

# Conservation of the Folding Mechanism between Designed Primordial $(\beta\alpha)_8$ -Barrel Proteins and Their Modern Descendant

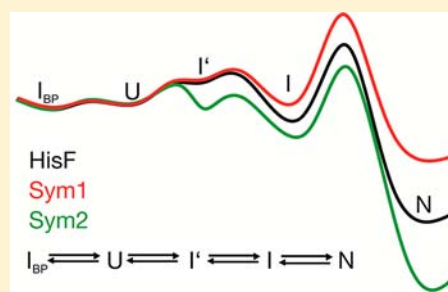
Linn Carstensen,<sup>†</sup> Josef M. Sperl,<sup>†</sup> Marco Bocola,<sup>†</sup> Felix List,<sup>†</sup> Franz X. Schmid,<sup>‡</sup> and Reinhard Sterner<sup>\*†</sup>

<sup>†</sup>Institut für Biophysik und physikalische Biochemie, Universität Regensburg, D-93040 Regensburg, Germany

<sup>‡</sup>Laboratorium für Biochemie und Bayreuther Zentrum für Molekulare Biowissenschaften, Universität Bayreuth, D-95440 Bayreuth, Germany

**S** Supporting Information

**ABSTRACT:** The  $(\beta\alpha)_8$ -barrel is among the most ancient, frequent, and versatile enzyme structures. It was proposed that modern  $(\beta\alpha)_8$ -barrel proteins have evolved from an ancestral  $(\beta\alpha)_4$ -half-barrel by gene duplication and fusion. We explored whether the mechanism of protein folding has remained conserved during this long-lasting evolutionary process. For this purpose, potential primordial  $(\beta\alpha)_8$ -barrel proteins were constructed by the duplication of a  $(\beta\alpha)_4$  element of a modern  $(\beta\alpha)_8$ -barrel protein, imidazole glycerol phosphate synthase (HisF), followed by the optimization of the initial construct. The symmetric variant Sym1 was less stable than HisF and its crystal structure showed disorder in the contact regions between the half-barrels. The next generation variant Sym2 was more stable than HisF, and the contact regions were well resolved. Remarkably, both artificial  $(\beta\alpha)_8$ -barrels show the same refolding mechanism as HisF and other modern  $(\beta\alpha)_8$ -barrel proteins. Early in folding, they all equilibrate rapidly with an off-pathway species. On the productive folding path, they form closely related intermediates and reach the folded state with almost identical rates. The high energy barrier that synchronizes folding is thus conserved. The strong differences in stability between these proteins develop only after this barrier and lead to major changes in the unfolding rates. We conclude that the refolding mechanism of  $(\beta\alpha)_8$ -barrel proteins is robust. It evolved early and, apparently, has remained conserved upon the diversification of sequences and functions that have taken place within this large protein family.



## INTRODUCTION

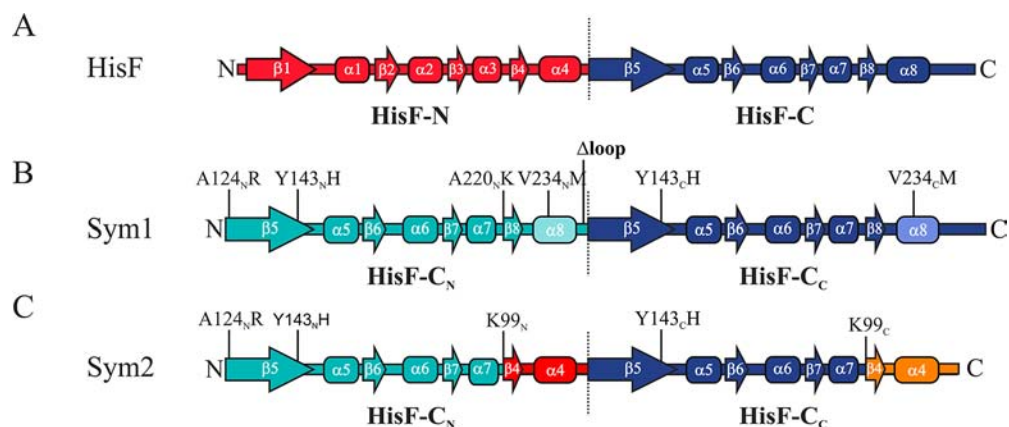
Natural proteins are usually characterized by high energy barriers between the denatured and the native state, which leads to cooperative folding and protection against unfolding, aggregation, and premature degradation.<sup>1</sup> In contrast, proteins designed in the laboratory lacking an “evolutionary history” often show low-cooperativity equilibrium transitions, and almost barrier-less rapid folding and unfolding kinetics.<sup>2–5</sup> The differences between natural and artificial proteins indicate that cooperative folding and distinct energy barriers are not intrinsic physicochemical properties of proteins but the consequence of natural selection.<sup>6</sup> Presumably, the selection pressure for efficient folding has been operative since the onset of protein evolution. Therefore, the comparison of a modern protein with its potential ancestors might provide insights into the evolution of folding mechanisms.

The  $(\beta\alpha)_8$ -barrel scaffold is one of the most ancient, frequent, and versatile protein structures and used by a multitude of enzymes that catalyze more than 60 different reactions from five out of the six EC classes.<sup>7–9</sup> The canonical  $(\beta\alpha)_8$ -barrel contains at least 200 amino acids and is composed of eight modules. Each module consists of a  $\beta$ -strand and an  $\alpha$ -helix linked by a  $\beta\alpha$ -loop; the individual modules are connected by  $\alpha\beta$ -loops. The eight strands form a closed parallel  $\beta$ -sheet, the barrel, which is surrounded by an outer layer of eight  $\alpha$ -helices.

In all characterized  $(\beta\alpha)_8$ -barrels, residues important for substrate specificity and catalysis are found at the C-terminal ends of the central  $\beta$ -strands and in the subsequent  $\beta\alpha$ -loops. This conserved location of the active site suggests that the contemporary  $(\beta\alpha)_8$ -barrels might have evolved from of a common ancestor.<sup>10–13</sup> Moreover, the modular structure of the fold points to an evolutionary precursor that has emerged by the duplication and fusion of individual  $\beta\alpha$ -entities. In fact, the  $(\beta\alpha)_8$ -barrel enzymes *N*'-[(*S*'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA) and imidazole glycerol phosphate synthase (HisF) display a weak 4-fold and a strong 2-fold internal symmetry, suggesting that they evolved by the stepwise duplication and fusion of  $(\beta\alpha)_2$ -quarter and  $(\beta\alpha)_4$ -half-barrels.<sup>14–18</sup> Previously we aimed to reconstruct the evolutionary pathway from a  $(\beta\alpha)_4$ -half-barrel to a stable  $(\beta\alpha)_8$ -barrel by using the C-terminal half-barrel of HisF (HisF-C) as a model. In a first step, a symmetric  $(\beta\alpha)_8$ -barrel was constructed by tandem fusion of two copies of HisF-C and then stabilized by several rounds of design and selection. The resulting artificial  $(\beta\alpha)_8$ -barrel protein originally termed HisF-C\*\*\*C<sup>17</sup> is renamed here as Sym1, and its two half-barrels are designated as HisF-C<sub>N</sub> and HisF-C<sub>C</sub> (Figure 1). Crystal structure analysis showed that the

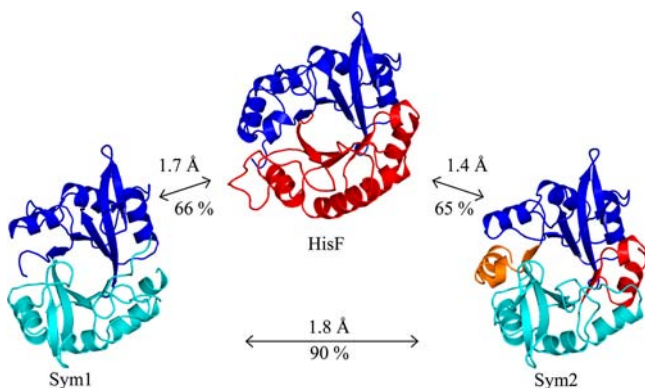
Received: May 22, 2012

Published: July 3, 2012



**Figure 1.** Design of the artificial  $(\beta\alpha)_8$ -barrel proteins Sym1 and Sym2 from two identical and fused half-barrels. (A) Secondary structure elements of wild-type HisF. The N-terminal half-barrel HisF-N [modules  $(\beta\alpha)_{1-4}$ ] is shown in red, the C-terminal half-barrel HisF-C [modules  $(\beta\alpha)_{5-8}$ ] in blue. (B) Design of Sym1. Two copies of HisF-C were fused and stepwise stabilized to reach Sym1 (formerly denoted as HisF-C\*\*\*C) by the indicated amino acid substitutions, and by shortening of the loop connecting the two half barrels.<sup>15–17</sup> (C) Design of Sym2. Sym2 was derived from Sym1 by replacing  $\beta\alpha$  module 8 with  $\beta\alpha$  module 4 in both half-barrels (red and orange).

$(\beta\alpha)_8$ -barrels of Sym1 and wild-type HisF are highly similar.<sup>17,19</sup> Sym1 is, however, less stable than HisF, and the termini of the two fused half-barrels are not resolved in the crystal structure (Figure 2).



**Figure 2.** Ribbon diagrams of the crystal structures of Sym1 (2w6r.pdb),<sup>17</sup> HisF (1thf.pdb)<sup>19</sup> and Sym2 (3og3.pdb). The invisible parts in the crystal structure of Sym1 are fully resolved in Sym2. Colors represent the origin and location of secondary structure elements as described in Figure 1. Amino acid sequence identities (%) and rmsd values (Å) of corresponding  $C\alpha$ -atoms as deduced from pairwise structure-based sequence alignments are indicated (HisF-Sym1: 184 superimposed  $C\alpha$  atoms; HisF-Sym2: 198 superimposed  $C\alpha$  atoms; Sym1-Sym2: 191 superimposed  $C\alpha$  atoms).

We have now stabilized Sym1 by replacing  $\beta\alpha$ -module 8 from HisF-C with the stable  $\beta\alpha$ -module 4 stemming from HisF-N, the N-terminal half of HisF (Figure 1). We solved the crystal structure of the resulting Sym2 protein (Figure 2) and compared it in its unfolding and refolding kinetics with Sym1 and the wild-type protein.<sup>20</sup> The three proteins show cooperative unfolding transitions and share a common sequential refolding mechanism, which comprises two productive intermediates and a final high energy barrier toward the native state. We conclude that this folding mechanism is an ancient property of  $(\beta\alpha)_8$ -barrel proteins.

## EXPERIMENTAL SECTION

**Expression of Genes and Purification of Recombinant Proteins.** Cloning of the genes *hisF* and *sym1* (formerly *hisF-C\*\*\*C*) into pET24a(+) was described previously.<sup>17,20</sup> The *sym2* gene was constructed and cloned into pET24a(+) as described in the Supporting Information. The genes were expressed in *E. coli* T7-Express cells (New England Biolabs). After induction with 0.5 mM IPTG, cells were grown for 4 h at 37 °C, and harvested. HisF was purified by heat precipitation of the host cell proteins, followed by anion exchange as described.<sup>21</sup> Sym1 and Sym2 were purified as described for Sym1 by metal chelate affinity chromatography using the C-terminal His<sub>6</sub>-tag.<sup>17</sup> According to SDS-page, all proteins were pure to more than 95%. At least 100 mg of protein per liter of culture was obtained and dialyzed against 50 mM Tris/HCl buffer (pH 7.5).

**Equilibrium Unfolding Transitions and Kinetics of Unfolding and Refolding.** The thermodynamics and kinetics of unfolding/refolding of 4  $\mu$ M protein were measured at 25 °C in 50 mM Tris/HCl buffer (pH 7.5) containing different concentrations of GdmCl, basically as described.<sup>20</sup> GdmCl (ultrapure) was purchased from MP Biomedicals (Illkirch, France), and its concentration was determined by the refractive index of the solution.<sup>22</sup> Loss/gain of secondary structure after manual mixing was followed by the far-UV circular dichroism (CD) signal at 225 nm using a JASCO model J815 CD spectrophotometer (path-length, 5 mm; bandwidth, 1 nm). Loss/gain of the tertiary structure after manual mixing was followed by the fluorescence emission at 320 nm (bandwidth, 5 nm) after excitation at 280 nm (bandwidth, 3 nm) with a JASCO model FP-6500 spectrofluorimeter. For rapid reactions, fluorescence emission was followed using a 320 nm cutoff filter after excitation at 280 nm in a stopped-flow SX.20MV spectrometer from Applied Photophysics (Leatherhead, UK).

To reach equilibrium, HisF and Sym2 were preincubated at the indicated concentrations of GdmCl for 10 and 20 days at 45 °C, respectively, and then incubated for three weeks at 25 °C. The samples of Sym1 reached equilibrium after 3 days at 25 °C. The transitions were analyzed according to the two- or three-state equilibrium model, assuming a linear dependency of the free-energy of unfolding on the GdmCl concentration.<sup>23</sup> The rates of conventional unfolding and refolding kinetics were determined by fitting monoexponential or double-exponential equations to the data points using the software GraFit6 from Erithacus (West Sussex, UK).

The unfolding limbs of I' and I of Sym2 were determined by stopped-flow interrupted refolding experiments. Unfolded Sym2 (132  $\mu$ M in 6 M GdmCl) was diluted 6-fold to 1 M GdmCl, and incubated for 0.5 or 7 s to allow for the conversion of U to I' or I. The sample was then diluted 11-fold into a concentration range of 1.6–4 M GdmCl, which resulted into the rapid conversion of I' and I into U. A

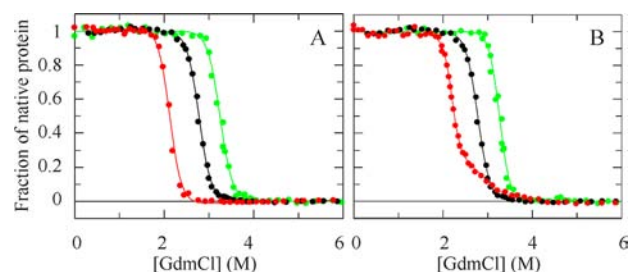
mono- or double-exponential equation with linear factor was fitted to the unfolding traces. The kinetics of the sequential  $U \rightarrow I' \rightarrow I \rightarrow N$  reaction was monitored in further stopped-flow fluorescence interrupted refolding experiments. Unfolded Sym2 (132  $\mu\text{M}$  in 6 M GdmCl) was again diluted 6-fold to 1.0 M GdmCl, and incubated for various times to allow for the conversion of  $U$  to  $I'$  or  $I$ . The sample was then diluted 11-fold to 2.6 M GdmCl, which resulted in the very rapid unfolding of  $I'$  and the fast unfolding of  $I$ . Double-exponential equations with linear factor were fitted to the kinetic traces. An equation describing the consecutive model ( $U \rightarrow I' \rightarrow I \rightarrow N$ ) was fitted to the obtained amplitudes using DynaFit (BioKin Ltd.).<sup>24</sup>

**Crystal Structure Determination of Sym2 and Structural Superimpositions.** Details of the crystallization and structure determination of Sym2 are described in the Supporting Information, the data-collection statistics are shown in Table S2. Structural superimpositions were performed using the program STAMP.<sup>25</sup>

## RESULTS AND DISCUSSION

**Design of Sym2 and Structural Comparison with HisF and Sym1.** In the crystal structure of Sym1, the  $\beta\alpha$ -modules 8 at the C-termini of HisF- $C_N$  and HisF- $C_C$  are not visible (Supporting Information Figure S1). We replaced these flexible and putatively labile regions by the stable  $\beta\alpha$ -module 4 from HisF-N (Figures 1, 2). The resulting Sym2 protein was produced, purified, and crystallized. Structure determination at 2.08 Å resolution revealed a  $(\beta\alpha)_8$ -barrel with high similarity to HisF and Sym1. The  $C\alpha$ -atoms of the three proteins, which show sequence identities between 65% (Sym1-HisF; Sym2-HisF) and 90% (Sym1-Sym2) superpose with a rmsd of 1.4–1.8 Å. Importantly, the regions at the C-termini of HisF- $C_N$  and HisF- $C_C$ , which could not be resolved in Sym1, are well-defined in the structure of Sym2 (Figure 2). Moreover, a stabilizing salt bridge cluster at the N-terminal face of the central  $\beta$ -barrel of HisF is also present in Sym1 and Sym2 (Supporting Information Figure S2). Interestingly, a protein similar to Sym2 was recently computationally designed using the ROSETTA software.<sup>26</sup>

**Sym2 is Thermodynamically More Stable than HisF and Sym1.** The thermodynamic stabilities of Sym1, HisF, and Sym2 were determined by GdmCl-induced equilibrium unfolding transitions. The loss of tertiary structure was probed by protein (tryptophan/tyrosine) fluorescence, the loss of secondary structure was probed by far-UV circular dichroism (CD). The unfolding and refolding curves for the individual proteins superimpose well, demonstrating that unfolding is reversible for all three proteins. However, equilibration is very slow, in particular for HisF<sup>20</sup> and Sym2 (Supporting Information Figure S3). The thermodynamic parameters derived by two-state analysis from the fluorescence-detected transitions (Figure 3A) are listed in Table 1. The three proteins show similar cooperativities ( $m$  values), indicating that they have comparably compact tertiary structures, but they differ strongly in stability ( $\Delta G_D$ ). Sym2 is significantly more stable, Sym1 significantly less stable than HisF. The  $\Delta G_D$  and  $m$  values derived from the CD-monitored unfolding transitions (Supporting Information Table S1) agree with those from the fluorescence-detected transitions. However, for Sym1 the main transition was followed by a second phase characterized by a small CD amplitude and low cooperativity, which indicates the population of a partially folded intermediate (Figure 3B). This second unfolding phase was not detected for Sym2 and HisF, presumably because their native states are still stable at the denaturant concentration at which the intermediate of Sym1 is populated.



**Figure 3.** Equilibrium unfolding transitions of Sym1 (red), HisF (black), and Sym2 (green). GdmCl-induced unfolding of 4  $\mu\text{M}$  protein was followed in 50 mM Tris/HCl buffer (pH 7.5) at 25 °C by (A) Trp/Tyr fluorescence (excitation, 280 nm; emission, 320 nm) and (B) far-UV CD at 225 nm. The continuous lines represent the fit of the two-state model (or three-state model in case of the far-UV CD signal of Sym1) to the normalized unfolding transitions, yielding the thermodynamic parameters listed in Tables 1 and Supporting Information Table S1.

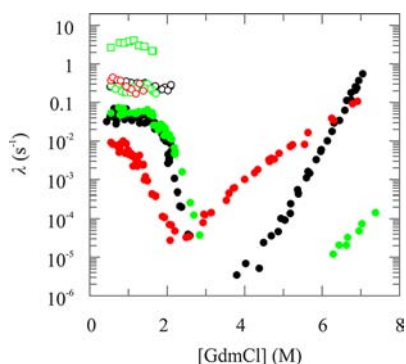
**Sym1, HisF, and Sym2 Form Burst-Phase Refolding Intermediates.** Folding and unfolding kinetics were followed by protein fluorescence and far-UV CD, after both manual and stopped-flow mixing. Refolding occurred in two (HisF, Sym1) or three phases (Sym2), whereas unfolding was monophasic for the three proteins. The fluorescence amplitudes of the unfolding and refolding reactions as well as the CD amplitude of the unfolding reaction accounted for the entire signal change as expected from the equilibrium transitions. However, the CD amplitude of the refolding reaction was 5-fold lower than expected even when followed after stopped-flow mixing. Apparently, at low GdmCl concentrations most of the ellipticity is recovered during the dead time of the experiment. Thus, the three proteins form burst-phase intermediates ( $I_{BP}$ ) within the first 5 ms of refolding. In all cases, the amplitude of the burst-phase reaction decreased with increasing GdmCl concentration in a sigmoidal manner (Supporting Information Figure S4), indicative of a cooperative unfolding of  $I_{BP}$ . A tentative two-state analysis testifies to a high content of compact and stable secondary structure (Table 1). Remarkably, for Sym1 the apparent CD unfolding transition of  $I_{BP}$  is congruent with the second, less cooperative phase of the equilibrium unfolding transition in Figure 3B, suggesting that, at moderate GdmCl concentrations,  $I_{BP}$  of Sym1 exists in equilibrium with the folded form of this protein. Taken all together, our findings suggest that the formation of  $I_{BP}$  early in folding is a common feature of Sym1, HisF, and Sym2.

**Sym1, HisF, and Sym2 Unfold with Extremely Different Rates.** The kinetics of refolding and unfolding of Sym1, HisF, and Sym2, measured as a function of the GdmCl concentration, are shown as chevron diagrams in Figure 4 and Figure S5. The rate constants determined from the fluorescence and far-UV CD signals of the individual proteins superimpose well, indicating that coupled changes in secondary and tertiary structure occur in all kinetic phases. Above 2 M GdmCl, refolding and unfolding are monophasic, indicating the absence of detectable kinetic intermediates. Sym1, HisF, and Sym2 differ strongly in the rate of unfolding. Sym1 denatures relatively fast even at moderate GdmCl concentrations, which allowed for the determination of the complete unfolding limb of the chevron diagram. In contrast, unfolding of HisF is very slow, and its rate could not be determined below 4 M GdmCl. The unfolding of Sym2 is further decelerated. It is roughly 1000-fold slower than the unfolding of HisF and cannot be

Table 1. Thermodynamic Stability Parameters for HisF, Sym1, and Sym2

	N $\leftrightarrow$ U			I <sub>BP</sub> $\leftrightarrow$ U		
	$\Delta G_D$ (kJ mol <sup>-1</sup> )	$m$ (kJ mol <sup>-1</sup> M <sup>-1</sup> )	[D] <sub>1/2</sub> (M)	$\Delta G_D$ (kJ mol <sup>-1</sup> )	$m$ (kJ mol <sup>-1</sup> M <sup>-1</sup> )	[D] <sub>1/2</sub> (M)
HisF	53.5 $\pm$ 1.7	19.1 $\pm$ 0.6	2.8	15.2 $\pm$ 2.4	7.9 $\pm$ 0.9	1.9
Sym1	42.8 $\pm$ 0.2	20.2 $\pm$ 0.4	2.1	13.2 $\pm$ 0.4	5.9 $\pm$ 0.2	2.2
Sym2	62.2 $\pm$ 3.9	19.2 $\pm$ 1.2	3.2	35.7 $\pm$ 6.5	21.4 $\pm$ 3.7	1.7

For the N  $\leftrightarrow$  U transition, the Gibbs free energy of denaturation ( $\Delta G_D$ ), the cooperativity ( $m$ ), and the denaturant concentration required to unfold 50% of the protein ( $[D]_{1/2}$ ) were obtained by analyzing the fluorescence-detected unfolding transitions (see Figure 3A) with the two-state model. For the I<sub>BP</sub>  $\leftrightarrow$  U transition, the parameters were obtained by analyzing the initial values of the refolding kinetics as monitored by the CD signal at 225 nm (see Supporting Information Figure S4).

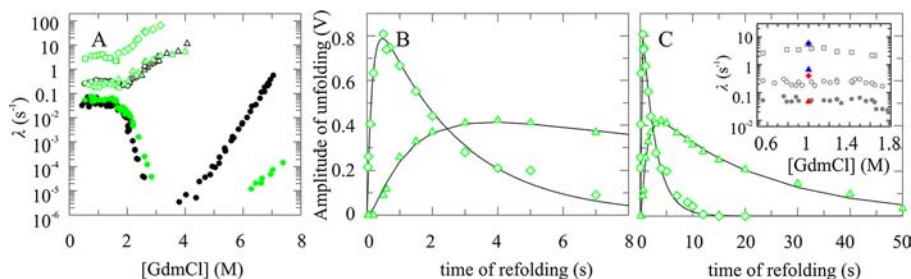


**Figure 4.** Folding kinetics of Sym 1 (red), HisF (black), and Sym2 (green). The dependence on GdmCl concentration of the apparent rate constant ( $\lambda$ ) is shown for the slow unfolding and refolding phases (filled circles), for the fast refolding phase (open circles), and for the additional very rapid refolding phase of Sym2 (squares). Chevron diagrams of the individual proteins are shown in Supporting Information Figure S5.

followed in reasonable time below 6 M GdmCl. The strong differences in the unfolding rates account to a large extent for the observed differences in the thermodynamic stabilities of the three proteins (Table 1). The extreme disparities between the unfolding rates of Sym1 and Sym2 are remarkable given their high amino acid sequence identity of 90%. Apparently, they are caused by the replacement of  $\beta\alpha$  module 8 in Sym1 with  $\beta\alpha$  module 4 in Sym2 (Figure 1, Supporting Information Figure S1). Our findings point to the significance of optimizing the

contact regions of the fused half-barrels as an important evolutionary step following the duplication and fusion of the ancestral ( $\beta\alpha$ )<sub>4</sub>-module.

**Sym1, HisF, and Sym2 Show Similar Refolding Kinetics.** The refolding kinetics of Sym1, HisF, and Sym2 are remarkably similar. At low denaturant concentrations a fast reaction with a time constant  $\tau$  of about 3 s is observed for all three proteins (Figure 4). This reaction is followed by a slower one, again with a common  $\tau$  of about 20 s for Sym2 and HisF. For Sym1 the slow reaction shows an increased time constant of 175 s, presumably because its folding intermediates are much less stable than those of HisF and Sym2. Sym2 shows an additional very rapid folding reaction with  $\tau = 0.27$  s, which could not be monitored for the other two proteins (Figure 4). Importantly, at low concentrations of GdmCl, the observed refolding rate constants for all observed kinetic phases of the three proteins are largely independent of the amount of denaturant. Such “roll-over” behavior is indicative of the formation of folding intermediates.<sup>27</sup> The reciprocal change of the amplitudes of the fast 3-s phase and the slow 20-s phase of HisF folding with increasing concentration of GdmCl (Supporting Information Figure S6B) was previously shown to reflect the formation of a GdmCl-sensitive on-pathway folding intermediate I and its subsequent transformation to the native state in a sequential folding mechanism.<sup>20</sup> For Sym2, the amplitudes of the three consecutive folding phases also change in a reciprocal manner with denaturant concentration (Supporting Information Figure S6C). This suggests that the folding of this stable artificial protein also occurs sequentially



**Figure 5.** Kinetic intermediates during the folding of HisF (black) and Sym2 (green). (A) Chevron diagram. Symbols for rate constants determined by single-mixing experiments were taken from Figure 4. The unfolding limb of the on-pathway intermediate I of HisF (black triangles) was determined by interrupted refolding experiments as described.<sup>20</sup> The unfolding limbs of the on-pathway intermediates I' (green diamonds) and I (green triangles) of Sym2 were determined by interrupted refolding experiments. After exposure to refolding conditions (1.0 M GdmCl) for 0.5 s to populate I' or for 7 s to populate I, the formed intermediates were unfolded at the indicated GdmCl concentrations. (B, C) Amplitudes of the fast (green diamonds,  $\tau = 56$  ms) and slow (green triangle,  $\tau = 2.6$  s) unfolding reaction of Sym2 in 2.6 M GdmCl after refolding for the indicated times in 1.0 M GdmCl. The amplitudes are shown for 8 s (B) and for 50 s (C) of refolding. Continuous lines represent the results of the analysis of the data with an equation describing a consecutive process (U  $\rightarrow$  I'  $\rightarrow$  I  $\rightarrow$  N). Inset to panel c: The rate constants for the formation and depletion of I' and I as deduced from the amplitude analysis (blue diamond, rate of formation of I',  $\tau = 0.17$  s; red diamond, rate of depletion of I',  $\tau = 2.5$  s; red open triangle, lag in the formation of I,  $\tau = 0.18$  s; blue triangle, rate of formation of I,  $\tau = 1.5$  s; red filled triangle, rate of depletion of I,  $\tau = 21.3$  s) are compared with the refolding rates of Sym2 determined by conventional refolding kinetics (gray symbols, data taken from Figure 4).

via two intermediates, but the existence of parallel folding pathways can formally not be ruled out. For Sym1 the slow folding phase dominates already at low GdmCl concentrations (Figure S6A), emphasizing that its folding intermediate shows a very low stability toward unfolding by GdmCl.

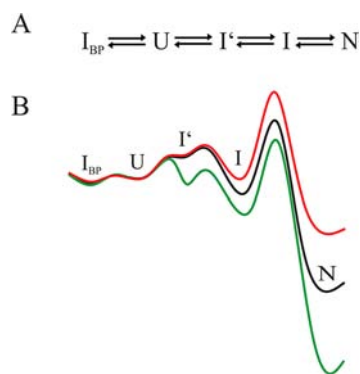
**A Common Sequential Folding Mechanism for Sym1, HisF, and Sym2.** For HisF, the formation of the folding intermediate I could previously be unambiguously confirmed by interrupted refolding assays.<sup>20</sup> This approach was now also used to follow the formation and the decay of intermediates during the refolding of Sym2. The time constants of the three folding phases of Sym2 at 1.0 M GdmCl are 0.27 s, 4.3 s, and 20 s (Figure 4). The product of the 0.27-s reaction should be well populated after 0.5 s of refolding whereas the product of the 4.3-s reaction should dominate after 7 s. Therefore, refolding was interrupted after 0.5 or 7 s, and the samples were transferred to unfolding conditions between 1.9 and 4 M GdmCl. The observed fluorescence decrease caused by unfolding after 0.5 or 7 s of refolding was mono- or biphasic, respectively, as shown for 2.9 M GdmCl in Supporting Information Figure S7. After 0.5 s of refolding, a single species, the intermediate I', was formed, which unfolded very rapidly ( $\tau = 0.02$  s). After 7 s of refolding, in addition to the 0.02-s unfolding reaction a second, 20-fold slower unfolding reaction ( $\tau = 0.39$  s) was monitored, suggesting that a further, more stable species, the intermediate I, had formed. Unfolding assays were performed between 1.9 and 4 M GdmCl, and under all conditions, the unfolding reactions of the intermediates, I' and I, were at least  $10^7$ -fold faster than the unfolding of native Sym2 (Figure 5A). In the transition region around 1.7 and 2.1 M GdmCl the unfolding rates of I' and I as measured in the interrupted refolding assays connect smoothly with the previously determined refolding rates, showing that the folding of these intermediates is quasi-reversible, because the final conversion to the native form of Sym2 is slow. Remarkably, the rate constants for formation and denaturation of the intermediate I are virtually identical for HisF and Sym2, suggesting that the on-pathway intermediates of the two proteins are comparably stable (Figure 5A).

To examine whether I' and I are on-pathway intermediates in a sequential reaction leading to native Sym2, the time course of their accumulation and depletion was analyzed in further stopped-flow interrupted refolding experiments. In these experiments, Sym2 was allowed to refold in 1.0 M GdmCl for variable time intervals, followed by the transfer to 2.6 M GdmCl. At this denaturant concentration, native Sym2 molecules remain folded (Figure 3), whereas the intermediates I' and I unfold with time constants of 0.05 and 0.48 s (Figure 5A, Supporting Information Figure S8). This large difference in rates allowed us to determine the unfolding amplitudes of I' and I simultaneously with high precision. The unfolding amplitudes are proportional to the concentrations of I' and I present at the time when refolding was interrupted, and therefore they trace the time course of the formation and depletion of the intermediates. The fraction of I' initially increases, passes through a maximum, and then decreases. The subsequent intermediate I forms with an initial lag, because the concentration of I' is zero at time zero. Then the fraction of I increases with the same rate as I' is depleted, and it decreases with the rate constant of the formation of the native state N (Figure 5B). Importantly, the rate constants for the formation and depletion of the various species obtained from the interrupted refolding assays coincided with those observed

directly by conventional refolding kinetics (Figure 5C). These findings provide compelling evidence for a sequential folding mechanism where I' is directly transformed to I and then further to native Sym2.

Interestingly, although I' is formed very rapidly, the overall folding process of Sym2 is not accelerated compared to HisF as the rate constants for the formation of I and N are virtually identical for the two proteins (Figure 5A). This observation implies a similar folding landscape for Sym2 and HisF with the modification that I' is energetically favored in the case of Sym2 but for an unknown reason not detectable in the case of HisF. For Sym1, an unambiguous interrupted folding analysis was not possible, because the fast folding reaction is characterized by low amplitudes (Supporting Information Figure S6A). Nevertheless, the observation of two refolding phases with roll-over behavior (Figure 4) and the amplitude profiles (Figure S6A) point to a similar sequential mechanism for Sym1 as well.

Taken all together, our findings suggest a common folding mechanism for Sym1, HisF, and Sym2 (Figure 6). All three



**Figure 6.** (A) Unifying folding mechanism for Sym1, HisF, and Sym2. An off-pathway equilibrium between the unfolded state (U) and the burst-phase intermediate ( $I_{BP}$ ) precedes the formation of the on-pathway intermediates I' and I. For HisF and Sym1, I' is assumed to be a high energy intermediate. (B) Energy diagram for the folding of Sym1 (red), Sym2 (green), and HisF (black). The different heights of the energy barriers indicate different folding/unfolding rates but are not at scale.

proteins form a burst-phase intermediate  $I_{BP}$  that contains a significant amount of secondary structure but is most likely off-pathway.<sup>20</sup> Then, very rapidly the on-pathway intermediate I' is formed, which is populated for Sym2 but cannot be detected in the case of HisF and Sym1, probably because it is a high-energy intermediate.<sup>28</sup> I' is directly converted into the second on-pathway intermediate I with similar rates for all three proteins. In the final and rate-limiting step of folding, I is transformed into the native states of Sym1, HisF, and Sym2. At non-native conditions (above 2 M GdmCl), both folding and unfolding of Sym1, HisF, and Sym2 are described by a single phase (Figure 4), most probably because I' and I are destabilized by the denaturant and do not accumulate.

## CONCLUSION

Many computational and experimental studies have focused on how complex protein structures might have evolved from simple polypeptide fragments.<sup>29–32</sup> How the mechanism of protein folding has evolved and changed over these long evolutionary periods has, however, received little attention. We approached this problem by focusing on the  $(\beta\alpha)_8$ -barrel,

which is among the most ancient, frequent, and versatile enzyme structures. Present-day  $(\beta\alpha)_8$ -barrel proteins probably have arisen by the duplication and subsequent combination of two  $(\beta\alpha)_4$ -half-barrels.<sup>15–18</sup> The artificial proteins Sym1 and Sym2 were constructed as models for such early fusion proteins by duplicating the C-terminal half barrel HisF-C of the modern protein HisF. Sym1 appears to be the more primordial version. The contact regions between the two half-barrels are still disordered, and its kinetic and conformational stability is low. In Sym2, the contact regions are optimized, which increases its stability to a level beyond the parent protein HisF. Notwithstanding the large differences in stability, the two models for a primordial  $(\beta\alpha)_8$ -barrel protein show the same folding mechanism, and they share this mechanism with the present-day HisF and other  $(\beta\alpha)_8$ -barrel proteins.<sup>20,33,34</sup> After a rapid equilibration with an off-pathway intermediate ( $I_{BP}$ ), the three proteins fold on a sequential pathway via partially folded intermediates (Figure 6). The most strongly populated intermediate I accumulates at identical positions along the pathway and is separated from the native state by a common activation barrier, which has a similar height for the two artificial constructs and the natural protein HisF. In this property, Sym1 and Sym2 differ from other designed proteins, which generally fold on complex energy landscapes without a significant common energy barrier.<sup>2–6</sup> These findings suggest that a synchronized and ordered sequential folding mechanism with a high energy barrier arose together with the emergence of the earliest  $(\beta\alpha)_8$ -barrel proteins and has essentially remained unaltered ever since. It will be interesting to see whether the folding mechanism of other protein structures has been conserved in a similar fashion during evolution.

## ■ ASSOCIATED CONTENT

### Supporting Information

A methods section describing the construction and cloning of the *sym2* gene, and the crystallization and structure determination of the Sym2 protein. Supplementary references. Figures showing a schematic representation of the  $\beta\alpha$ -composition of Sym1, HisF, and Sym2 (Figure S1), important structural features of Sym2 in comparison to HisF and Sym1 (Figure S2), GdmCl-induced equilibrium unfolding transitions of (A) HisF and (B) Sym2 (Figure S3), equilibrium unfolding transitions of (A) Sym1, (B) HisF, and (C) Sym2 in comparison with the unfolding transitions of their burst-phase intermediates ( $I_{BP}$ ) (Figure S4), GdmCl-dependence of the apparent rate constants ( $\lambda$ ) for refolding and unfolding of (A) Sym1, (B) HisF, and (C) Sym2 (chevron diagrams; Figure S5), amplitudes of the very rapid, fast and slow refolding phases of (A) Sym1, (B) HisF, and (C) Sym2 (Figure S6), unfolding kinetics of the putative folding intermediates  $I'$  and  $I$  of Sym2 after interrupted refolding (Figure S7), detection of the folding intermediates  $I'$  and  $I$  of Sym2 by interrupted refolding experiments (Figure S8). Tables listing the thermodynamic parameters of Sym2, HisF, and Sym1 determined by far-UV CD (Table S1) and data collection and refinement statistics for the structure determination of Sym2 (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

Reinhard.Sternier@biologie.uni-regensburg.de

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Dr. Jochen Reinstein (Max Planck Institute for Medical Research, Heidelberg) for stopped-flow CD measurements, Christiane Endres and Barbara Kellerer (University of Regensburg) for help with protein purification and crystallization of Sym2, and Dr. Gabriel Zoldák (Technical University of Munich) for discussion.

## ■ REFERENCES

- (1) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.
- (2) Sadqi, M.; de Alba, E.; Pérez-Jiménez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4127.
- (3) Gillespie, B.; Vu, D. M.; Shah, P. S.; Marshall, S. A.; Dyer, R. B.; Mayo, S. L.; Plaxco, K. W. *J. Mol. Biol.* **2003**, *330*, 813.
- (4) Scalley-Kim, M.; Baker, D. *J. Mol. Biol.* **2004**, *338*, 573.
- (5) Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. *Cell* **2007**, *128*, 613.
- (6) Sanchez-Ruiz, J. M. *Biophys. Chem.* **2010**, *148*, 1.
- (7) Caetano-Anollés, G.; Kim, H. S.; Mittenthal, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9358.
- (8) Wierenga, R. K. *FEBS Lett.* **2001**, *492*, 193.
- (9) Sterner, R.; Höcker, B. *Chem. Rev.* **2005**, *105*, 4038.
- (10) Copley, R. R.; Bork, P. *J. Mol. Biol.* **2010**, *303*, 627.
- (11) Nagano, N.; Orengo, C. A.; Thornton, J. M. *J. Mol. Biol.* **2002**, *321*, 741.
- (12) Reardon, D.; Farber, G. *FASEB J.* **1995**, *9*, 497.
- (13) Pujadas, G.; Palau, J. *Biologica (Bratislava)* **1999**, *54*, 231.
- (14) Richter, M.; Bosnali, M.; Carstensen, L.; Seitz, T.; Durchschlag, H.; Blanquart, S.; Merkl, R.; Sterner, R. *J. Mol. Biol.* **2010**, *398*, 763.
- (15) Höcker, B.; Claren, J.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16448.
- (16) Seitz, T.; Bocola, M.; Claren, J.; Sterner, R. *J. Mol. Biol.* **2007**, *372*, 114.
- (17) Höcker, B.; Lochner, A.; Seitz, T.; Claren, J.; Sterner, R. *Biochemistry* **2009**, *48*, 1145.
- (18) List, F.; Sterner, R.; Wilmanns, M. *ChemBioChem* **2011**, *12*, 1487.
- (19) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. *Science* **2000**, *289*, 1546.
- (20) Carstensen, L.; Zoldak, G.; Schmid, F. X.; Sterner, R. *Biochemistry* **2012**, *51*, 3420.
- (21) Thoma, R.; Obmolova, G.; Lang, D. A.; Schwander, M.; Jenö, P.; Sterner, R.; Wilmanns, M. *FEBS Lett.* **1999**, *454*, 1.
- (22) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.
- (23) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063.
- (24) Kuzmič, P. *Anal. Biochem.* **1996**, *237*, 260.
- (25) Russell, R. B.; Barton, G. J. *Proteins* **1992**, *14*, 309.
- (26) Fortenberry, C.; Bowman, E. A.; Proffitt, W.; Dorr, B.; Combs, S.; Harp, J.; Mizoue, L.; Meiler, J. *J. Am. Chem. Soc.* **2011**, *133*, 18026.
- (27) Lorenz, T.; Reinstein, J. *J. Mol. Biol.* **2008**, *381*, 443.
- (28) Bachmann, A.; Kiefhaber, T. *J. Mol. Biol.* **2001**, *306*, 375.
- (29) Claren, J.; Malisi, C.; Höcker, B.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3704.
- (30) Alva, V.; Remmert, M.; Biegert, A.; Lupas, A. N.; Söding, J. *Protein Sci.* **2010**, *19*, 124.
- (31) Yadid, I.; Kirshenbaum, N.; Sharon, M.; Dym, O.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7287.
- (32) Lee, J.; Blaber, S. I.; Dubey, V. K.; Blaber, M. *J. Mol. Biol.* **2011**, *407*, 744.
- (33) Forsyth, W. R.; Bilsel, O.; Gu, Z.; Matthews, C. R. *J. Mol. Biol.* **2007**, *372*, 236.
- (34) Forsyth, W. R.; Matthews, C. R. *J. Mol. Biol.* **2002**, *320*, 1119.