Frustration and folding of a TIM barrel protein

Kevin T. Halloran1, Yanming Wang1, Karunesh Arora1, Srinivas Chakravarthy2, Thomas C. Irving3, Osman Bilse1, Charles L. Brooks III1,2, and C. Robert Matthews1,2

1Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605; 2Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, MI 48109; and 3Biophysics Collaborative Access Team (BioCAT), Center for Synchrotron Radiation Research and Instrumentation and Department of Biological and Chemical Sciences, Illinois Institute of Technology, Chicago, IL 60616

Edited by William F. DeGrado, University of California, San Francisco, CA, and approved June 27, 2019 (received for review January 22, 2019)

Triosephosphate isomerase (TIM) barrel proteins have not only a conserved architecture that supports a myriad of enzymatic functions, but also a conserved folding mechanism that involves on- and off-pathway intermediates. Although experiments have proven to be invaluable in defining the folding free-energy surface, they provide only a limited understanding of the structures of the partially folded states that appear during folding. Coarse-grained simulations employing native centric models are capable of sampling the entire energy landscape of TIM barrels and offer the possibility of a molecular-level understanding of the readout from sequence to structure. We have combined sequence-sensitive native centric simulations with small-angle X-ray scattering and time-resolved Förster resonance energy transfer to monitor the formation of protein-folding intermediates. Simulations with small-angle X-ray scattering and time-resolved Förster resonance energy transfer to monitor the formation of protein-folding intermediates. Simulations reveal the presence of a major and 2 minor folding channels not detected in experiments. Frustration in folding, i.e., backtracking in native contacts, is observed in the major channel at the initial stage of folding, as well as late in folding in a minor channel before the appearance of the native conformation. Similarities in global and pairwise dimensions of the early intermediate, the formation of structure in the central region that spreads progressively toward each terminus, and a similar rate-limiting step in the closing of the β-barrel underscore the value of combining simulation and experiment to unravel complex folding mechanisms at the molecular level.

Significance

The ways in which proteins fold to their native state remain a challenging and important question in biology. Large proteins often populate meta-stable intermediate states before reaching their native conformation, making them excellent targets for understanding the sequence-structure code. We studied folding of indole-3-glycerol phosphate synthase, a βα8 TIM barrel protein, using SAXS and FRET experiments and coarse-grained computer simulation to gain insights into structures of early folding intermediates. The folding mechanism first proposed by experimental studies is elaborated by a more complete mechanism involving 3 folding channels revealed by simulation. We provide connections between the off-pathway intermediate and one on-pathway rate-determining intermediate detected by experiment with corresponding meta-stable states in the major folding channel found in the simulations.


Published online July 25, 2019.

Published under the PNAS license.

1K.T.H. and Y.W. contributed equally to this work.
2To whom correspondence may be addressed. Email: brookscl@umich.edu or c.robert.matthews@umassmed.edu.

This article contains information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1900880116/-/DCSupplemental.

16378-16383 | PNAS | August 13, 2019 | vol. 116 | no. 33

www.pnas.org/cgi/doi/10.1073/pnas.1900880116
folding events in the microsecond time range and have allowed us to examine the earliest events in the folding of a candidate TIM barrel protein (13). In the present study, pair-wise distance measurements from time-resolved Förster resonance energy transfer (trFRET) and global size and shape measurements from small-angle X-ray scattering (SAXS), combined with coarse-grained computer simulations, are employed to probe the earliest events in the folding of the Sulfolobus solfataricus indole-3-glycerol phosphate synthase (SsIGPS) TIM barrel. Surprisingly, the global collapse of the unfolded chain to a misfolded, off-pathway intermediate occurs within 50 μs. The simulations reveal the potential complexities of this exceedingly rapid reaction and show that the rate-limiting step in the folding of SsIGPS is the frustration encountered by the competition between the N- and C-terminal β-strands to close the 8-stranded β-barrel.

**Results**

The SsIGPS TIM barrel is a representative member of the most common architecture for enzymes in biology. The 8 alternating β and α elements are arranged sequentially as a central parallel-stranded β-barrel encompassed by an α-helical shell (Fig. 1A). Its folding mechanism has previously been shown to begin with the submillisecond formation of an off-pathway intermediate, IBP, followed by 2 on-pathway intermediates, IA and IB, before reaching the native state (Fig. 1B). The IBP intermediate has an apparent stability of 3.5 kcal-mol⁻¹, is rich in secondary structure, and displays strong protection against exchange of amide hydrogens with solvent in the central (βα)₄ module within 75 ms (14). As the folding reaction proceeds, the protection expands to encompass (βα)₂₋₅ in IA and (βα)₁₋₄ in IB. The fully folded TIM barrel appears in the final step of folding.

**Measuring Global Dimensions by SAXS.** To obtain global insights into the structures of the intermediates, SAXS profiles were obtained under equilibrium and kinetic refolding conditions. At equilibrium, the native state of the protein has a radius of gyration (Rg) of ~18 Å, and the unfolded state has an estimated Rg of 46 Å by linear extrapolation from high denaturant conditions (Fig. 2A). The IA intermediate, highly populated at 4 M urea (14), self-associates at the 80-μM concentration required for reliable SAXS measurements, precluding an estimate of its Rg (SI Appendix, Fig. S1).

The Rg of the submillisecond burst-phase intermediate IBP was determined by a 10-fold dilution from 8 M urea, using a custom, single-piece microfluidic mixer. Within 150 μs, the dead time of the mixer, the Rg of SsIGPS is 26 ± 1.5 Å (Fig. 2B). The absence of change in Rg out to 4 ms, where IBP can be detected by stop-flow circular dichroism, shows that the IBP intermediate appears within 150 μs. The conclusion that the SAXS-detected burst-phase species is a discrete thermodynamic state, and not a collapsed form of the unfolded state, is supported by the observation of a Rg that is insensitive to the urea concentration up to 2 M urea (Fig. 2A). A collapsed form of the unfolded state would have been expected to swell with increasing urea concentration (15).

Transformation of the scattering curve from the native state of SsIGPS in 0.8 M urea to a dimensionless Kratky plot (16) shows the parabolic shape typical of globular structure. The maximum in the plot occurs at (√3, 1.1) as expected for the Guinier approximation (Fig. 2C) (16). At 8 M urea, SsIGPS has an extended random coil-like structure with the expected hyperbolic plateau shape at high scattering angle × Rg (qRg). By contrast, the dimensionless Kratky plot for the continuous flow refolding curve shows that the IBP state has a peak shift on the qRg axis to approximately (2, 1.25). This behavior deviates from the Guinier approximation and shows that the protein has regions that are not yet fully globular. The pair distribution function, P(r), for IBP confirms a large collapse of the chain from the unfolded state (U) to IBP within 150 μs (Fig. 2D). The maximum distance between any 2 atoms, Dmax, concomitantly decreases from 130 to 80 Å, and the significant shoulder at ~70 Å shows that IBP is not fully globular.

**Pair-Wise Dimensional Analysis by trFRET.** To complement the global dimensional data obtained by SAXS, 2 sets of pair-wise distances were measured by trFRET experiments on SsIGPS, using tryptophan and AEDANS. One FRET pair was positioned to monitor barrel closure by measuring the distance distribution between α₁ and α₂, W63-C/AEDANS238. The second pair was positioned to monitor the formation of the strongly protected segment, W112-C140, containing the strongly protected (βα)₂₋₅ module (14). A collapsed form of the unfolded state would have been expected to swell with increasing urea concentration (15).

**Fig. 1.** (A) A ribbon representation of the structure of SsIGPS (PDB code 2C32) prepared by PyMol. The FRET pairs employed to study the central (βα)₄ segment, W112-C140, containing the strongly protected (βα)₂₋₅ module and the N- and C-termini, W63-C238, are highlighted with W112 and W63 residues in blue and the C140 and C238 residues in red. (B) The reaction diagram of SsIGPS (14). The barrier heights were estimated using Kramer’s formalism with a prefactor of 1 μs. The blue arrows indicate aspects of the free energy landscape probed in this study by simulations, and the red arrow indicates the focus of the present experiments.

**Fig. 2.** (A) Rg as a function of urea concentration for the unfolding of SsIGPS (black circles). The estimated Rgs of the native and unfolded states in water are indicated by linear extrapolation of the native and unfolded baselines. The estimated Rg after 150 μs of refolding at several final urea concentrations in the native baseline region (red circles). (B) Rg as a function of folding time at 0.8 M urea. The Rgs of the unfolded and native states in water are indicated. (C) Dimensionless Kratky plots of the unfolded (blue), IBP intermediate (red), and native state (black). The arrows indicate the maxima in the plots for the N and IBP species. (D) The P(r) of the unfolded (blue), IBP intermediate (red), and native states (black). The dashed lines represent the P(r) for these states calculated from the simulations.
The average Trp lifetimes for donor-only (DO) and donor-acceptor (DA) samples for both the \( \alpha_{\beta}\alpha \) and \( \alpha_{\beta}\alpha \) pairs were analyzed by 2D maximum entropy modeling (2D-MEM) \( (17, 18) \) for the refolding of SsIGPS containing the \( \alpha_{\beta}\alpha \) pair shows a decrease to a non-native-like lifetime of 4.5 ns for the DO sample and 3.3 ns for the DA sample within the dead time of the mixer (50 \( \mu s \)) \( (SI \ Appendix, \ Fig. \ S2A \ and \ B) \). As was the case for Rg, there are no significant changes in lifetimes for both the DO and DA samples from \( \sim 50 \mu s \) out to \( \sim 1 \) ms. Similar behavior was observed for the \( \alpha_{\beta}\alpha \) FRET pair during refolding jumps to 0.8 M urea, demonstrating a global collapse of unfolded SsIGPS within 50 \( \mu s \).

**Maximum Entropy Modeling.** The trFRET data for both the \( \alpha_{\beta}\alpha \) and \( \alpha_{\beta}\alpha \) pairs were analyzed by 2D maximum entropy modeling (2D-MEM) \( (17, 18) \) for the unfolded state (8 M urea, 100 \( \mu s \)), and the native state (0 M urea) \( (Fig. \ 3A \ and \ B \) and \( SI \ Appendix, \ Fig. \ S2C) \). The unfolded state for both FRET pairs show very similar maximum amplitudes from 12 to 35 \( \AA \), the distances most sensitive to FRET for the Trp-AEDANS pair \( (R_0 = 22 \, \AA) \). The maximum normalized amplitude in the native state for the \( \alpha_{\beta}\alpha \) FRET pair \( (Fig. \ 3A) \) is \( \sim 13 \, \AA \), in good agreement with the calculated distance between residues 63 and 238 in the crystal structure, 9.3 \( \AA \). The maximum normalized amplitude after 100 \( \mu s \) for the \( \alpha_{\beta}\alpha \) FRET pair revealed the presence of 2 distinct distributions of distances. One distribution of distances is more compact than native, and the other is more extended than native but more compact than the unfolded state. By contrast, the \( \alpha_{\beta}\alpha \) pair after 100 \( \mu s \) shows a single peak around 20 \( \AA \) \( (Fig. \ 3B) \) that is similar to the native protein but has a broader distribution.

**Ensemble Averaged Folding Properties from Simulations.** To gain deeper insight into the development of structure during the folding of SsIGPS, we complemented the present and previous experimental studies \( (8, 19) \) with a native centric simulation to sample the entire folding landscape.

Native-centric coarse-grained \( \beta \)-G\( \beta \) model \( (20) \) re-folding simulations were initiated from an unfolded ensemble of structures sampled from simulations at high temperature \( (\text{see Materials and Methods}) \) for further details), and 100 independent 2,000 time units \( (1 \, \text{time unit} = 10,000 \, \text{dynamics steps}) \) folding trajectories were sampled in the analysis. Because the underlying model is coarse-grained, the landscape is smoother and the folding timescales are compressed and thereby do not directly correspond to the times observed in experiments. However, we anticipate that the time ordering, as well as the relative lag times between folding phases, should reflect what is observed in kinetic experiments \( (21) \).

The progress of the folding reaction for each trajectory was monitored by the Rg and the fraction of total native contacts \( (Q) \). Over the time courses of the 100 trajectories, persistent values were observed for \( Q_0 \) of 0.3, 0.5, 0.6, 0.8, and 0.9 \( (Fig. \ 4A) \). The initial \( (0.3) \) and final \( (0.9) \) values correspond to the unfolded and native forms of the protein with the intermediate plateaus \( (0.5, 0.6, 0.8) \) suggesting the presence of partially folded states. Examination of the entire set of trajectories revealed that only a small fraction of the simulations reached the native state within 2,000 time units \( (SI \ Appendix, \ Fig. \ S3) \). The majority of the simulations reached a \( Q_0 \) of 0.8. Plateaus at similar time steps were observed for Rg \( (Fig. \ 4B) \), beginning with the unfolded state at \( \sim 45 \, \text{Å} \) and progressively decreasing to 18 \( \text{Å} \) for the native state.

Decomposition of \( Q_0 \) into contributions \( (Q) \) from the N- and C-terminal halves, \( \alpha_{\beta}(\beta \alpha)_{-1} \) and \( (\beta \alpha)_{-5} \), reveals frustration in folding \( (Fig. \ 4C) \). The striking anticorrelated gain/loss in native contacts in N- and C-terminal halves was observed at \( Q_0 = 0.50 \) to 0.65 and \( Q_0 = 0.75 \) to 0.85, where the intermediate states persist, which suggests that the frustration in these 2 regions might be related to those intermediate states. Further decomposition of \( Q_0 \) into the four \( (\beta \alpha)_{-1} \) modules of stability \( (22) \) pinpoints the major sources of frustration \( (Fig. \ 4D) \). At \( Q_0 = 0.5 \) the source of frustration derives from \( (\beta \alpha)_{-1} \) competing with \( (\beta \alpha)_{-5} \), and to a lesser extent \( (\beta \alpha)_{-6} \). The frustration event at \( Q_0 = 0.6 \) is \( (\beta \alpha)_{-1} \) driving folding while \( (\beta \alpha)_{-4} \) and \( (\beta \alpha)_{-6} \) lose native contacts. The final frustration event at \( Q_0 = 0.8 \) is mainly the competition between the \( (\beta \alpha)_{-2} \) and \( (\beta \alpha)_{-7,8} \) modules; however, the competition also affects \( (\beta \alpha)_{-4} \) and \( (\beta \alpha)_{-5} \).

An examination of the contact probability maps at different times gives further insight into the folding mechanism \( (SI \ Appendix, \ Fig. \ S4) \). The central region \( (\text{residue} \sim 30 \text{residues}) \) formed most of its contacts within 400 time units compared with the total simulation time of 2,000 time units. Then, more contacts were formed in the C-terminal region within 1,000 time units. At the end of the simulation, most of the low-probability contacts were those formed between \( \alpha_6 \beta_1 \) (residues \( 1 \text{~to~} 40 \)) and other regions, suggesting that many trajectories ended up with a structure with unfolded \( \alpha_6 \beta_1 \).

**Multiple Folding Pathways Revealed from Simulations.** To obtain further insights into the molecular events that occur during the folding of SsIGPS, we examined each trajectory in detail. Three significant folding pathways were found, based on the assembly order of secondary structural units \( (Fig. \ 5) \). The classification of different states found in all trajectories was based upon both their distinct Q0 and Rg values \( (SI \ Appendix, \ Table \ SI) \) and their visual differences in structure \( (Fig. \ 5 \) and \( SI \ Appendix, \ Fig. \ S5) \).

The U state initially collapses to a single intermediate, Ic, with a well-folded central region \( (\text{residues} \sim 50) \), \( (\beta \alpha)_{-5} \). The Ic state then partitions into the I1A state, the I3 state, or the I1 state with transition probabilities of 81, 9, and 10%, respectively, entering 3 separate folding channels.

The I1 pathway is characterized by the formation of an extremely stable Ic state after Ic. The Ic state transited to I1A by spreading its folded structure from \( (\beta \alpha)_{-5} \) to \( (\beta \alpha)_{-8} \), leaving an unfolded \( \alpha_6 \) tethered by \( \beta_1 \). The I1A state has a 7-stranded barrel with native-like contacts between helices \( \alpha_6 \) and \( \alpha_8 \) that prevent the incorporation of \( \alpha_6 \) and \( \beta_1 \) into the barrel architecture. I1A persists in the great majority of the trajectories with only a small fraction escaping to dock \( \alpha_6 \) across the bottom of the barrel to form the I1A state. The I1A state has an unfolded \( \beta_1 \) with 2 of its ends fixed on the folded \( \beta \)-barrel. I1A then rapidly folds to the native state by the insertion of \( \beta_1 \) into the \( \beta \)-barrel through the
channel between \( \alpha_1 \) and \( \alpha_6 \). To test the stability of the I1A state, we further sampled another set of 100 trajectories beginning from the I1A state for 8,000 time units. Even with a quadrupled simulation time, only 17% of the trajectories reached the native state (SI Appendix, Fig. S7), confirming the extremely long lifetime of the I1A state.

In comparison with the I1 pathway, in which \( \alpha_0 \beta_1 \) is the last to fold, both the I2 and the I3 pathway require the \( \alpha_0 \beta_1 \) element to fold before the closure of the \( \beta \)-barrel. In the I2 pathway, Ic incorporates the \( \alpha_0 \beta_1 \) element but excludes the C-terminal \( (\alpha \beta)_7 \) elements to form the I2 state. I2 then readily folds to the native state by incorporating the C-terminal \( (\alpha \beta)_7 \). In the I3 pathway, Ic first forms 2 partially folded \( \alpha_2(\alpha \beta)_3 \) and \( \alpha_6(\alpha \beta)_7 \) before the docking of \( \alpha_0 \beta_1 \) on \( \alpha_6(\alpha \beta)_7 \) to form the I3 state. The I3 state has 2 partially folded halves of the \( \beta \)-barrel, the \( \alpha_2(\alpha \beta)_3 \) subdomain, and the \( (\alpha \beta)_9 + \alpha_6(\alpha \beta)_7 \) subdomain, linked by unfolded \( \beta_0 \alpha_1 \beta_2 \) and \( \beta_6 \) strands. Interestingly, the I1 state, fully connected by contacts from head to tail, then folds to native by cooperatively merging the 2 partially folded halves. More structural details of the I1A, I1B, I2, and I3 intermediates are shown in SI Appendix, Fig. S5.

To examine the relationship between the 2 regions of frustration (\( Q_0 = 0.50 \) to 0.65 and \( Q_0 = 0.75 \) to 0.85) found in the ensemble averaged analysis (Fig. 4C) and the 3 folding pathways, the \( Q_0 \) vs. \( Q_0 \) data for the N- and C-terminal halves of SsIGPS of the 3 folding pathways were compared with the data of all trajectories, as shown in SI Appendix, Fig. S6. The \( Q_0 = 0.50 \) to 0.65 and \( Q_0 = 0.75 \) to 0.85 regions, representing the early and final folding stages, respectively, differ in their sources of frustration. The great similarity of the \( Q_0 \) vs. \( Q_0 \) plots at \( Q_0 = 0.50 \) to 0.65 between all trajectories and the I1 pathway (SI Appendix, Fig. S6 A and B) indicates that the global frustration at \( Q_0 = 0.50 \) to 0.65 is primarily contributed by the frustration in the major I1 channel.

However, the global frustration at \( Q_0 = 0.75 \) to 0.85 is different, since no significant backtracking events were observed in the same region of all 3 folding channels (SI Appendix, Fig. S6 B–D). The 3 folding channels, differing in their assembly order of the protein, have very different values of \( Q_{\text{C-ter}} \) and \( Q_{\text{N-ter}} \) at different times. In the I1 and I2 channels, the C-terminal \( (\alpha \beta)_7 \) and the N-terminal \( \alpha \beta_1 \), respectively, are the last to fold. Before reaching the final stage of folding at \( Q_0 = 0.75 \) to 0.85, the \( Q_{\text{C-ter}} \) of the I1 channel reaches \( \sim 0.9 \), which is significantly larger than that of the I2 channel (\( \sim 0.7 \)). During most of the 2,000 time units the I1 pathway trajectories (74 of 81) were trapped in the extremely stable I1 state (\( Q_0 \sim 0.788 \)), which precludes sufficient sampling at \( Q_0 > 0.788 \) in the I1 channel and makes the I2 and I3 channels dominant in this region. Therefore, the backtracking of \( Q_{\text{C-ter}} \) at \( Q_0 = 0.75 \) to 0.85 is a result of the significantly smaller \( Q_{\text{C-ter}} \) of the I1 pathway and the sparse sampling of the I1 pathway in this region.

**Discussion**

A combined experimental and computational study of the folding reaction of the SsIGPS TIM barrel has revealed insights into the structures of partially folded states and the potential role of frustration that occurs in simulations of the folding reaction.

**Mechanistic Analysis.** Previous experimental studies of the folding kinetics generated a 5-species model, \( \text{IBP} \rightleftharpoons \text{U} \rightleftharpoons \text{I}_1 \rightleftharpoons \text{I}_2 \rightleftharpoons \text{N} \) (Fig. 6) (11). Current native-centric simulations predict a more complex model involving partitioning between 3 different pathways to reach the native conformation (Fig. 5). The data from both models can be used to generate “kinetic species” plots to

![Fig. 5. Multiple folding pathways discovered by simulations, the upper right legend shows the transition probabilities from Ic to I1A, I1B, I2, and I3. Additional structural details for I1A, I1B, I2, and I3 are shown in SI Appendix, Fig. S5. The gray contours show the overlay of ~50 protein conformations, sampled from the corresponding states. See Movies S1–S3 for animations.](image-url)
compare the flow of material from the U state to the N state during a folding reaction (Fig. 6).

The 2 kinetic species plots are remarkably similar in several respects, but obviously differ with respect to the number of folding channels. The CF-FL and CF-SAXS measurements report the collapse of the unfolded chains within 50 μs; however, the distance distribution of the α1-α8 FRET pair (Fig. 3A) indicates the presence of 2 states. One state is more compact than the native state, implying a nonnative structure, and the other more expanded than native but more compact than the unfolded state. Because the experimental kinetic species plot (Fig. 6A) predicts the simultaneous presence of the IGBP and IA states after a few milliseconds, with the IGBP state predominant, we presume that the overly compact state corresponds to IA and the expanded state to IGBP. Examination of the predictions of the simulations after ~200 time units (Fig. 6B and SI Appendix, Fig. S11) supports this interpretation when comparing the populations of the Ic and IA states. The subsequent increase in the population of the IA state in experiments is also mimicked by the IGBP state in the simulations. Particularly striking is the correspondence of the very long lifetimes of both the IGBP and IA states, consistent with their rate-limiting roles in folding by both experiment and simulations. Both experiments and the major refolding channel in the simulations then reveal a final intermediate, IB and I1B, respectively, before proceeding to the native state. The partitioning of Ic into 3 channels in the simulations is not evident in the experimental data. However, channels 2 and 3 each carry only ~10% of the population and would be difficult to detect experimentally. Both experiments and the major refolding channel in the simulations then reveal a final intermediate, IB and I1B, respectively, before proceeding to the N state.

**Structural Analysis.** The pairwise and global dimensional analysis provided by CF-trFRET and CF-SAXS enables a direct comparison with the results of the simulations on the structures of the unfolded state, U, and the IGBP/IA intermediates that appear in microseconds.

**U state.** SAXS measurements of the unfolded state in high concentrations of urea (Fig. 2A), when extrapolated to the absence of denaturant, yield an estimated Rg in water, 46 ± 5 Å, that is consistent with a random-coil ensemble for a chain of 226 amino acids (23). Remarkably, native-centric simulations of the U state (Fig. 4B) obtained the same estimate of Rg, ~45 Å, and both approaches revealed the breadth of the unfolded manifold of conformers (Figs. 2D and 4B).

**I GBP/IA state.** The trFRET data for the α1-α8 pair show that the microsecond folding reaction partitions into 2 distinct distributions with different degrees of contraction (Fig. 3A). One ensemble is more compact than that for the N state and the other much more expanded. Unfortunately, the limitations of FRET measurements of distance outside efficiencies of 0.2 to 0.8 preclude estimates of the relative populations of these distributions. As described above, these results are consistent with the previous global analysis of the folding of SsIGPS that found U partitioning into the IGBP and IA states (Fig. 6A). Although the α1-α8 pair distances of the Ic state from simulations (SI Appendix, Fig. S9A) do not capture the overall compact conformations that appear after 50 μs (Fig. 3A), native-centric simulations are incapable of detecting nonnative structures. As the time steps increase, more compact states appear at ~20 and ~10 Å (SI Appendix, Figs. S8A and S10B), reflecting the progression of the folding reaction toward the IGBP, I1B, and N states. The α1-α8 FRET pair show a single distribution centered near that for the N state (Fig. 3B), but the greater breadth of which indicates a larger, dynamic ensemble. The SAXS data reveal a denaturant-independent Rg of 26 Å below 2 M urea after 150 μs (Fig. 2A), demonstrating that this species is not a collapsed form of the unfolded state (15). The Kraty plot after 150 μs (Fig. 2C) is not consistent with a fully globular structure, and the associated P(r) (Fig. 2D) shows a peak at ~30 Å and a tail at longer distances.

The simulations show a remarkable degree of correspondence with experimental results for the structures that appear at the initial stage of folding for SsIGPS. The IGBP intermediate has a native-like structure in the (βα)2-5 segment, consistent with the formation of a secondary structure detected by previous HDX-mass spectrometry (HDX-MS) experiments (8, 11) and the distance measured for the α1-α8 FRET pair. The N- and C-terminal segments are unstructured and give rise to the tail at high values of the P(r) (Fig. 2D), very similar to that seen in the SAXS data. Both the IGBP and IA states are also mimicked by the I1A state in the simulations. Particularly striking is the correspondence of the very long lifetimes of both the IGBP and IA states, consistent with their rate-limiting roles in folding by both experiment and simulations. Both experiments and the major refolding channel in the simulations then reveal a final intermediate, IB and I1B, respectively, before proceeding to the N state.

**Frustration in Folding.** Two major regions of topological frustration were found in the ensemble averaged analysis of the simulation, shown in the Q1 vs. Q2 plots (Fig. 4C). Multiple asynchronous folding pathways complicate the descriptions of the frustration in folding.

**Frustration at Q1 = 0.50 to 0.65.** The frustration at Q1 = 0.50 to 0.65, where the IGBP/I1 state persists, is mainly contributed by the backtracking events in the dominate I1 channel. The backtracking event of the QN-ter at Q1 = 0.50 (Fig. 4C) corresponds primarily to the unfolding of the (βα)2-8 element (Fig. 4D). This conclusion is consistent with experimental results in which some premature structures in the IGBP/I1 state are required to unfold before reaching the productive folding pathway. The 2 minor pathways I2 and I3 show different outcomes of the IGBP/I1 state in this region. The I2 pathway shows backtracking in the C terminus while the I3 pathway shows no obvious frustration. Interestingly, the different sources of frustration in the I1 and I3 pathways reflect alternative forms of an incomplete TIM barrel. The major I1 pathway excludes the N terminus while the minor I2 pathway excludes the C terminus. Both contain the central (βα)2-5 region that is protected against HDX (11) and, evidently, is capable of propagating structure in either direction.

**Frustration at Q1 = 0.75 to 0.85.** The frustration at Q1 = 0.75 to 0.85 is a combined result of the 3 folding pathways that differ in their assembly order of the protein and therefore does not represent
actual loss of structures. The global backtracking of $Q_{\text{C-ter}}$ at $Q_t = 0.75$ to 0.85 is caused mainly by the $I_2$ channel in which the C-terminal (αL)₆ elements of the protein are the last to fold, thereby lowering the global $Q_{\text{C-ter}}$ value. Although no evidence of backtracking of the $I_{1A}$ state ($Q_t \sim 0.79$) was found, it remains possible that $I_{1A}$ may first partly unfold so that αL₆ folds before the barrel closure to reduce the tremendous entropic cost required to make the transition from $I_{1A}$ to $I_{1B}$.

Conclusions
A combined experimental and computational study of the folding reaction for a TIM barrel protein has yielded remarkable agreement between these studies’ complementary views of a complex process. The mechanism defined by the major refolding pathway in simulations agrees closely with the mechanism determined by a variety of experiments. Striking similarities include the formation of stable structure in the central region of the sequence early in folding and a rate-limiting step before the formation of an 8-stranded β-barrel. Global and pairwise distance measurements of the early intermediate find a very similar degree of compactness, likely with disordered tails at both termini. In contrast to the experiments, the simulations reveal the presence of 2 minor channels that delineate alternative pathways to the native conformation. The frustration in folding detected by the simulations results in the formation of a pair of cryptic TIM barrels that differ in the exclusion of either the N- or C-terminal segments of the protein, consistent with the presence of intermediates observed in experiments. Although it is likely that these incomplete barrels contain nonnative structures inaccessible to native centric simulations, the remarkable similarities in the minima on the experimental and computational folding free energy surfaces argue that they are dominated by native-like structures.

Materials and Methods
Protein Production. Protein was expressed in DE3 cells and purified using metal affinity, ion exchange, and sizing chromatography before labeling with 1,5-I-AEDANS. See SI Appendix for full details.

SAXS. SAXS measurements were performed at the BioCAT beamline at the Advanced Photon Source, Argonne, IL. Equilibrium SAXS measurements were performed by interfacing an autosampler running custom software to the standard quartz sample capillary (24). Kinetic experiments were performed as previously described (24) with the exception that flow to the quartz mixer for the kinetic experiments was controlled by syringe pumps (Harvard Apparatus) at a total flow rate of 4 to 5 mL min⁻¹. Scattering images were reduced using scripts provided by BioCAT and analyzed as previously described (18, 24).

Gō Model Simulations. The Cβ-based native centric coarse-grained model of SiGIPS was generated by an in-house script using Protein Data Bank (PDB) code 2C3Z. Molecular dynamics simulations were performed using the CHARMM package (25). Snapshots were saved every 10,000 time steps (1 time unit). The 2,000 time unit unbiased simulation sampled 100 independent trajectories at 230 K with unfolded initial configurations sampled at 510 K. See SI Appendix for further details.

ACKNOWLEDGMENTS. We thank Sagar Kathuria for expert assistance with the CF-SAXS and CF-FRET experiments. We thank Noah Cohen and Sagar Kathuria for helpful discussions and Divya Kapoor for help in creating the mutant constructs. This work was supported by National Institute of General Medical Sciences (NIGMS) Grant 5 R01 GM023303; the National Science Foundation (NSF) Division of Biological Infrastructure Grant 1353942; and the NSF Chemistry Division Grant 1506273 (to C.L.B.). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by Grant 9 P41 GM103622 from the NIGMS of the National Institutes of Health (NIH). Use of the Pilatus 3 T detector was provided by Grant 1S10OD018090-01 from NIGMS. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the NIGMS or the NIH.