_1, α_6, β_7 and α_7 elements through one rate-limiting kinetic step to form I_b and then, in a final rate-limiting step, closes the barrel to produce $\alpha_0(\beta\alpha)_{1\text{-}8}$. The implied sequential development of structure is portrayed on a ribbon diagram of sIGPS in Figure 4a.

Relationship between structures of folding intermediates and the sequence of sIGPS

The heterogeneous HX protection pattern observed for the pair of stable intermediates in sIGPS cannot be explained by a previous analysis in which the contributions of all 20 side chains to stability were considered.[27] Nonpolar side chains from all eight β-strands define global hydrophobic clusters located inside and outside the barrel; however, no basis for the selective protection observed in the present study was apparent. By contrast, a recent mutational analysis of the equilibrium folding intermediate for αTS suggested that large aliphatic side chains play a primary role in the stability of its equilibrium intermediate.[28] Those results led to the BASiC hypothesis, which proposes that large hydrophobic clusters formed primarily by **B**ranched **A**liphatic **Si**de **C**hains, i.e. isoleucine, leucine and valine, play crucial roles in guiding the early events in folding and in stabilizing the on-pathway equilibrium intermediate for αTS.[28,29] The high preponderance of these branched aliphatic side chains in all $(\beta\alpha)_8$ barrels,[30] including sIGPS where the ILV residues as a group comprise 29% of the amino acid sequence, raises the possibility that clusters of ILV side chains might also play important roles in defining the structures of I_a and I_b in sIGPS.

An examination of the crystal structure for the truncated version of sIGPS[25] shows four ILV clusters containing 54 of the 64 ILV residues (Figure 4b). These clusters are distinguished by

a network of contacts between their aliphatic side chains at a cut-off distance of 4.2 Å and by the near absence of contacts between these networks. This relatively short cut-off distance was chosen to reflect van der Waals interactions that might be expected to make greater contributions to stability. Cluster 1, containing 11 ILV residues, is located in the center of the barrel towards the N-termini of the strands and includes long-range contacts between α0 and the several of the β strands. This cluster has an average of 2.8 ILV side chains in contact with each ILV residue and has a contact surface area of 534 $Å^2$ (Figure 4b, orange). Cluster 2, containing 8 residues, located at the interface between β_2 and α_2, has an average contact density of 2.5 contacts per residue and a contact surface area of 442 $Å^2$ (Figure 4b, magenta). Cluster 3, containing 27 residues, is also located at the interface between the strands and the helices and spans a large portion of the molecule from α_3 to α_7. It is the most densely packed cluster, containing an average of 3.1 contacts per residue, and has a contact surface area of 1720 $Å^2$ (Figure 4b, cyan). Cluster 4 is the smallest cluster, containing 8 residues primarily from side chains in α_8. Cluster 4 is less densely packed, with an average of 2.0 contacts per residue, and it has the smallest contact surface area of 353 $Å^2$ (Figure 4b, yellow). The only inter-cluster contact involves Leu106, which contacts Leu83 in Cluster 1 and Ile108 in Cluster 3. The spaces between the four clusters are occupied by other large nonpolar side chains or by nonpolar segments of nominally polar side chains. The dominant role for ILV side chains in stabilizing these hydrophobic clusters is implied by the fact that, as a group, they represent 75% of the total number of large nonpolar side chains in sIGPS.

The comparison of the HD exchange average protection patterns (Figure 4a) and the ILV clusters (Figure 4b) for sIGPS can also be visualized in a two-dimensional contact map of the large nonpolar side chains (Figure 5). The highest continuous density of ILV-ILV self contacts is found in the region from β_2 to β_6, with contributions from Cluster 2 and the N-terminal portion of Cluster 3 from α_3 to β_6. This segment is coincident with the most strongly protected region by hydrogen exchange (Figure 5, blue box) and corresponds to the I_a equilibrium species. The I_b equilibrium species, spanning the region from α_1 to α_7 (Figure 5, green box), corresponds closely to that defined by side chains from Clusters 2 and 3. The apparent absence of protection against HD exchange for the peptide corresponding to α0 for either I_a or I_b (Table 3) rules out a significant role for Cluster 1 in stabilizing either species. The absence of protection in the region corresponding to Cluster 4 may reflect its small size and paucity of contacts with the other ILV clusters.

The partitioning of Cluster 3 in the I_a intermediate may reflect the larger number of short-range ILV-ILV contacts, i.e. contacts near the diagonal, in the N-terminal region (residues 90-160) compared to the C-terminal region (residues 160-200). The formation of short-range contacts would entail a smaller chain entropy penalty than the long-range contacts prevalent at the C-terminus of Cluster 3. Apparently, the larger entropy penalty for recruiting the C-terminal region of Cluster 3 is only overcome by the formation of additional structure and stability as I_a folds to I_b. Similarly, the absence of a role for Cluster 1 in stabilizing the intermediates may reflect the distribution of its resident side chains across the entire sequence (Figure 5).

Phenylalanine, tyrosine, and methionine residues are prevalent at the edges of Clusters 2 and 3 (Figure 5), suggesting that they may play a role in defining the boundaries of the ILV clusters and, in the case of Clusters 2 and 3, fusing adjacent ILV clusters into a single entity in both the I_a and I_b species. All three of these side chains are inherently more polarizable than their aliphatic counterparts, ILV, and, therefore, might serve as an interface between the solvent and the hydrocarbon core of the ILV clusters. In that role, phenylalanine, tyrosine and methionine might fuse α1 into the I_b intermediate (Figures 2 and 5). When shared between two clusters, e.g. Clusters 2 and 3, F, Y and M might fuse them into the single thermodynamic entity found in I_a and I_b. Mutational analysis will be required to assess the role of adjacent non-ILV side chains in defining the properties of ILV-dominated clusters.

The very good agreement between the location of two of the four ILV clusters and the regions of greatest average protection against HX for sIGPS provides strong support for the hypothesis that ILV clusters play a significant role in defining the structures of partially folded states in TIM barrel proteins.

General implications for the folding mechanisms of TIM barrel proteins

Accumulating evidence supports the hypothesis that the common structural motif of TIM barrel proteins defines the conserved folding free energy surface, including the almost uniform existence of a stable intermediate.[5] However, comparison of the cores of structure in the equilibrium folding intermediates of sIGPS and αTS identified by HX-MS and peptide mapping shows that these cores are not conserved. The minimal core of structure in sIGPS, defined by the I_a intermediate, involves $(\beta\alpha)_{2\text{-}5}\beta_6$, and the maximal core, defined by I_b, involves $\alpha_1(\beta\alpha)_{2\text{-}7}$. By contrast, the core of structure in the single intermediate for αTS measured by HX-MS involves $(\beta\alpha)_{1\text{-}4}$.[18] Although an equilibrium intermediate is preserved for these two sequences of very low identity, 13%, different elements of secondary structure combine to stabilize this species. This result is similar to those for other TIM barrel proteins where the structures of the intermediates vary from sequence to sequence (see Introduction).

What does appear to be common among the structures of all of the TIM barrel intermediates reported thus far, however, is a minimal βαβαβ module. On this essential template, other elements of secondary structure are added in a sequential fashion to form one or more stable intermediates whose folding to the native conformation represents one of the rate-limiting steps in the folding of TIM barrel proteins. To a first approximation, the final step in folding appears to correspond to the closure of the β-barrel with the completion of its cylindrical network of hydrogen bonds and the packing of the surface helices.

The modular assembly of what appear to be α and β elements of secondary structure for sIGPS is reminiscent of the foldon hypothesis advanced by Englander and his colleagues for cytochrome c.[31-33] For this small helical protein, it has been argued that the folding mechanism corresponds to the progressive accumulation of individual helical elements and loops, i.e. foldons.[32] The results of a recent native-centric simulation of the folding are consistent with this proposal.[34] The foldon hypothesis has also been argued to provide a general structural explanation[35] for the folding mechanisms of several other proteins, including RNase H,[36] apocytochrome b_{562}[37] and dimeric yeast TIM.[21] Unfortunately, the inherently lower resolution of the HX-MS method of mapping average protection against exchange, compared to the site-specific NMR and MPAX[21] methods, precludes a definitive assessment of the foldon hypothesis for sIGPS.

The close correspondence between the location of hydrophobic clusters defined by the side chains of branched aliphatic amino acids, isoleucines, leucines and valines, and the segments of the sequence that define the stable intermediates for sIGPS and for αTS[18] suggests that these particular amino acids play key roles in defining the folding free energy surface. Further, this correspondence implies that native-like sub-structures appear during folding reactions and, thereby, enable the acquisition of structure and stability to reach their native conformations. The very large fraction of ILV residues in all TIM barrels suggests that clusters of their aliphatic side chains might also guide the folding reactions of this entire class of proteins. This conjecture is supported by the existence of large hydrophobic clusters in the crystal structures of many other TIM barrels, including yeast TIM and PRAI (S. Kathuria and Y. Wu, unpublished results). In the latter instances, the agreement between the location of the ILV clusters and the cores of structure[15,21] is quite striking. Ongoing HX-MS experiments to map the structures of the stable intermediates in these and other TIM barrel proteins by HX-MS will provide a robust test of this hypothesis.

Materials and Methods

Reagents

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

Protein expression and purification

A truncated version of sIGPS containing a deletion of the first 26 residues, corresponding to helix 00 in the crystal structure, was used in these studies to enhance solubility and avoid aggregation encountered with the full-length protein.[8,22] The protein expression and purification were performed as described previously with minor revisions.[7] A final size-exclusion purification step on a Superdex-75 gel filtration column was typically required to obtain high purity protein for mass spectrometric studies. Purified protein was stored as a precipitate in 95% ammonium sulfate at 4 °C. The purity of sIGPS was confirmed by SDS-PAGE and mass spectrometry assays.

Preparation of samples for mass spectrometry analysis

sIGPS unfolding was initiated by dissolving 10 μl of sIGPS in 90 μl of urea solution (0-8 M urea, 10 mM potassium phosphate, pH 7.8). Following incubation at 25 °C overnight, 900 μl of buffer containing $^{2}H_{2}O$ in the same concentration of deuterated urea was added to initiate isotopic labeling for 5 s. 0.025% TFA in 5% acetonitrile was then added to drop the pH of the sample to 2.7, and quenched samples were immediately frozen in liquid nitrogen and stored at −80 °C until analysis. A non-deuterated, 0% reference sample served as the reference for native sIGPS, and the 100% exchanged reference sample was prepared by incubating sIGPS in 8 M deuterated urea overnight at pH 7.8.

Isotope analysis by HPLC ESI-MS

Deuterium levels in intact sIGPS were determined by HPLC-ESI-MS. The HPLC system consisted of a pump (Rheos Flux Instruments 2000, Leap Technologies Inc, Carrboro, NC) connected to a reversed phase column (Vydac: C_4, 1 mm × 10 mm). A five minute gradient (2%-98% acetonitrile, 0.05% TFA) was used for elution at a flow rate of 50 μl/min. 50 μl samples were used for each injection. The results for intact protein were obtained using an ion trap LCQ ESI mass spectrometer (Thermal Electron, previously Finnegan, San Jose, CA). All components of the HPLC system, from the injector to the electrospray probe, were submerged in ice.

Pepsin digestion of sIGPS for MALDI-TOF MS analysis

Pepsin digestion was carried out in a batch mode with immobilized pepsin at an enzyme to substrate W/W ratio of 2:1. Immobilized pepsin slurry (100 μl) was initially washed and spun 4 times in a 1.5 ml 0.2 μ Millipore microfiltration centrifuge tube with 500 μl chilled 0.025% TFA. Quenched protein samples were added to the washed pepsin resin and digested for 4 min on ice with occasional mixing. The digested samples were centrifuged for a few seconds at 14,000 g and immediately chilled in liquid nitrogen. The samples were stored at −80 °C until use. Peptic fragments were identified by MS/MS using LCQ ESI mass spectrometry (Thermal Electron, San Jose, CA).

Preparation of samples for MALDI-TOF MS analysis

The MALDI plate was prepared with minor revision according to the method of Komives et al.[38,39] The MALDI plate was chilled on ice in a small plastic box for 10 min before use, and the condensation was blown off by dry nitrogen before sample application. 0.5 μl of matrix of 10 mg ml^{-1} α-cyano-4-hydroxycinnamic acid was first applied on the target plate. Frozen samples of protein or peptides were quickly defrosted and spotted on the target plate. The plate was immediately put into a desiccator, where spots would dry in 1-2 min under a moderate vacuum in the presence of desiccant. The plate was transferred immediately to the mass spectrometer to minimize the back exchange. Matrix-assisted laser desorption ionization time-of-flight (MALDI) mass spectra were obtained on a Micromass M@LDI-LR instrument in the reflectron mode. A sampling rate of 0.5 ns was used with a source voltage of 15 kV, a pulse voltage of 2175.0 V and a micro-channel plate voltage of 1850 V. The matrix suppression delay was set to 400 mass units and delayed extraction was used with a nominal pulse delay of 500 ns. The results of 50-100 laser shots were averaged to generate each mass spectrum.

Analysis of spectral data from LC-ESI MS and MALDI-TOF MS

Data analysis methods have been previously reported.[18,24] Deuterium incorporation for intact protein and peptide fragments, D, was determined according to Equation 1,

$$D = \frac{m - m_{0\%}}{m_{100\%} - m_{0\%}} \times N \quad (1)$$

where N is the total number of exchangeable amide hydrogens in the peptide, m is the mass of the analyzed sample, $m_{0\%}$ is the mass of the undeuterated sample and $m_{100\%}$ is the mass of the fully-deuterated sample. The mass of the sample, m, is measured by determining the centroid of the isotopic envelope. The centroid was then compared to the centroids of fully undeuterated and fully deuterated sIGPS to correct for deuterium loss during analysis.

Visualization of ILV clusters from the crystal structure of sIGPS

The distance between side chain carbon atoms of the selected residue types was computed using an in-house program and a typical van der Waals cut-off distance of 4.2 Å. The number of contacts between any two residues was counted and exported to Pymol[40] for 3D visualization of the clusters formed by networks of side chain-side chain clusters. The crystal structure of the truncated IGPS from *S. solfataricus* (PDB code 2C3Z)[25] was used for the cluster calculations; comparable results were obtained when the full-length crystal structure was used (PDB code 1A53).[41] Contact surface area was calculated using the program SPACE.[42]

Acknowledgements

This work was supported by N.I.H. grant GM23303 to C.R.M. 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 visualizing the hydrophobic clusters, and we also thank Drs. R. Forsyth, R. Simler, R. Vadrevu and Y. Wu for many insightful discussions.

Abbreviations

αTS, alpha subunit of tryptophan synthase from *E. coli*; CD, circular dichroism; ESI, electrospray ionization; HPLC, high performance liquid chromatography; HD, hydrogen-deuterium; HX-MS, hydrogen-deuterium exchange mass spectrometry; I_a and I_b, partially-folded states in sIGPS; ILV, isoleucine, leucine and valine residues; MALDI, matrix assisted laser desorption ionization; FMY, phenylalanine, methionine and tyrosine residues; MPAX, misincorporation proton-alkyl exchange; MS, mass spectrometry; N, native state; PRAI, N-

(5′-phosphoribosyl)anthranilate isomerase; sIGPS, indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*; TIM, triose phosphate isomerase; TFA, trifluoroacetic acid; U, unfolded state.

References

1. Vega MC, Lorentzen E, Linden A, Wilmanns M. Evolutionary markers in the $(\beta/\alpha)_8$-barrel fold. Curr. Opin. Chem. Biol 2003;7:694–701. [PubMed: 14644177]
2. Gerlt JA, Raushel FM. Evolution of function in $(\beta/\alpha)_8$-barrel enzymes. Curr. Opin. Chem. Biol 2003;7:252–264. [PubMed: 12714059]
3. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol 2002;321:741–765. [PubMed: 12206759]
4. Wierenga RK. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett 2001;492:193–198. [PubMed: 11257493]
5. Sterner R, Hocker B. Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. Chem. Rev 2005;105:4038–4055. [PubMed: 16277370]
6. Gualfetti PJ, Bilsel O, Matthews CR. The progressive development of structure and stability during the equilibrium folding of the α subunit of tryptophan synthase from *Escherichia coli*. Protein Sci 1999;8:1623–1635. [PubMed: 10452606]
7. 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]
8. Sanchez del Pino MM, Fersht AR. Nonsequential unfolding of the α/β barrel protein indole-3-glycerol-phosphate synthase. Biochemistry 1997;36:5560–5565. [PubMed: 9154940]
9. Jasanoff A, Davis B, Fersht AR. Detection of an intermediate in the folding of the $(\beta\alpha)_8$-barrel N-(5′-phosphoribosyl)anthranilate isomerase from *Escherichia coli*. Biochemistry 1994;33:6350–6355. [PubMed: 8193151]
10. Bilsel O, Zitzewitz JA, Bowers KE, Matthews CR. Folding mechanism of the α-subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. Biochemistry 1999;38:1018–1029. [PubMed: 9893998]
11. Higgins W, Fairwell T, Miles EW. An active proteolytic derivative of the α subunit of tryptophan synthase. Identification of the site of cleavage and characterization of the fragments. Biochemistry 1979;18:4827–4835. [PubMed: 389284]
12. Miles EW, Yutani K, Ogasahara K. Guanidine hydrochloride induced unfolding of the α subunit of tryptophan synthase and of the two α proteolytic fragments: Evidence for stepwise unfolding of the two α domains. Biochemistry 1982;21:2586–2592. [PubMed: 7046790]
13. Eder J, Kirschner K. Stable substructures of eightfold βα-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. Biochemistry 1992;31:3617–3625. [PubMed: 1567820]
14. Zitzewitz JA, Gualfetti PJ, Perkons IA, Wasta SA, Matthews CR. Identifying the structural boundaries of independent folding domains in the α subunit of tryptophan synthase, a β/α barrel protein. Protein Sci 1999;8:1200–1209. [PubMed: 10386870]
15. Akanuma S, Yamagishi A. Identification and characterization of key substructures involved in the early folding events of a $(\beta/\alpha)_8$ -barrel protein as studied by experimental and computational methods. J. Mol. Biol 2005;353:1161–1170. [PubMed: 16216267]
16. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R. Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. Nat. Struct. Biol 2001;8:32–36. [PubMed: 11135667]
17. Hocker B, Claren J, Sterner R. Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. Proc. Natl Acad. Sci. USA 2004;101:16448–16453. [PubMed: 15539462]
18. Rojsajjakul T, Wintrode P, Vadrevu R, Robert Matthews C, Smith DL. Multi-state unfolding of the α 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]

19. Pan H, Smith DL. Quaternary structure of aldolase leads to differences in its folding and unfolding intermediates. Biochemistry 2003;42:5713–5721. [PubMed: 12741828]
20. Pan H, Raza AS, Smith DL. Equilibrium and kinetic folding of rabbit muscle triosephosphate isomerase by hydrogen exchange mass spectrometry. J. Mol. Biol 2004;336:1251–1263. [PubMed: 15037083]
21. Silverman JA, Harbury PB. The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel. J. Mol. Biol 2002;324:1031–1040. [PubMed: 12470957]
22. 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]
23. Krishna MM, Hoang L, Lin Y, Englander SW. Hydrogen exchange methods to study protein folding. Methods 2004;34:51–64. [PubMed: 15283915]
24. 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]
25. Schneider B, Knochel T, Darimont B, Hennig M, Dietrich S, Babinger K, Kirschner K, Sterner R. Role of the N-terminal extension of the $(\beta\alpha)_8$-barrel enzyme indole-3-glycerol phosphate synthase for its fold, stability, and catalytic activity. Biochemistry 2005;44:16405–16412. [PubMed: 16342933]
26. 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]
27. Gromiha MM, Pujadas G, Magyar C, Selvaraj S, Simon I. Locating the stabilizing residues in $(\alpha/\beta)_8$ barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. Proteins 2004;55:316–329. [PubMed: 15048825]
28. 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. 2006in press
29. Wu Y, Vadrevu R, Yang X, Matthews CR. Specific structure appears at the N terminus in the sub-millisecond folding intermediate of the α subunit of tryptophan synthase, a TIM barrel protein. J. Mol. Biol 2005;351:445–452. [PubMed: 16023136]
30. Branden, C.; Tooze, J. Introduction to Protein Structure. 2nd edit. Garland Publishing, Inc.; New York: 1999.
31. Bai Y, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: Native-state hydrogen exchange. Science 1995;269:192–197. [PubMed: 7618079]
32. Maity H, Maity M, Englander SW. How cytochrome c folds, and why: Submolecular foldon units and their stepwise sequential stabilization. J. Mol. Biol 2004;343:223–233. [PubMed: 15381432]
33. Krishna MM, Maity H, Rumbley JN, Lin Y, Englander SW. Order of steps in the cytochrome c folding pathway: Evidence for a sequential stabilization mechanism. J. Mol. Biol 2006;359:1410–1419. [PubMed: 16690080]
34. Weinkam P, Zong C, Wolynes PG. A funneled energy landscape for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments. Proc. Natl. Acad. Sci. USA 2005;102:12401–12406. [PubMed: 16116080]
35. Maity H, Maity M, Krishna MM, Mayne L, Englander SW. Protein folding: the stepwise assembly of foldon units. Proc Natl Acad Sci USA 2005;102:4741–4746. [PubMed: 15774579]
36. 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]
37. Chu R, Pei W, Takei J, Bai Y. Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. Biochemistry 2002;41:7998–8003. [PubMed: 12069590]
38. Mandell JG, Falick AM, Komives EA. Identification of protein-protein interfaces by decreased amide proton solvent accessibility. Proc. Natl. Acad. Sci. USA 1998;95:14705–14710. [PubMed: 9843953]
39. Mandell JG, Falick AM, Komives EA. Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry. Anal. Chem 1998;70:3987–3995. [PubMed: 9784743]
40. Delano, WL. The Pymol molecular graphics system. DeLano Scientific; San Carlos, CA: 2002.

41. Hennig M, Darimont BD, Jansonius JN, Kirschner K. The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from *Sulfolobus solfataricus* with substrate analogue, substrate, and product. J. Mol. Biol 2002;319:757–766. [PubMed: 12054868]

42. Sobolev V, Eyal E, Gerzon S, Potapov V, Babor M, Prilusky J, Edelman M. SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. Nucleic Acids Res 2005;33:W39–43. [PubMed: 15980496]

Scheme 1.

Figure 1.

Representative mass spectra of the +25 charge state of sIGPS equilibrated in 0-8 M urea followed by pulse-labeling in urea/$^{2}H_2O$ for 5 s at 25 °C and pH 7.8. Isotope exchange was quenched on ice with TFA/acetonitrile at pH 2.7, and samples were analyzed by HPLC ESI-MS. The m/z values of ions representing the native, N, intermediate, I, and unfolded, U, sIGPS are 1027, 1030 and 1032.2, respectively, and the corresponding peaks are labeled.

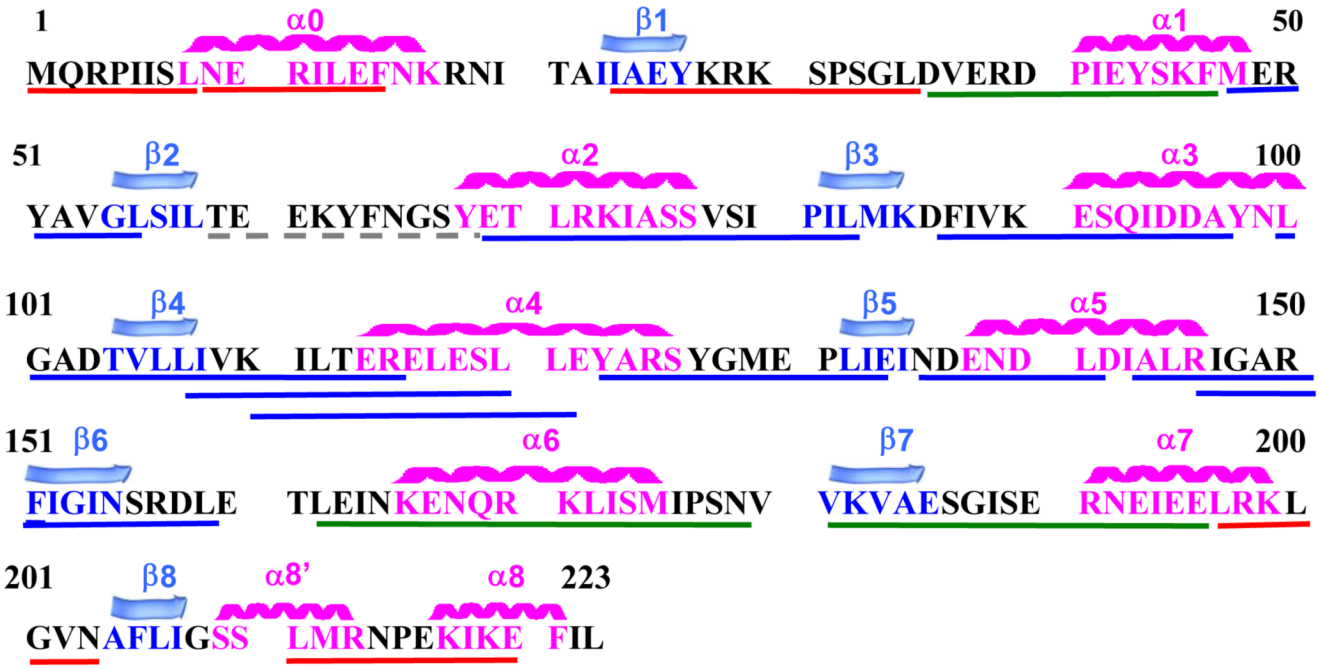

Figure 2.

Amino acid sequence of sIGPS displaying secondary structure (above) and peptide coverage (below). Regions of the sequence corresponding to β-strands are colored blue, and α-helical segments are colored fuchsia. More highly-protected peptides (≥ 49% average protection) are underlined in blue, those with lower protection (23-27% average protection) are underlined in green, and unprotected peptides (≥ 11% average protection) at 5 M urea are underlined in red. The dashed gray region is the unstructured loop region linking β2 and α2, whose protection against HD exchange cannot be measured even in the native state.

Figure 3.
Representative mass spectra of peptides displaying the four different types of exchange behavior observed at 5 M urea and at 25 °C, pH 7.8: (a) unprotected at 5 M urea, residues 211-220, (b) EX1 mechanism with lower average protection (23-27%), residues 180-196, (c) EX2 mechanism with higher average protection (≥49%), residues 143-151, and (d) EX1 mechanism with higher average protection (≥49%), residues 107-120. The isotope envelopes representing the unfolded reference at higher m/z in 8 M urea and the folded reference at lower m/z in 0 M urea are shown as dotted and dashed lines, respectively

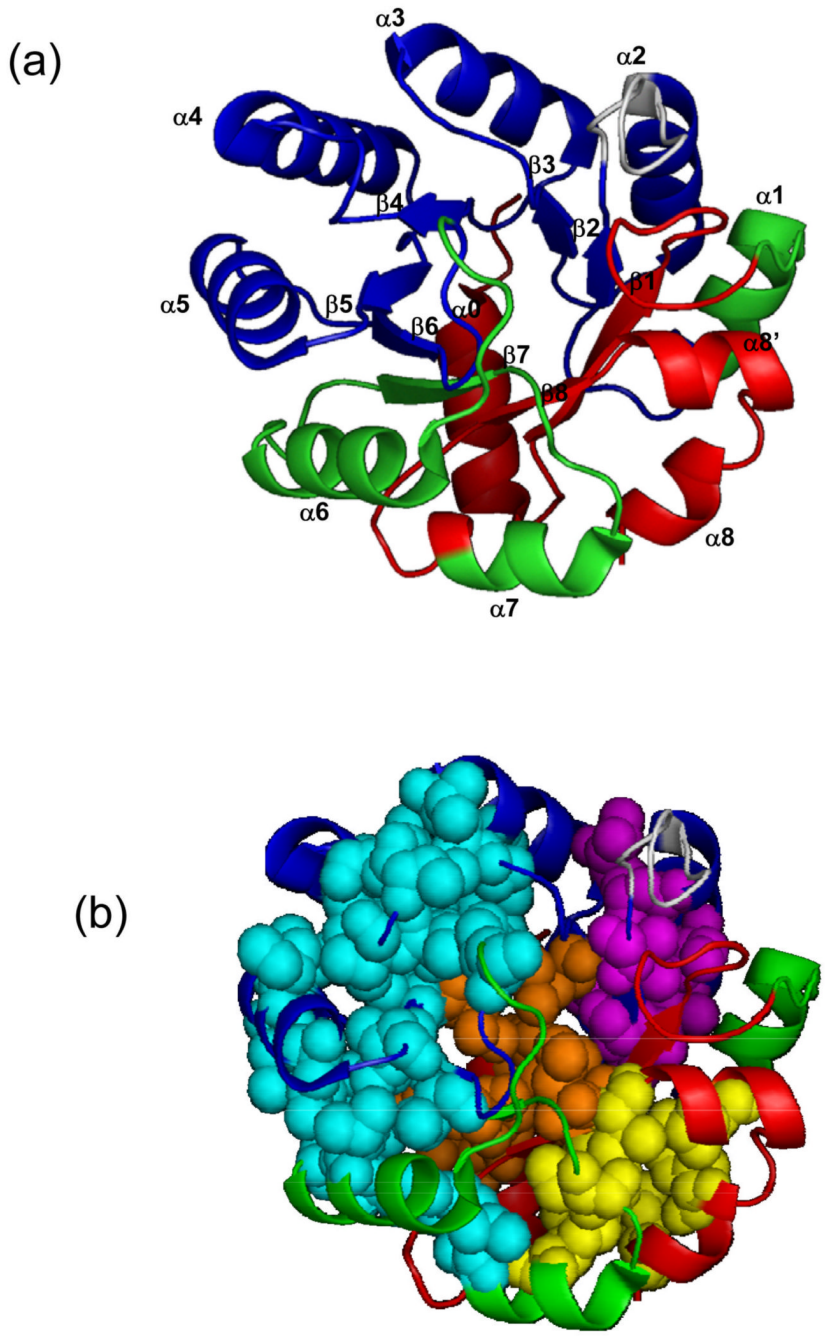

Figure 4.

Comparison of the HD average protection pattern for sIGPS pulse-labeled at 5 M urea with the position of hydrophobic clusters of ILV residues for sIGPS. (a) Structured regions in the equilibrium intermediate determined from the HD exchange mass spectrometry experiment. More highly protected regions (≥49%) are shown in blue, regions with lesser protection (~25%) are shown in green, and unprotected regions (≤ 11%) are shown in red. (b) Four hydrophobic clusters formed by ILV residues obtained using a 4.2 Å cut-off: cluster 1 (orange); cluster 2 (magenta); cluster 3 (cyan) and cluster 4 (yellow). Structures were generated using Pymol[40] and Protein Data Bank entry 2C3Z.pdb.[25]

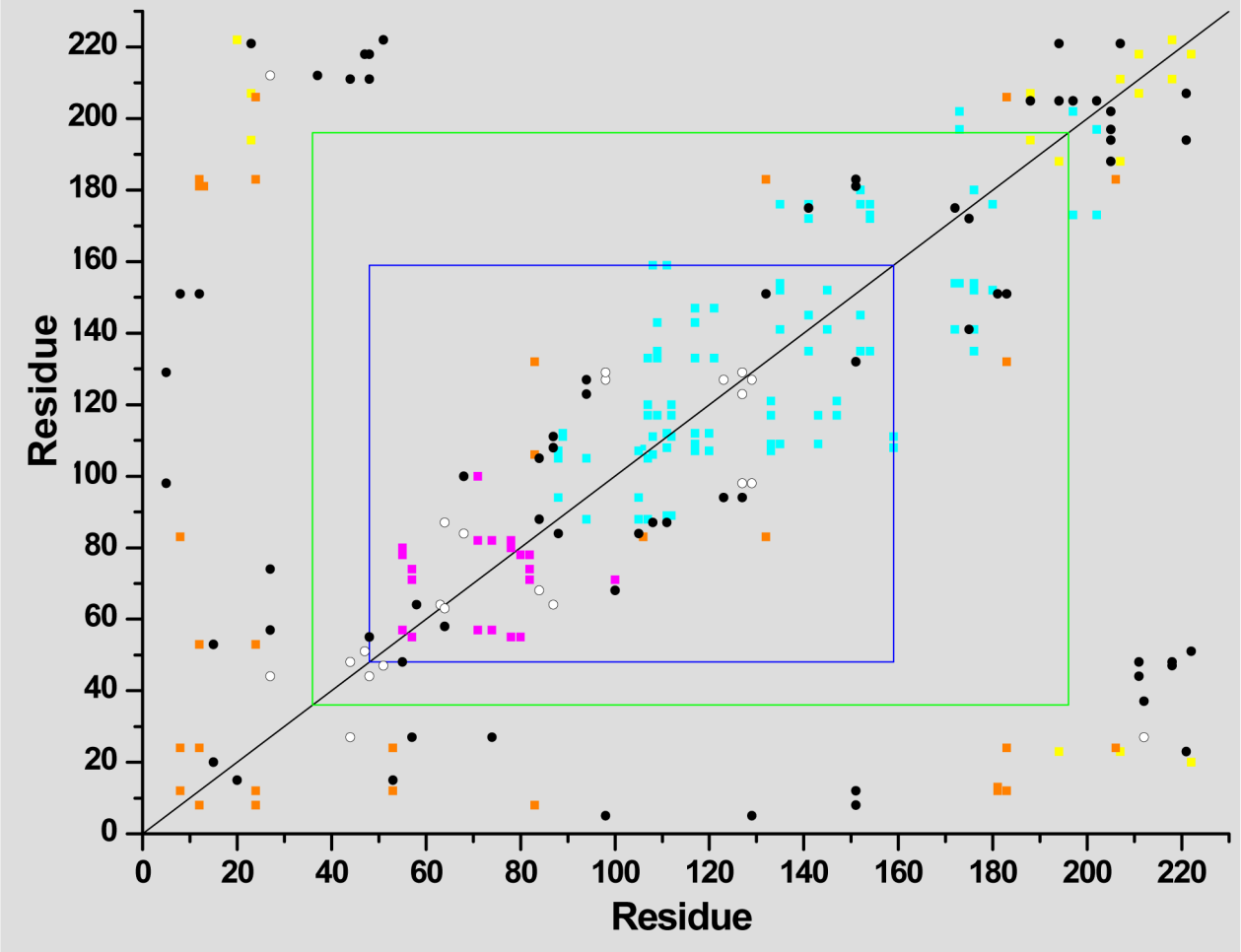

Figure 5.

Two-dimensional contact map at 4.2 Å highlighting pairwise hydrophobic contacts for sIGPS. The ILV to ILV contacts are shown with colored squares according to the color scheme for the clusters in Figure 4b, the ILV to FMY contacts are shown with open circles, and the FMY to FMY contacts are shown with closed black circles. sIGPS contains no tryptophan side chains. The blue box encompasses the most strongly protected region and represents structured regions in the I_a intermediate, and the green box encompasses regions that are at least moderately protected and represents structured regions in the I_b intermediate.

Table 1

Molecular masses and relative populations of native, intermediate, and unfolded species at various urea concentrations[a]

[Urea] (M)	Native		Intermediate		Unfolded	
	Mass[b]	Abundance[c]	Mass[b]	Abundance[c]	Mass[b]	Abundance[c]
0	25639	100				
2.0	25641	91	25715	9	25760	0
4.0	25641	51	25716	41	25762	8
4.5	25649	36	25715	43	25761	21
5	25649	4	25717	52	25762	44
5.5	25649	0	25715	28	25763	72
8.0			25715	0	25763	100
0% ref	25577					
100% ref					25769	

[a] Data correspond to the peaks shown in Figure 1, and experimental conditions are described in the caption to Figure 1.

[b] Average molecular masses were determined from the centroids of envelopes of isotope peaks representing N, I and U.

[c] Abundances were determined from the relative areas of the envelopes of isotope peaks. Errors are < 10% and reflect the reproducibility of the experiment and uncertainties resulting from fitting the peak areas to Gaussian functions.

Table 2

Determination of the number of protected amide hydrogens in the stable equilibrium intermediate for sIGPS

	N	I	U
Measured molecular mass[a]	**25639**	**25715**	**25763**
Deuterium level (measured)[b]	**62**	**138**	**186**
Deuterium level (adjusted)[c]	**70**	**155**	**209**
Number of amides protected[d]		**54<#<131**	

[a] Average masses taken from Table 1.

[b] Determined by subtracting the mass of the 0% reference state (25577) from the measured molecular masses.

[c] Calculated by multiplying the measured deuterium levels by 216/192, which is the ratio of the total number of amide hydrogens (223 amino acids – 7 prolines) to the difference between the 100% and 0% reference measurements (25769 – 25577).

[d] The lower limit is calculated by subtraction of the adjusted deuterium level in the I state from the deuterium level in the U state. The upper limit is obtained by subtracting the difference between the deuterium levels of the I and N states, 85, from the total number of exchangeable amides, 216.

Table 3

Deuterium levels in peptic fragments derived from pulse-labeled and quenched sIGPS equilibrated at various urea concentrations[a] of urea at pH 7.8 and 25 °C.

Fragment	Secondary Structure[b]	Urea Concentration (M)				Percent Protected[c]		Exchange Behavior at 5 M Urea
		0	4	5	8	4 M Urea	5 M Urea	
1-8		2	4.3	5.7	6.0	42	8	NP
9-15	α_0	0.8	4.5	5.4	6.0	29	11	NP
24-35	β_1	3.9	8.6	10.7	11.0	34	4	NP
36-47	α_1	4.5	6.8	8.7	10.0	58	23	EX1
48-55	$(\alpha_1)/\beta_2$	1.1	2.0	2.8	7.0	84	71	EX2
59-68	disordered	9.0	9.0	9.0	9.0	NP[d]	NP	NP
69-83	α_2/β_3	1.8	6.0	7.5	13.0	63	49	EX1
87-97	α_3	0.6	1	1.7	10.0	95	88	EX1
100-115	$\beta_4/(\alpha_4)$	2.6	5.3	7.8	14.0	76	54	EX1
107-120	β_4/α_4	2.1	2.1	4.0	13.0	100	83	EX1
109-121	α_4	2.2	2.5	4.3	12.0	97	78	EX1
123-134	α_4/β_5	1.4	2.2	3.1	10.0	91	80	EX1
135-141	β_5/α_5	0.3	1.2	1.7	5.0	81	70	EX1
143-151	α_5/β_6	0.8	1.9	3.2	7.0	84	61	EX2
146-159	$(\alpha_5)/\beta_6$	1.1	3.8	7.2	13.0	77	49	EX2
162-179	α_6	5.0	10.1	12.3	15.0	49	27	EX1
180-196	β_7/α_7	6.8	10.6	12.9	15.0	53	25	EX1
197-203	α_7	0.4	3.5	6.0	6.0	45	0	NP
211-220	α_8	2.4	6.0	8.0	8.0	36	0	NP

[a] Deuterium levels were adjusted for labeling in 90% ^2H_2O.

[b] The elements of secondary structure contained within each peptide. Secondary structural elements in parentheses are only partially represented by one or two residues.

[c] To determine the percent of protection at 4 and 5 M urea, the deuterium levels at those urea concentrations were normalized against the difference in the deuterium levels in the 0 and 8 M urea samples.

[d] Not Protected.